SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(89th Edition)

MAY 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 Ninth** 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 **May, 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 learning architecture of RA-DLNet for visual sentiment analysis, ***6.Ashima Yadav*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

2. A deeply coupled ConvNet for human activity recognition using dynamic and RGB images, ***6.Tej Singh*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

3. A Multi-Phase Ensemble Model for Long Term Hourly Load Forecasting, ***8.Kushagra Bhatia***, ***8.Rajat Mittal***, ***8.Nisha*** and ***3.Madan Mohan Tripathi***, Electrical, DTU

4. A novel approach for multi-cue feature fusion for robust object tracking, ***8.Ashish Kumar***, Gurjit Singh Walia and ***3.Kapil Sharma***, IT, DTU

5. A Parallel Military Dog based Algorithm for Clustering Big data in Cognitive Industrial Internet of Things, Ashish Kumar Tripathi, ***3.Kapil Sharma[1]***, Manju Bala, ***3.Akshi Kumar[2]***, Varun G Menon and Ali Kashif Bashir, [1]IT and [2]Computer, DTU

6. Adaptive Droop based Control for Active Power Sharing in Autonomous Microgrid for Improved Transient Performance, Surya Prakash, ***8.1.Vaibhav Nougain*** and Sukumar Mishra, Electrical, DTU

7. An off-center-fed compact wideband antenna with truncated corners and parasitic patches for circular polarization, ***3.1.Sounik Kiran Kumar Dash***, Qingsha S. Cheng and Taimoor Khan, DTU

8. Analyzing the inter-relation between workplace spirituality and constructive deviance, ***3.Naval Garg*** and Anubhuti Saxena, USME, DTU

9. APPLICATIONS OF DIFFERENTIAL SUBORDINATIONS TO CERTAIN CLASSES OF STARLIKE FUNCTIONS, ***6.Shagun Banga*** and ***3.S. Sivaprasad Kumar***, Mathematics, DTU

10. Autism Detection using r-fMRI: Subspace Approximation and CNN Based Approach, ***3.Daya Gupta***, ***8.Ishani Vij*** and ***8.Mansi Gupta***, Computer Science, DTU

11. Channel Characterization for Devices in a Turbulent Diffusive Environment: A Mobile Molecular Communication Approach, Nilay Pandey, Sandeep Joshi, Ranjan K. Mallik and ***7.1.Brejesh Lall***, Electronics, DTU

12. Chemical composition and thermoanalytical analysis of the nest of yellow wasp and mud dauber, ***8.1.Anurag Sanghmitra***, ***8.1.Anurag Aswal***, ***8.1.Ankita Maurya***, ***8.Apurv Arora***, Anuradha Saha, ***6.Temesgen Hailu*** and ***3.Rajinder K Gupta***, Applied Chemistry, DTU

13. Classification of Cuisines from Sequentially Structured Recipes, ***8.Tript Sharma***, Utkarsh Upadhyay and Ganesh Bagler, Mechanical, DTU

14. Design of Tall Chimney for Thermal Power Plant, ***8.1.Karan Dhall***, Civil, DTU

15. Domineering games with minimal number of moves, ***8.Rohan Karthikeyan*** and ***8.Siddharth Sinha***, Mathematics, DTU

16. Effect of Friction Stir Processing on Microstructure and Mechanical Properties of TIG Welded Joint of AA6061 and AA7075, ***6.Husain Mehdi*** and ***3.R. S. Mishra***, Mechanical, DTU

17. Flavones Fluorescence-Based Dual Response Chemosensor for Metal Ions in Aqueous Media and Fluorescence Recovery, Nisha Fatma, ***3.Mohan Singh Mehata***, Nupur Pandey and Sanjay Pant, Physics, DTU

18. GA based optimization of tri-bological behaviour of diamond coated tungsten carbide, Ramakant Rana, ***3.Qasim Murtaza*** and R.S. Walia, Mechanical, DTU

19. HandAid: A Seven DoF Semi-Autonomous Robotic Manipulator, ***8.Vipul Garg***, ***8.Tript Sharma***, ***8.Aakash Kumar*** and ***3.Vikas Rastogi***, Mechanical, DTU

20. Hierarchical Clustering of World Cuisines, ***8.Tript Sharma[1]***, Utkarsh Upadhyay, ***Jushaan Kalra[2]***, Sakshi Arora, Saad Ahmad, Bhavay Aggarwal and Ganesh Bagler, [1]Mechanical and [2]Computer, DTU

21. Illumination Estimation for Nature Preserving Low-Light Image Enhancement, ***6.Kavinder Singh*** and ***3.Anil Singh Parihar***, Computer Science, DTU

22. Impact of Variability on Processor Performance in Negative Capacitance FinFET Technology, Hussam Amrouch, ***8.1.Girish Pahwa***, Amol D. Gaidhane, Chetan K. Dabhi, Florian Klemme , Om Prakash and Yogesh Singh Chauhan, Electronics, DTU

23. IMPLEMENTATION OF ELLIPTIC – CURVE CRYPTOGRAPHY, ***8.Monika***, ***8.Tanya Tomar***, ***8.Vineet Kumar*** and ***8.Yogesh Kumar***, Mathematics, DTU

24. In-Depth Analysis of Structures, Materials, Models, Parameters, and Applications of Organic Light-Emitting Diodes, SHUBHAM NEGI, ***3.POORNIMA MITTAL*** and BRIJESH KUMAR, Electronics, DTU

25. Intrusion detection systems: A Comparison of soft computing techniques, ***3.Anamika Chauhan*** and ***3.Kapil Sharma***, IT, DTU

26. Key Components of Rectenna System: A Comprehensive Survey, Daasari Surender, Taimoor Khan, Fazal A. Talukdar, ***3.Asok De***, Yahia M.M. Antar and Al. P. Freundorfer, Electronics & Communication, DTU

27. Learning Interpretable Hidden State Structures for Handwritten Numeral Recognition, ***3.Seba Susan*** and ***7.Jatin Malhotra***, IT, DTU

28. Material removal analysis of hybrid EDM-assisted centrifugal abrasive flow machining process for performance enhancement, ***6.Parvesh Ali***, R. S. Walia, ***3.Qasim Murtaza*** and ***3.Ranganath Muttanna Singari***, Mechanical, DTU

29. Modeling and Predicting Piped Water Theft using Machine Learning Approach, ***7.Simon Peter Khabusi*** and ***3.Rajni Jindal***, Computer Science, DTU

30. Movie trailer classification using deer hunting optimization based deep convolutional neural network in video sequences, ***3.Prashant Giridhar Shambharkar***, and M. N. Doja, Computer Science, DTU

31. Neural Machine Translation for Low-Resourced Indian Languages, ***8.Himanshu Choudhary***, ***8.Shivansh Rao*** and ***3.Rajesh Rohilla***, Electronics, DTU

32. Nutritional Profile Estimation in Cooking Recipes, ***8.Jushaan Kalra***, Devansh Batra, Nirav Diwan and Ganesh Bagler, Computer, DTU

33. Observational constraints on holographic dark energymodel with matter creation, ***6.Ajay Kumar*** and ***3.C.P. Singh***, Mathematics, DTU

34. Optimization of machining parameters during cryogenic turning of AISI D3 steel, ***6.ANURAG SHARMA***, ***3.R C SINGH*** and ***3.RANGANATH M SINGARI***, Mechanical, DTU

35. Real Time Experimental control and Design of FOP□□ and L-PID Controllers tuned by Invasiveweed optimization for Trajectory Control of TRAS, ***7.Hem Prabha*** and ***8.Rajul Kumar***, Electrical, DTU

36. Relegated, Stigmatized and Peripheral Voices: A Study of Unsung Mythological Women Characters in Chitra Banerjee Divakaruni's The Forest of Enchantments, ***3.Poorva Gulati***, DTU

37. Sentiment Analysis using Machine Learning and Deep Learning, Yogesh Chandra and ***8.Antoreep Jana***, IT, DTU

38. Standoff Quantum Cascade Laser Coupled Acoustic Sensor for the Detection of Hazardous Chemicals, ***6.Mohit Mann***, ***3.A.S. Rao*** and Ramesh C Sharma, Applied Physics, DTU

39. Theoretical analysis for transmission of Gaussian and sine time irradiance of electromagnetic beam in collisional dusty plasmas, ***6.Ruchi Sharma*** and ***3.Suresh C. Sharma***, Physics, DTU

40. Thermal Reliability of GaN-BTG-MOSFET for High-Performance Applications in Integrated Circuits, ***7.Harshit Soni***[1], ***8.Pranav M Tripathi***[2], ***3.M.M. Tripathi***[1], Ajay Kumar and ***3.Rishu Chaujar***[2], [1]Electrical and [2]Physics, DTU

41. Two Quadrant Analog Voltage Divider and Square-Root Circuits Using OTA and MOSFETs, ***6.Ajishek Raj***, ***3.D. R. Bhaskar*** and ***3.Pragati Kumar***, Electrical, DTU

42. Venom-Derived Bioactive Compounds as Potential Anticancer Agents: A Review, ***7.Arpita Roy*** and ***3.Navneeta Bharadvaja***, Biotechnology, DTU

1. *Chancellor*

2. *Pro Vice Chancellor*

2.1. *Ex Pro Vice Chancellor*

3. *Faculty*

3.1. *Ex Faculty*

4. *Teaching-cum-Research Fellow*

4.1. *Asst. Librarian*

5. *Alumni*

6. *Research Scholar*

6.1. *Ex Research Scholar*

7. *PG Scholar*

7.1. *Ex PG Scholar*

8. *Undergraduate Student*

8.1. *Ex Undergraduate Student*

Multimedia Systems
https://doi.org/10.1007/s00530-020-00656-7

REGULAR PAPER

A deep learning architecture of RA-DLNet for visual sentiment analysis

Ashima Yadav[1] · **Dinesh Kumar Vishwakarma**[1]

Received: 6 November 2019 / Accepted: 8 May 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
Visual media has become one of the most potent means of conveying opinions or sentiments on the web. Millions of photos are being uploaded by the people on famous social networking sites for expressing themselves. The area of visual sentiment analysis is abstract in nature due to the high level of biasing in the human recognition process. This work proposes a residual attention-based deep learning network (RA-DLNet), which examines the problem of visual sentiment analysis. We aim to learn the spatial hierarchies of image features using CNN. Since the local regions in the image convey significant sentiments, we apply residual attention model, which focuses on crucial sentiment-rich, local regions in the image. The significant contribution of this work also includes an exhaustive analysis of seven popular CNN-based architectures such as VGG-16, VGG-19, Inception-Resnet-V2, Inception-V3, ResNet-50, Xception, and NASNet. The impact of fine-tuning on these CNN variants is demonstrated in visual sentiment analysis domain. The extensive experiments on eight popular benchmark data sets are conducted and the performance is measured in terms of accuracy. The comparison of accuracy with similar state-of-the-art exhibits the superiority of the proposed work.

Keywords Convolutional neural network · Deep learning · Residual attention model · Visual sentiment analysis

1 Introduction

The amount of multimedia data on social networks is tremendously increasing as social networks have become a significant part of everyone's life. The users are expressing their opinions virtually on every topic or event in the form of text, images, emoticons, audios, and videos. This has produced a strong vision for inspecting the opinions of users from different modalities. Moreover, the sentiment expressed by visual modalities like emoticons, images, and videos can boost the sentiment extracted by the text modality [1]. Nowadays, visual media are rising as users feel more convenient to express their opinion in the form of emoticons or images. For example, Fig. 1 displays the image tweets which are collected from Twitter. The first set of images corresponds to the positive sentiment, and the second set of images display the negative sentiment.

With the growing demand for deep learning models, CNNs are extensively used in the area of computer vision and have shown breakthrough in various tasks like action recognition [1], person re-identification [2], human pose estimation [3], object detection and localization [4, 5], and object segmentation [6]. Object recognition involves identifying the objects in a given image, whereas visual sentiment analysis (VSA) refers to the task of recognizing the objects, scenes along with their emotional context [7]. The problem of universal object recognition is well defined, whereas VSA is abstract in nature due to the high level of subjectivity in the human recognition process [8]. For object recognition, the images belonging to the same class often share similar visual features. However, in VSA, different sentiments can be expressed from the objects of the same class. For example, images of "angry cat" and "happy cat" belongs to the same class "cat", but expresses different visual sentiments. Hence, the VSA methods need to learn these subtle differences in the images of the same class. Another challenge can be seen in [9], where different objects convey the same sentiment (either positive or negative), resulting in high intra-class variance. The general object recognition methods have

Communicated by Y. Zhang.

✉ Dinesh Kumar Vishwakarma
dvishwakarma@gmail.com

Ashima Yadav
ashimayadavdtu@gmail.com

[1] Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Delhi, India

Published online: 25 May 2020

Fig. 1 Visual sentiments

limited performance. Moreover, several pre-trained CNN models like Alexnet, VGG16, and Resnet have been trained on the ImageNet dataset (containing around 1.2 million images) for object recognition and detection task. For VSA, no such models exist which are pre-trained for extracting the visual sentiments, and developing such a model requires massive cost and time. Hence, transfer learning is applied to solve this problem by transferring the knowledge from the source to the target domain [10]. The representation of images learned from CNNs on big data sets are passed to other image recognition tasks, where little data are available. Motivated by this approach, researchers are now using transfer learning for VSA.

1.1 Research objectives

The main goal of this work is to develop a unified framework that helps in identifying the positive and negative sentiments along with eight different types of emotions namely *Amusement, Anger, Awe, Contentment, Disgust, Excitement, Fear, and Sadness* in the images. Inspired by the results of applying transfer learning in computer vision, the features learned by training deep CNNs on large data sets like ImageNet [11] are applied in the area of VSA. This eradicates the need for training a model from scratch. Chu et al. [12] have discussed the best approaches for fine-tuning the visual classifiers on the target domain. They concluded that the best practice for fine-tuning a network on a new dataset (which has less amount of data available) was to copy all leaving the last layer of the network. Hence, inspired by this approach, we have applied transfer learning for extracting the sentiments and emotions from images by proposing an efficient RA-DLNet architecture for VSA. The proposed architecture is a unique integration of CNN for learning the spatial hierarchies of image features and residual attention, which generates attention-aware features that change adaptively as layers go deeper and deeper. Our CNN architecture is a combination of feedforward series of stacked inception-like modules called cells (Normal cell and Reduction cell). The prime advantage of these cells is that they can handle the images of arbitrary spatial dimensions and filter depth. Since the local regions in the image convey significant sentiments, we apply residual attention, which focuses on crucial sentiment-rich, local regions of the image. This attention block enhances the vital features by suppressing the noise without significant information loss. Hence, the model can capture different types of attention extensively which refines the feature maps gradually, thus guiding the feature learning process. The significant contributions are summarized as follows:

- We propose an effective residual attention-based deep learning network (RA-DLNet) using CNN and Residual attention model for classifying the visual sentiments.
- We perform extensive analysis on seven popular CNN-based architectures: VGG-16, VGG-19, Inception-Resnet-V2, Inception-V3, ResNet-50, Xception, and NASNet to demonstrate the impact of fine-tuning on VSA problem.
- The performance of the proposed RA-DLNet architecture is evaluated by performing extensive experiments on eight publically available benchmark datasets: Twitter I (3-agree, 4-agree, 5-agree), Twitter II, CMU-MOSI, CMU-MOSEI, Flickr, and Instagram for binary sentiment classification, and on Artphoto, F&I dataset for multi-class emotion classification.
- The exhaustive experiments validate that the proposed RA-DLNet has surpassed the state-of-the-art methods on all benchmark datasets.
- Finally, to quantify the effect of some vital layers of the RA-DLNet architecture, we conduct layer ablation study and provide visualizations in the form of colormaps and heat maps, which establish the effectiveness of the proposed architecture.

The remaining manuscript is structured as follows: Sect. 2 outlines the previous work done in the area of VSA. Section 3 presents the proposed architecture. The experimental settings and evaluation results are demonstrated in Sect. 4. Finally, Sect. 5 discusses the conclusion followed by future work in this area.

2 Related work

Research in the area of VSA is taking place for several years. This section explores the significant work and discusses the importance of visual attention for VSA.

2.1 Visual sentiment analysis

The area of VSA is quite challenging. The work done in this area can be divided into low-level features, mid-level features, and deep learning-based features, as shown in Fig. 2. This section gives a detailed analysis of the crucial work done in each of the categories.

Earlier efforts for VSA focused on several low-level features. Siersdorfer et al. [13] used SIFT-based visual terms with RGB Histogram to get the prominent features from Flickr images. The text metadata embedded in the image expressed the sentiment conveyed by them. A sentiment value was allotted to the image on the basis of textual metadata which was combined with low-level features for the final classification with SVM. Vonikakis et al. [14] proposed a method to generate slideshow of the family photos by concentrating on the colors of the picture, Gist of the scene, emotions, and the clicked time. Jia et al. [15] used Flickr

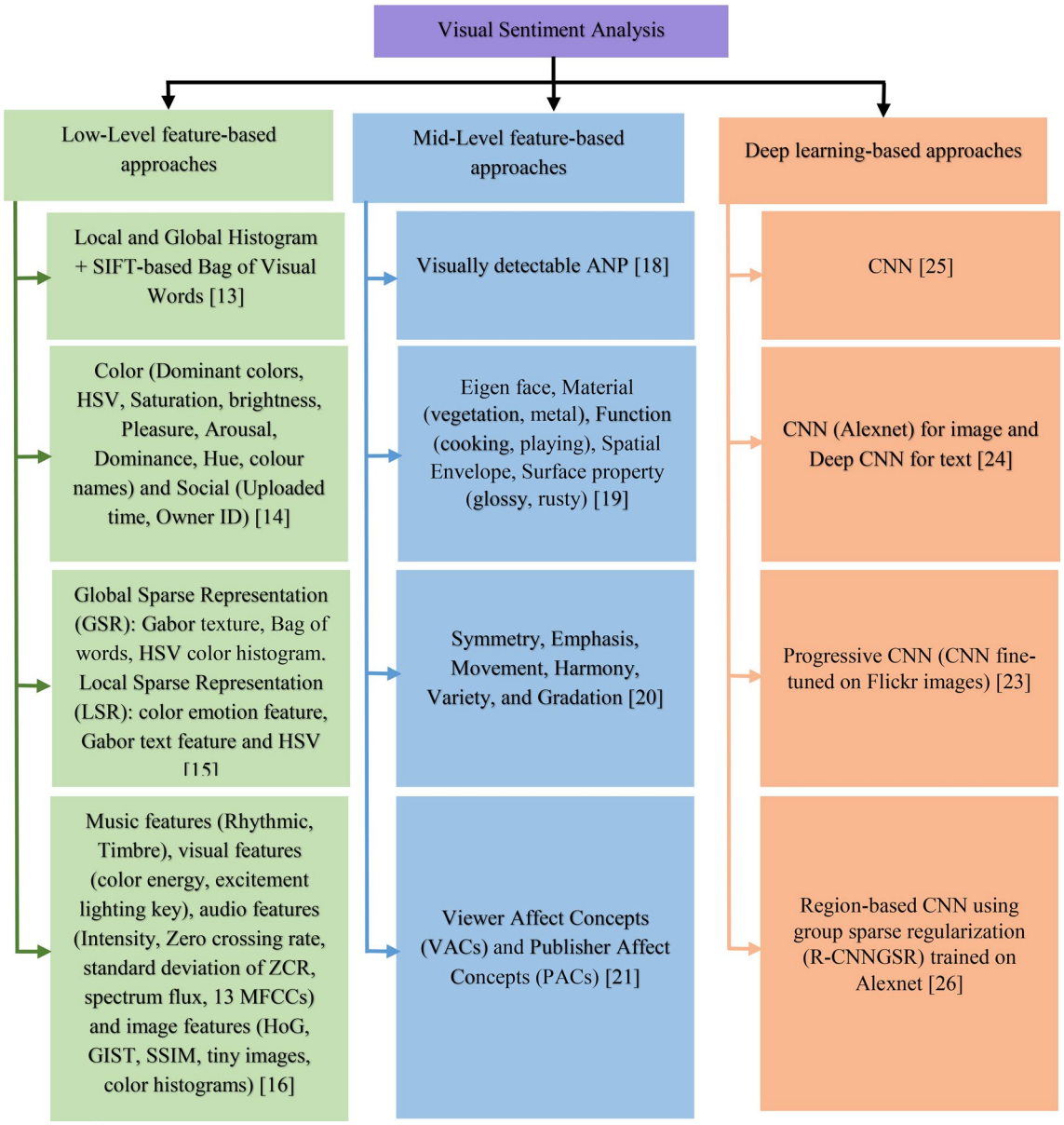


Fig. 2 Taxonomy of visual sentiment analysis

images to analyze the features which reflect the effects of the author in the images. They used image color features, which include HSV, arousal and dominance, saturation and brightness, color names, and used social correlation among images. They applied the partially-labeled factor graph model (PFG) for predicting the affects in them. Furthermore, they performed the Feature Contribution Analysis to analyze how each of the features helped to infer the affect from the images. Experimental results showed that HSV features gave a better precision, while saturation and brightness gave better F1 measure and Recall.

Li et al. [16] proposed Bilayer Sparse Representation (BSR)-based system, which contains global sparse representation (GSR) and local sparse representation (LSR). The former extracts color histogram, texture, bag of words, and the latter performs image segmentation to generate uniform image regions that apply GSR features. Since an image may represent more than one emotion, [17] presented an approach for tagging multiple emotions in the multimedia data. First, they extracted multimedia features from the data which includes: music features (rhythmic features, timbre features), visual features (color energy, visual excitement, lighting key), audio features (Spectrum flux, Average energy intensity, etc.), and image features (HoG features, GIST feature, SSIM features, etc.). Next, they measured emotion tags using binary relevance. Finally, a three-layer RBM (TRBM) model was used to capture the high-order dependencies in the labels and the measurements obtained in the former step. The results prove that the method can extract the various emotions present in the images.

The major drawback of low-level features was that they were unable to extract the sentiments from images. Hence, mid-level feature-based representation became popular which gave a mid-level description of the images. SentiBank [18] provided adjective noun pairs (ANP) for every image which was used by the SVM and logistic regression classifiers. Sentribute [19] generated 102 mid-level attributes by extracting the low-level features from images. Zhao et al. [20] extracted principles-of-art emotion features (PAEF) like variety, gradation, movement, etc., from the International Affective Picture System to calculate the final sentiment score. Chen et al. [21] presented a novel approach to convey the affects intended by a publisher from the visuals of images. They applied Sentibank to extract the 1200 Publisher Affect Concepts (PACs) from images and various data mining methods for discovering 446 Viewer Affect Concepts (VACs) from the comments. All the experiments were conducted on Flickr.

The deep learning-based architectures are increasingly being applied in VSA [22, 23]. You et al. [24] fine-tuned CNN on Flickr images. Chen et al. [25] used multiple modalities in the form of text, image, and emoticon from microblogs, and proposed a weak supervised deep learning model called WS-MDL to focus on numerous modalities. Yang et al. [26] employed CNN for learning the sentiments by extracting the global features along with local features from the image. The objectness score was calculated by creating a bounding box around every image. The objectness score along with the sentiment score was joined to generate the affective regions. Xiong et al. [27] designed a framework based on the pre-trained VGGNet model and group sparse regularization (R-CNNGSR) to obtain an initial sentiment prediction. The sentiment regions were detected by combining the sentimental features with the full image to estimate the final sentiment score of the image. This approach can also be applied in the future for multi-label sentiment classification.

2.2 Visual attention

Sometimes, images may be noisy with lots of immaterial information that could hinder the performance of the model. Attention-based systems were widely used for these problems. Li et al. [28] applied a 3D attention model to extract the spatial and temporal features of a video and get the total highlight score for video highlight detection. Zhao et al. [29] exploited the LSTM-based attention network for fine-grained classification. The proposed recurrent soft attention framework focused on crucial regions of the image.

Recent studies on sentiment analysis apply attention mechanisms for affective analysis. She et al. [8] proposed a weakly supervised coupled network to detect sentiments using the detection branch and classify them using the classification branch. The former generated the regions which evoke certain emotions. The soft attention map gives the probability of emotion for each part of the image. The final sentiment maps highlight the important part of the image which serves as crucial information for the sentiment classification process. This information was combined with deep features to fetch the semantic features of the image. Experimental results on several data sets show the efficacy of soft attention systems for classification. However, their approach gave a low precision value in the case of images with smaller objects.

Fan et al. [30] developed a novel dataset (EMOd) for capturing the relation between the emotional content and the visual attention mechanism of human by including eye-tracking data and annotating the image context related to scenes, objects, emotions, and semantics. Furthermore, they developed a CNN model that compute a 1024-dimensional feature weights of each image. This helped to select the relevant semantic features of each image. The experimental results demonstrate that emotional content in the images strongly attracts human visual attention. Similarly, [31] applied the VGG-19 model along with saliency maps to represent the visual attention in the images. The results

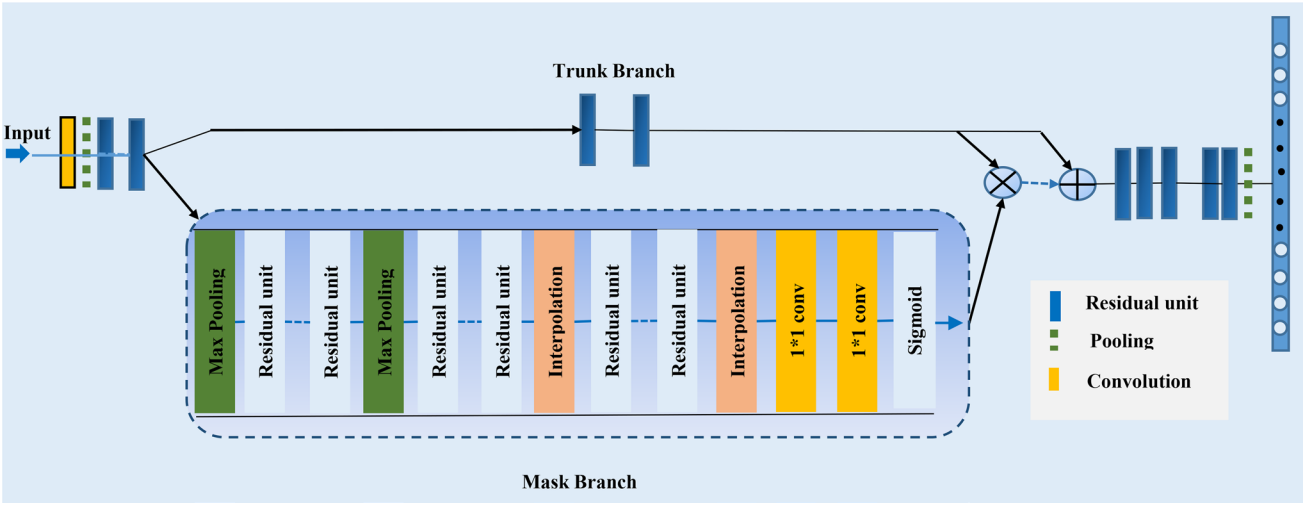


Fig. 3 Residual attention network [36]

show that the negative sentiments were influenced by focal information, whereas the positive sentiments by focal and context-based information. Lee et al. [32] proposed two subnetworks that extracted facial features and contextual features separately. The features from both the networks were fused to build a context-based emotion recognition system. Similarly, Bawa et al. [33] developed a technique based on two models. The first model extracted the global features, and the second model learned the emotional state of different subjects in an image. The features from both the model were learned using the LSTM model.

3 Proposed methodology

Visual sentiment classification can be modeled as an image classification problem. CNN architectures have shown effective results for such problems. The main drawback of CNN is that they need a huge dataset for their training. The datasets used in VSA are not large enough to properly train the CNN architectures. Hence, to resolve this problem, we use a transfer learning-based approach, which transfers the information learned by a pre-trained network on large data sets like ImageNet [11]. The pre-trained models are fine-tuned according to the target dataset. This process involves initializing the network weights according to a pre-trained model. Finally, the final layer of the model is swapped with a new one, containing the same neurons as the number of labels (or classes) in the target dataset. Hence, the fine-tuning process is prevalent for image classification tasks.

Inspired by these ideas, we have used fine-tuning for our VSA task. Our work is based on the popular architecture of NASNet [34]. The proposed attention mechanism is described in Sect. 3.1, and the overall architecture is discussed in Sect. 3.2.

3.1 Proposed attention mechanism

Earlier approaches were based on the encoder-decoder system for language modelling, image captioning, and machine translation. The encoder-decoder units were composed of recurrent neural networks or bi-directional recurrent neural networks (LSTM or GRU). The encoder module processes an input word sequence to compress it into a constant length context vector, known as sentence embedding. The decoder was initialized with the context vector to generate the required output. The disadvantage of this model was that it was incapable of remembering the long sentences. Hence, attention mechanisms [35] were developed to resolve these problems. In this mechanism, the decoder selectively focuses on information-rich phrases of the source sequence. This was achieved by modifying the context vector to focus on the encoder's hidden state, decoder's hidden state, and alignment between source and target. [35] proposed an encoder–decoder system with an additive attention mechanism. The encoder accepts a series of vectors $x = (x_1, x_2, \ldots .., x_T)$ which maps to the series of encoder's hidden states $h_1, h_2, \ldots, h_T$. Each hidden layer contains the information about the full input sequence focusing on those parts which surround the ith word of the input sequence.

This work applies the Attention mechanism, which is inspired by the Residual Attention network [36]. In [36], the author has developed a system that was built by assembling multiple attention modules in a feedforward structure, which is different from the previous work of formulating attention as a sequential process. The proposed architecture

applies a single attention module to concentrate on the most significant regions of the image.

Figure 3 shows the complete view of the attention module. Each module is separated into two branches. The first branch, known as the *trunk branch,* helps in the feature engineering process. The second branch, called a *mask branch,* helps in picking the relevant features of the image to focus and learn the attention-based features. They prevent the trunk branch parameters from getting updated with the wrong gradients computed from noisy labels. The mask branch consists of a sequence of residual units in-between the pooling layers. Bi-linear interpolation is applied for upsampling the output. Finally, the sigmoid layer is added after two convolutional layers to normalize the output in [0, 1]. The output of the Attention module is calculated using Eq. (1):

$$H_{i,c}(X) = A_{i,c}(X) * T_{i,c}(X). \tag{1}$$

Here, $A_{i,c}(X)$ denotes the attention mask, $c \in \{1,2,\ldots,C\}$ is the channel's index, i varies over the spatial positions, $T_{i,c}(X)$ represents the original features generated by the trunk branch, and * denotes element-wise product (dot product) between trunk branch features and mask branch features. The gradient of the mask for input features is calculated using Eq. (2):

$$\frac{\partial A(X,\theta)T(X,\emptyset)}{\partial \emptyset} = A(X,\theta)\frac{\partial T(X,\emptyset)}{\partial \emptyset}, \tag{2}$$

where θ denotes the parameters of the mask branch and $\varnothing$ denotes the parameters of the trunk branch. However, the repeated dot product between trunk branch features and mask branch features leads to a significant performance drop for useful and useless information. Hence, Eq. (1) is changed as follows:

$$H_{i,c}(x) = \big(1 + A_{i,c}(X)\big) * T_{i,c}(X) \tag{3}$$

or

$$H_{i,c}(x) = T_{i,c}(X) + \big(A_{i,c}(X) * T_{i,c}(X)\big). \tag{4}$$

From Eq. (4), it can be seen that an element-wise product of trunk branch features and mask branch features is performed followed by element-wise addition with trunk branch features. Thus, when $A_{i,c}(x) = 0$, the $H_{i,c}(x)$ will approximate to the original features $T_{i,c}(x)$. This process is called residual attention learning where the soft mask unit is constructed as the same mapping. Hence, the mask branch tries to improve the trunk branch features by refining the feature maps, which makes the network robust to noisy inputs.

3.2 RA-DLNet architecture

The architecture of the proposed RA-DLNet framework for VSA is shown in Fig. 4. The proposed architecture is divided

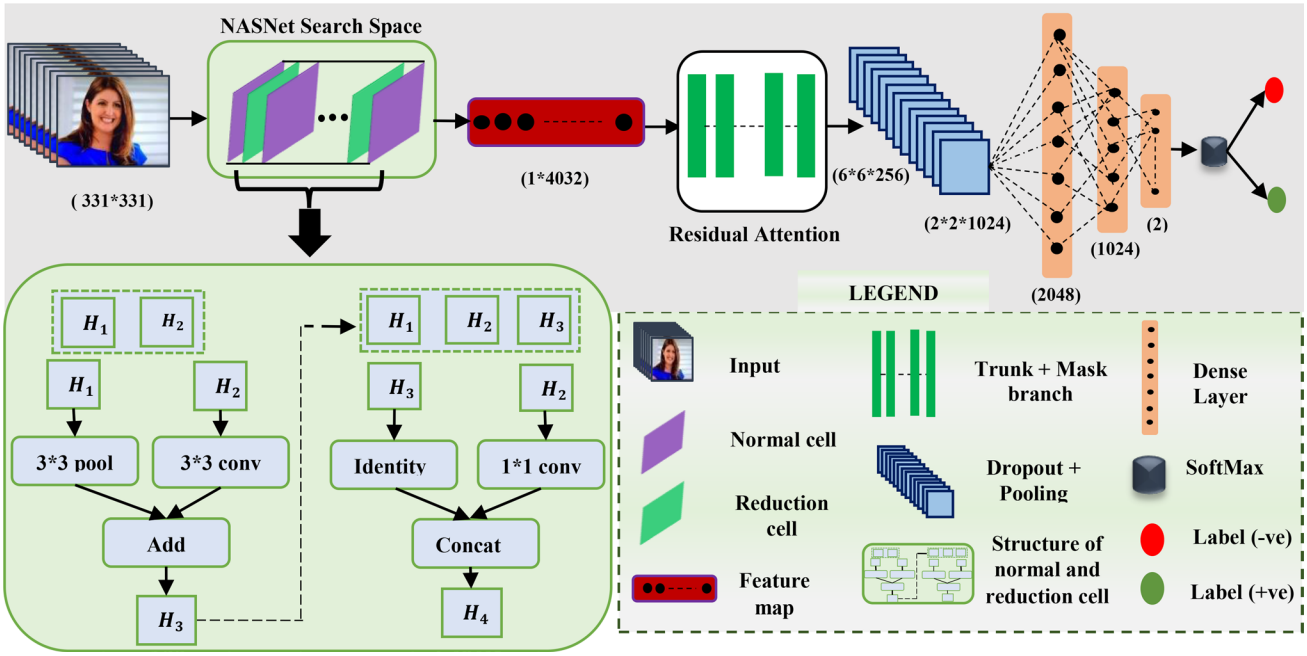


Fig. 4 Proposed RA-DLNet architecture

into three main components: The first part consists of applying a pre-trained model. We have implemented the NASNet architecture, which is inspired by neural architecture search (NAS) framework [37]. In the NAS framework, a controller RNN recursively trains a child network on a validation set for finding the good architectures. The child network will generate different accuracies, which helps to calculate the gradient for updating the controller. Hence, the controller will eventually improve its searching process for learning the most appropriate architectures. The controller RNN used in our work is the LSTM model with 100 hidden units. The Proximal Policy optimization [38] is used for updating the parameters of the controller RNN. The objective function of PPO is defined as in Eq. (5):

$$P^{CPI}(\theta) = \hat{E}_t\left[\min\left(\kappa_t(\theta)\hat{A}_t, \text{clip}\left(\kappa_t(\theta), 1-\in, 1+\in\right)\hat{A}_t\right)\right]. \tag{5}$$

The *CPI* superscript denotes the conservative policy iteration constraint which controls the excessive policy updates. The $\hat{A}_t$ denotes the advantage function that estimates the relative value of the selected action in the current state. The $\kappa_t(\theta)$ defines the ratio between a newly updated policy output and the old policy output of network, which is shown in Eq. (6):

$$\kappa_t(\theta) = \frac{\Pi_\theta(a_t|s_t)}{\Pi_{old}(a_t|s_t)}. \tag{6}$$

Here, $\Pi_\theta\left(a_t|s_t\right)$ signifies the policy that takes observed states from the environment as input and suggests actions to take as output. Hence, in Eq. (5), the term $\kappa_t(\theta)\hat{A}_t$ pushes the policy towards actions that yield a high positive advantage over the baseline. The second term contains the truncated version of $\kappa_t(\theta)$ by applying a clipping operation between $1-\in$ and $1+\in$.

The architecture of NASNet is composed of different convolutional cells, called normal cell and reduction cell. As discussed above, the controller RNN continuously searches for the best architectures of these cells. Each of the cells is made up of B (in our case, $B=5$) blocks where each block performs the operations which are described in Algorithm 1 and further relate to the different outputs of the softmax layer. Since this architecture has attained state-of-the-art on prevalent ImageNet classification task, hence, we are motivated to analyze its performance for capturing the visual sentiments from the images. As shown in Fig. 4, the input image of size 331×331 passes through a series of reduction cells and normal cells. The former returns a feature map whose height and width are reduced by two and the latter returns a feature map with the same dimension. Both the cell shared the same structure with different weights, as shown in Fig. 4.

Algorithm 1 Working of controller RNN for generating the structure of convolutional cells in NASNet

For i = 1 to 5 **do**

1. Select two hidden states H_1, H_2 from the set of previously created hidden states.
2. Select an operation (identity, convolution, and pooling) to be applied on both the hidden states to generate a new output state H_{o1} *and* H_{o2}.
3. Combine H_{o1} *and* H_{o2} by using element-wise addition or concatenation to create a new hidden state H_3.
4. Append the state H_3 to the existing list of hidden states for making it as potential input for the next blocks.

end For

return the concatenated value of the final cell output

We extract the final feature maps of size 1×4032 and pass it to the attention mechanism module, which was discussed in Sect. 3.1. As discussed, the attention mechanism concentrates on the different regions of the input image to get a finer sentiment prediction. The final feature vector is passed into a dropout layer with 0.5 probability, followed by an average pooling layer. We have used three dense layers for softer dimensionality reduction. The first layer contains 2048 neurons and the second layer comprises 1024 neurons. The final dense layer is composed of two neurons in the case of binary sentiment classification and eight neurons in the case of emotion detection with a softmax classifier for the final classification. Algorithm 2 describes the training of RA-DLNet architecture.

Algorithm 2 Training of RA-DLNet architecture for VSA

Input: Training model (M), X = $\{x_1, x_2, \ldots, x_n\}$ are the set of training images of size 331*331, Y = $\{y_1, y_2, \ldots, y_n\}$ are the set of sentiment labels of X.

1. Partition X and Y randomly into five groups: $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_5, Y_5)\}$.
2. **For** i = 1 to 5 **do**
3. Fine-tune the model M using NASNet Architecture (Algorithm 1) to obtain the feature map F_i
4. Calculate the attention over the obtained feature map F_i of M
5. Apply dropout with 0.5 probability and global average pooling layer.
6. Add fully connected layers to M and apply the softmax classifier to calculate the sentiment probabilities.
7. Calculate the performance of M on the set of testing images.
8. **end for**
9. **return** the average performance of M on the set of testing images.

Table 1 Twitter dataset

Classes	5-Agree	At least 4-agree	At least 3-agree	Twitter II
# of positive samples	581	689	769	470
# of negative samples	301	427	500	133
Total	882	1116	1269	603

Fig. 5 Sample images from **a** Twitter dataset and **b** Artphoto dataset

Table 2 ArtPhoto dataset details

Emotion	Amusement	Anger	Awe	Contentment	Disgust	Excitement	Fear	Sadness	Total
# of Samples	101	77	102	70	70	105	115	166	806

4 Experimental results

To assess the performance of the proposed architecture, we experimented on widely used benchmark datasets for sentiment and emotion classification. We have separated the data sets into two categories: small-scale data sets and large-scale data sets according to the number of image samples. The evaluation parameters such as Precision, Recall, F1 score, and Accuracy are measured on these datasets. The upcoming sub-sections give the complete details of these datasets.

4.1 Small-scale data sets

The images on Twitter are related to the current trending topics. The users are most likely to upload the images which are related to their personal experiences. Hence, we experimented on the four most popular Twitter data sets for binary sentiment prediction: Twitter I [24] image dataset, which itself is composed of three sub-data sets, namely 3-agree, 4-agree, 5-agree dataset, and Twitter II [18] image dataset. Amazon Mechanical Turk (AMT) annotation was used for creating the sentiment label for the images. For Twitter I dataset, five AMT annotators labeled the images of the dataset. Thus, "5-agree" indicates that all the annotators gave identical sentiment response for a given image. Twitter II dataset is composed of images from 21 different hashtags. Three different AMT annotators were asked to label the images. Finally, the tweets with unanimous labels were included in the released dataset. Table 1 gives the complete details about the data set.

The major challenge with these datasets is that the images are noisy. Some of the images uploaded by the users are not captured clearly. Moreover, some of the images are quite challenging for the classifier as they contain multiple images within a single image or contains either text only or a combination of text + image both. Figure 5a shows some sample images from the Twitter dataset.

For multi-class emotion detection, we have work with ArtPhoto [39] dataset which consists of 806 artistic photos that are collected from some of the art sharing sites and categorized into eight discrete emotions namely, amusement, anger, awe, contentment, disgust, excitement, fear, and sadness. The complete details of the dataset are given in Table 2.

Some of the sample images from this dataset is shown in Fig. 5b. The photos evoke a certain emotion by combining and manipulating specific visual characteristics like colors, image position, etc.

4.2 Large-scale data sets

We assess the performance of our model on several large-scale data sets which include CMU-MOSI [40], CMU-MOSEI [41], Flickr [42], Instagram [42] for binary sentiment prediction, and F&I [43] for multi-label emotion classification. CMU-MOSI consists of 2199 opinionated video clips, and CMU-MOSEI contains 17,859 clips. Both the data sets were annotated with sentiment classes in the range of $[-3, 3]$. The distribution of samples for binary sentiment classification is as same as in [44], in which for CMU-MOSI, frames $[-3, 0)$ were taken as negative and frames $[0, 3]$ as positive and for CMU-MOSEI, frames with $[-3, 0)$ were taken as negative, and frames $(0, 3]$ as positive. Since both these data sets have primarily been used for multimodal sentiment classification and the purpose of our work is to focus on visual sentiments, hence, we have only focused on extracting the sentiments from videos (ignoring the audio and textual modality). The raw features were obtained from CMU Multimodal SDK [45]. The Flickr and Instagram sentiment data sets contain 60,745 (48,139 positive and 12,606 negative) and 42,856 (33,076 positive and 9780 negative) images, respectively, as some of the images were not available, because they were removed by the users. The Flickr & Instagram (F&I) emotion dataset contains 23,308 images annotated with eight discrete emotions namely, amusement (4942), anger (1266), Awe (3151), contentment (5374), disgust (1658), excitement (2963), Fear (1032), and sadness (2922). These are by far the largest datasets in this domain.

Table 3 Classification results (%) of RA-DLNet on benchmark datasets

Dataset	P	R	F1	ACC
3-Agree	88.2	78.9	83.2	81.3
4-Agree	86.7	85.1	85.8	83.2
5-Agree	**92.5**	95.7	**93.8**	**89.1**
Twitter II	80.7	**99.5**	89.1	81.2
ART Photo	76.3	86.5	80.9	71.8
CMU-MOSI	64.7	73.3	68.7	67.7
CMU-MOSEI	82.1	70.5	75.9	77.2
Flickr	88.6	90.6	89.6	83.4
Instagram	84.7	96.6	90.2	82.7
F&I	74.7	72.8	73.7	72.8

4.3 Implementation details

This section describes the evaluation settings, classification results, baseline comparison, layer ablation study, visualizations, and an empirical study of NASNet with other CNN architectures.

4.3.1 Evaluation settings

All the experiments were performed on 64-bit Windows 10 machine with 128 GB RAM on NVIDIA Titan RTX GPUs. Keras framework was used to build the network using Python 3. Data augmentation techniques like zooming, shearing, horizontal flipping, and rescaling were applied to avoid overfitting. We have used Adam as an optimizer with an initial learning rate of 0.001. Xavier uniform initializer is used to set the weights of the new layers and the weights of the pre-trained model were used to train the remaining layers.

4.3.2 Classification results

We have employed precision (P), recall (R), F1 score (F1), and accuracy (ACC) metrics to measure the performance of the RA-DLNet on all the data sets. Fivefold cross-validation is used to represent each class equally across the individual test fold and assess the model's performance. The final results were computed by averaging the values of each fold.

The classification results are shown in Table 3 and the highest value is highlighted in bold text.

- For the Twitter I and Twitter II dataset, we have taken 80% samples for training and 20% for testing with batch size as 32. The model was trained with binary cross-entropy loss for 50 epochs.
- For Artphoto dataset, since we have an unbalanced number of images per emotion category, we followed *one vs. all* setup strategy, where we separate each category against the rest of the others in turn. The model was trained with categorical cross-entropy loss for 50 epochs.
- For the CMU-MOSI dataset, the training, validation, and testing set consists of 1284 clips, 229 clips, and 686 clips, respectively. Similarly, the training, validation, and testing set for CMU-MOSEI dataset include 12,787, 3634, and 1438 clips respectively. The model was trained with

Table 5 Performance comparison of proposed architecture on Twitter II dataset (all the results are in %)

References	Method	P	R	F1	ACC
[18]	SentiBank	–	–	–	70.0
[46]	Hybrid SentiBank + late fusion	78.1	100	87.6	78.1
[26]	VGGNet + fusion	–	–	–	80.4
[8]	WCSNet	–	–	–	81.3
RA-DLNet	CNN + residual attention	**80.7**	**99.5**	**89.1**	**81.2**

Table 4 Performance comparison of proposed architecture on Twitter I (All the results are in %)

References	Method	3-Agree				4-Agree				5-Agree			
		P	R	F1	ACC	P	R	F1	ACC	P	R	F1	ACC
[13]	GCH	67.8	83.6	74.9	66.0	68.7	84.0	75.6	66.5	70.8	88.8	78.7	68.4
	LCH	71.6	73.7	72.6	66.4	72.5	75.3	73.9	67.1	76.4	80.9	78.6	71.0
	GCH + BoW	68.3	83.5	75.1	66.5	70.3	84.9	76.9	68.5	72.4	90.4	80.4	71.0
	LCH + BoW	72.2	72.6	72.3	66.4	75.1	76.2	75.6	69.7	77.1	81.1	79.0	71.7
[18]	SentiBank	72.0	72.3	72.1	66.2	74.2	72.7	73.4	67.5	78.5	76.8	77.6	70.9
[19]	Sentribute	73.3	78.3	75.7	69.6	75.0	79.2	77.1	70.9	78.9	82.3	80.5	73.8
[1, 2, 24]	CNN	73.4	83.2	77.9	71.5	77.3	85.5	81.1	75.5	79.5	90.5	84.6	78.3
[24]	PCNN	75.5	80.5	77.8	72.3	78.6	84.2	81.1	75.9	79.7	88.1	83.6	77.3
[48]	Fine-tuned CaffeNet + oversampling	–	–	–	–	–	–	–	–	–	–	–	83.0
[9]	DCAN + ReKL	–	–	–	–	–	–	–	–	–	–	–	83.8
[47]	Fine-tuned CaffeNet + oversampling	–	–	–	74.9	–	–	–	78.7	–	–	–	83.0
[46]	Hybrid SentiBank + Late fusion	73.9	80.9	77.2	71.1	76.5	82.3	79.2	73.4	80.4	86.4	83.3	77.2
[31]	VGG19 + saliency map	–	–	–	–	–	–	–	–	91.0	89.0	90.0	87.0
[49]	SentiNet-A	82.0	80.0	81.0	77.7	85.0	83.0	84.0	80.7	89.0	87.0	88.0	85.1
[7]	GoogleNet	85.1	78.4	81.6	78.7	84.9	83.0	83.9	80.7	92.6	86.9	89.7	86.1
[50]	CNN	–	–	–	67.77	–	–	–	–	–	–	–	–
[26]	VGGNet + Fusion	–	–	–	81.0	–	–	–	85.1	–	–	–	88.6
RA-DLNet	CNN + Residual Attention	**88.2**	**78.9**	**83.2**	**81.3**	**86.7**	**85.1**	**85.8**	**83.2**	**92.5**	**95.7**	**93.8**	**89.1**

binary cross-entropy for 100 epochs with a batch size of 128.

- For Flickr and Instagram data sets, we have taken 90% as training samples and 10% as testing samples, as specified in [42]. The model was trained with binary cross-entropy loss for 100 epochs with a batch size of 128.
- For F&I data sets, we have used 80% as training samples, 5% as validation, and 15% as testing samples, as specified in [43]. The model was trained with categorical cross-entropy loss for 100 epochs with a batch size of 128.

From Table 3, we observe that in the Twitter domain, 5-agree dataset gives the best accuracy rate of 89.1% and an F1 score of 93.8%, whereas the Twitter II dataset gives an accuracy of 81.2% with an F1 score of 89.1%. The intuition behind this behavior is that the Twitter II dataset contains more noisier images (blur images) and challenging images in the form of multiple images within a single image or images with either text or a combination of both text + image, as discussed in Sect. 4.1.1. However, when we compare our results with the earlier state-of-the-art methods in the upcoming Sect. 4.2.3, we observe that our proposed architecture performs better than state-of-the-art methods. In the Artphoto dataset, the proposed architecture obtained an accuracy of 71.8% for multi-class emotion classification.

For large-scale data sets, CMU-MOSI and CMU-MOSEI have shown an accuracy of 67.7% and 77.2% respectively, whereas in Flickr and Instagram sentiment data sets, the accuracy obtained is 83.4% and 82.7%, respectively. Hence, the proposed model has effectively characterized the facial expressions in the images. Although, these data sets were cleaner as compared to Twitter I and Twitter II data sets, yet the model misclassified some cases where text or letters were embedded in the images. For F&I emotion dataset, the model has achieved an accuracy of 72.8%. Due to more number of samples corresponding to each emotion, the network could learn more and hence improves the classification performance as compared to the Artphoto dataset. Deeper insights on the performance of each emotion class are shown in the next section.

4.3.3 Baseline comparison

This section compares the proposed RA-DLNet architecture with state-of-the-art works on all the eight datasets.

Twitter dataset The results of RA-DLNet on Twitter I and Twitter II datasets are compared with state-of-the-art works in Tables 4 and 5, respectively. Bold values represent our results. We have used the following baselines for comparison.

1. *Low-level feature-based approaches* Siersdorfer et al. [13] used Global Color Histogram (GCH), Local Color Histogram (LCH), and SIFT-based bag of visual terms. The GCH-based approach selects $4 \times 4 \times 4$ RGB histogram for representing the features, whereas the LCH-based approach first distributes each image into 16 different blocks to compute the 64-bit RGB histogram for each block. Finally, they applied SIFT descriptors to learn the visual word dictionary and used a bag of visual word features (BoW) for every image. A combination of GCH + BoW and LCH + BoW was also employed for combining the features of an image.
2. *Mid-level feature-based approaches* SentiBank [18] is a visual sentiment ontology comprising of 1,200 ANPs to generate 1200 responses for any test image. Sentibute [19] applies semantic features using scene-based attributes, instead of using ANPs. Finally, the two mid-level features were classified by SVM [46] fused the sentiment specific value with ANP values to compute the final classification score.
3. *Deep learning-based approaches* The deep learning approaches based on transfer learning for VSA are used for comparison with the proposed architecture. They are as follows:

- You et al. [24] designed CNN-based architecture is consisting of two convolutional layers with several dense layers for VSA. Progressive CNN (PCNN) was used to fine-tune the network.
- Campos et al. [47, 48] fine-tuned CaffeNet, with five convolutional layers and three dense layers for VSA. They also used oversampling to remove the dataset bias.
- Wang et al. [9] proposed DCAN (Deep Coupled Adjective and Noun) network to extract the descriptiveness and objectiveness features. They applied transfer learning with pre-trained Alexnet and Rectified Kullback–Leibler loss (ReKL) to build a scalable network.
- Song et al. [49] applied VGGNet-based saliency detection mechanism for VSA.
- Islam et al. [7] used transfer learning by applying the weights and biases from GoogleNet, which was pre-trained for VSA. Fan et al. [31] applied the VGG-19 network and devoted a full channel for saliency maps or focal object masks.

The experimental results show that the proposed architecture has outperformed the baseline results on all the evaluation metrics. The results reveal that deep learning-based approaches have outperformed the best performing low-level and mid-level approaches like Sentribute [19]. The layers of NASNet extract the local regions of the input image, whereas the attention mechanism helps to filter out the irrelevant regions and capture the global informative areas in the

Fig. 6 Performance comparison of proposed architecture on Artphoto dataset with similar state-of the-art

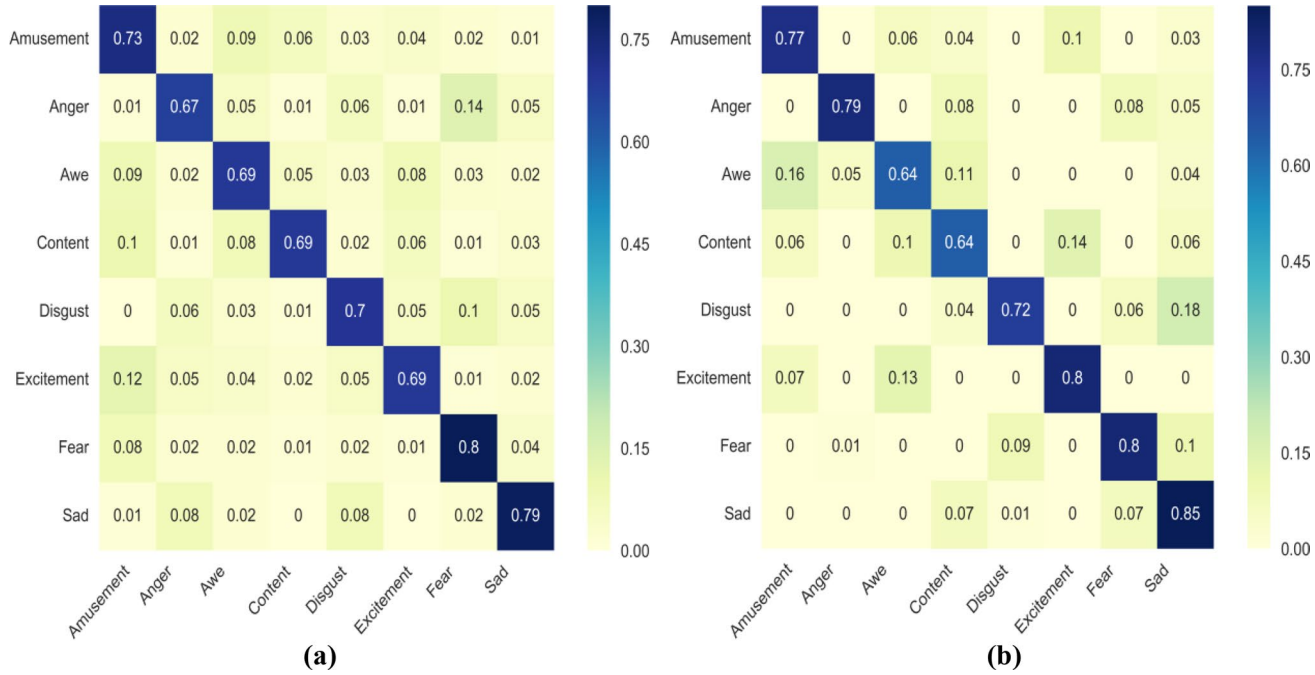


Fig. 7 Confusion matrix for **a** Artphoto dataset, **b** F&I dataset

images. Moreover, attention can also deal with noise in the dataset. Hence, the residual attention has further complimented the deep learning CNN-based architecture (NASNet) to enhance the VSA process. This shows the importance of attention mechanisms for extracting visual sentiments from the images.

Artphoto dataset Since this dataset suffers from unequal data distribution problems per emotion category, hence, we calculate the true positive rate per class averaged over positive and negative samples, as discussed in [39]. The following baselines are used for comparison:

- Machajdik et al. [39] extracted low-level features, which include color, texture, composition, and content from the images.
- Zhao et al. [20] extracted emotion features that were derived from different aspects like harmony, balance, proportion, variety, etc.

Table 6 Performance comparison of proposed architecture on CMU-MOSI dataset (all results are in %)

References	Method	ACC	F1
[55]	TFN	69.4	71.4
[56]	GME-LSTM-A	52.3	57.3
[57]	Bi-LSTM	53.8	53.5
[58]	DCC-CAE	59.4	59.3
[59]	CIA based RNN	51.31	50.48
[44]	DCCA	44.0	44.5
[60]	ICCN	48.10	49.06
[61]	HvnLBP-TOP	63.9	65.9
RA-DLNet	CNN + residual attention	**67.7**	**68.7**

Table 7 Performance comparison of proposed architecture on CMU-MOSEI dataset (all results are in %)

References	Method	ACC	F1
[57]	Bi-LSTM	59.6	59.5
[58]	DCC-CAE	67.3	67.5
[59]	CIA based RNN	75.04	69.13
[62]	GRU-based attention	76.5	69.1
[44]	DCCA	58.75	59.23
[60]	ICCN	59.25	59.90
RA-DLNet	CNN + residual attention	**77.2**	**75.9**

- Rao et al. [51, 52] used probabilistic latent semantic with the bag of visual words for generating a mid-level representation from the images.
- Wang et al. [53] extracted aesthetic features like color difference, shape, composition, etc. for affective image analysis.
- Liu et al. [54] used pre-trained deep CNN for generating semantic orientation from the images.

From Fig. 6, we see that our approach has outperformed the state-of-the-art results majorly on each of the emotion types. This improvement may be credited to the fact that the residual attention mechanism can capture the different attention-aware features from the different regions of the image. However, for *sad* emotion, the work done by Liu et al. [54] architecture has shown better results. Figure 7a shows the confusion matrix for eight classes on Artphoto dataset. As seen in figure, some of the emotions are difficult to classify such as anger and sad. This behavior is intuitive, because the images in both categories are dull with dark colors.

CMU-MOSI and MOSIE dataset As discussed in Sect. 4.1.2, we have focused on extracting the visual sentiments from the videos. Hence, we have compared our results with similar state-of-the-art works on video modality only. Tables 6 and 7 show the comparative results of the proposed RA-DLNet model on CMU-MOSI & MOSEI datasets, respectively. Our results are highlighted with bold.

Table 8 Performance comparison (ACC) of proposed architecture on Flickr and Instagram sentiment datasets (all results are in %)

References	Method	Flickr	Instagram
[18]	SentiBank	69.2	66.5
[63]	DeepSentiBank	70.1	67.1
[42]	Latent correlations	74.7	73.6
[64]	Multi-attentive pyramidal model	80.9	81.9
[65]	Weakly supervised coupled networks	81.3	81.8
[8]	WSCNet	81.3	81.8
RA-DLNet	CNN + residual attention	**83.4**	**82.7**

The following baselines are used for comparison:

- Dumpala et al. [57] extracted visual features using FACET framework and OpenFace toolkit. They applied the Bi-directional LSTM-RNN classifier for capturing the contextual features.
- Dumpala et al. [58] proposed DCC-CAE model which combines deep canonical correlation analysis along with cross-modal encoder.
- Chauhan et al. [59] proposed CIA (Context aware Interactive Attention)-based RNN framework to capture the interaction between different modalities using a contextual attention model.
- Akhtar et al. [62] developed a GRU-based framework based on context-based inter-modal attention framework.
- Sun et al. [44] developed Deep Canonical Correlation Analysis (DCCA) to learn the embeddings of data for combining audio, visual, and text modality, and improve the sentiment analysis process.
- Sun et al. [60] proposed Interaction Canonical Correlation Network (ICCN), which learns the hidden relationships between all three modalities.
- Zadeh et al. [55] developed Tensor Fusion Network (TFN), which models the interactions between different modalities by computing a threefold cartesian product from the embeddings of the modality.
- Chen et al. [56] proposed Gated Multimodal embedding LSTM with temporal attention (GME-LSTM-A), which tackles the noisy modalities and attention focus on the most crucial time steps.
- Li et al. [61] proposed the HvnLBP-TOP approach for combining handcrafted features on three orthogonal planes followed by bi-LSTM for final sentiment classification.

From the above table, it can seen that our proposed algorithms achieved significantly good results as compared with

Table 9 Performance comparison of proposed architecture on F&I emotion dataset (all results are in %)

References	Method	ACC
[18]	Sentibank	49.2
[63]	DeepSentibank	51.5
[66]	VGG_Net	67.4
[67]	CNN-RNN	73.0
[68]	Deep metric learning	67.6
[69]	CycleEmotionGAN	66.8
[8]	WSCNet	70.0
[70]	CNN	71.7
RA-DLNet	CNN + residual attention	**72.8**

Table 10 ACC (%) on all the datasets with (80:10:10) split

Dataset	# of training samples	# of validation samples	# of testing samples	Accuracy
3-Agree	1015	127	127	**82.8**
4-Agree	892	112	112	**84.14**
5-Agree	705	88	89	**91.01**
Twitter II	493	52	58	**81.6**
ArtPhoto	645	80	81	**62.4**
CMU-MOSI	1759	219	221	**73.1**
CMU-MOSEI	14,287	1785	1787	**79.7**
Flickr	48,596	6074	6075	**87.35**
Instagram	34,285	4286	4285	**85.10**
F&I	18,646	2330	2330	**80.98**

several baselines on both the datasets. We have reported only the results of video modality for the baseline comparison. The accuracy reported on CMU-MOSI and CMU-MOSIE dataset is 67.7% and 77.2%, respectively, which shows that the model has outperformed the baseline results.

Flickr and instagram sentiment datasets Table 8 compares the accuracy of our work with earlier baseline results on Flickr and Instagram sentiment datasets. The following methods are used for baseline comparison.

- Borth et al. [18] provided Adjective Noun Pairs (ANP), which were used as mid-level features for every image and fed into the SVM and logistic regression classifiers.
- Chen et al. [63] introduced Sentibank 2.0 called DeepSentiBank which used CNN-based deep networks for visual sentiment classification.
- Katsurai et al. [42] applied latent correlations among multiple views (visual, textual, and sentiment) using SentiWordNet. They obtained a latent embedding space to maximize the correlations among multiple views.
- He et al. [64] proposed Multi-Attentive Pyramidal model (MAP) which gives the association between local features of the image and the sentiment information represented by them.
- Yang et al. [65] and She et al. [8] developed weakly supervised coupled networks for detecting and classifying the sentiments using the detection branch and classification branch respectively. The final sentiment maps highlight the important part of the image which serves as crucial information for the sentiment classification process. However, the network fails to capture the sentiments in complex scenes and no interest regions are also detected in many images.

In Table 8, we see that the deep learning approaches are increasingly being applied for visual sentiment classification as they have outperformed the low-level and mid-level strategies. Bold values represent our results. However, the proposed RA-DLNet has outperformed the state-of-the-art deep learning methods for Flickr and Instagram sentiment datasets by learning more discriminative features for this task.

F&I emotion dataset Table 9 compares the accuracy of our work with earlier baseline results. We have included the following methods for baseline comparison:

- Yang et al. [66] developed multi-task CNN based on VGGNet to establish the relationship between multiple emotion categories for visual emotion prediction.
- Zhu et al. [67] proposed a unified CNN–RNN-based framework to exploit the dependence between the features.
- Yang et al. [68] explored the relationship between the emotion labels by applying CNN to optimize the classification loss and sentiment loss jointly.
- Zhao et al. [69] proposed an unsupervised approach that uses an adversarial model called CycleEmotionGan to have a similar distribution of images in the source and target domain.
- Zhang et al. [70] developed a novel method to integrate low-level visual features in the image like color, and shape with high-level information extracted by layers of CNN.

The results in Table 9 show that the proposed method has outperformed all the baseline results. Bold value represents our result. This shows that applying attention mechanism for capturing the local information has significantly improved the performance for image emotion classification. To get more insights into the classification results for each emotion, Fig. 7b shows the confusion matrix for the F&I dataset. Our approach shows a balanced performance, especially for negative emotions like anger, sad, and fear which is the most

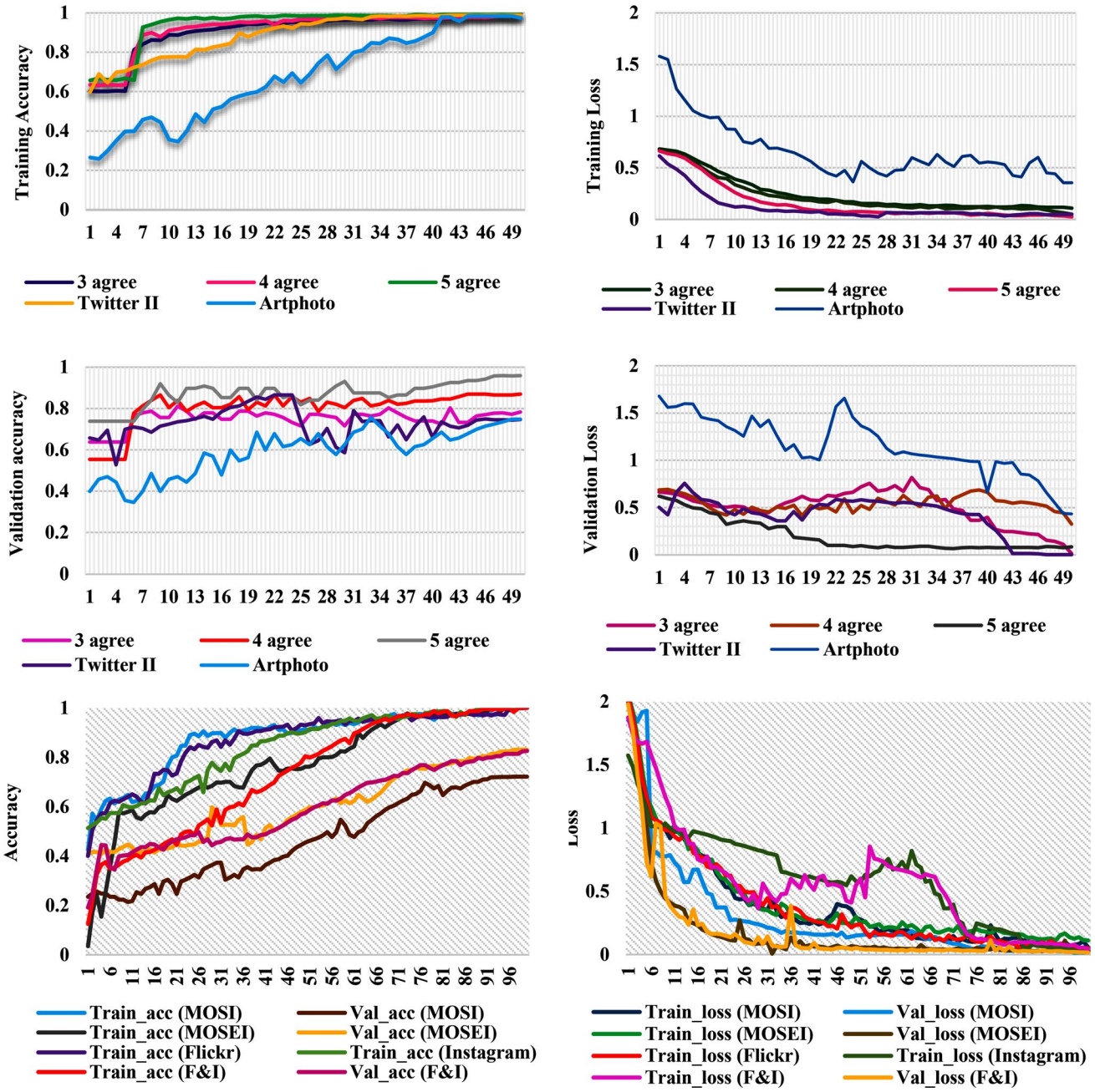


Fig. 8 Training and validation curves for each dataset

Table 11 ACC results (%) for layer ablation study on Flickr dataset

NasNet CNN	Trunk branch	Soft mask branch	Dropout	Pooling	Dense_1 (2048)	Dense_2 (1024)	Dense_3 (2)	SoftMax
✓			✓	✓	✓	✓	✓	45.1
			✓	✓	✓	✓	✓	42.2
✓	✓	✓		✓	✓	✓	✓	64.1
✓	✓	✓	✓	✓		✓	✓	71.3
✓	✓	✓	✓	✓	✓		✓	76.6
✓	✓	✓	✓	✓	✓	✓	✓	**83.4**

common problem for visual classifiers as these emotions are hard to be discriminated against by the classifiers.

4.3.4 Further evaluation results

The proposed architecture is further evaluated by training on 80% of the samples, validating on 10% samples, and finally testing on the remaining 10% samples. The final results with this split (80:10:10) are given in Table 10 and the accuracy achieved is higlighted in bold.

From Table 10, we see that with this split (80:10:10), the proposed RA-DLNet is performing even better and have crossed the results obtained by the fivefold cross-validation accuracy. However, for Artphoto dataset, the accuracy is dropped from 71.8 to 62.4%. This may be articulated by the fact that for multi-label emotion classification, the model did not get enough training images per emotion category for learning. In CMU-MOSI and CMU-MOSEI dataset, the accuracy increases by 6% and 2% respectively and for Flickr and Instagram sentiment datasets, the accuracy improves by 4% and 3% respectively. In F&I dataset, the accuracy considerably increases from 72.8 to 80.98%. The training accuracy, validation accuracy, training loss, and validation loss curves for all the datasets can be visualized by the graph plots shown in Fig. 8. As discussed, the network is trained for 50 epochs on small-scale datasets and for 100 epochs on large-scale datasets using 80:10:10 split as described above. From the accuracy plots, we infer that the training accuracy has reached almost 100% for every dataset. Moreover, as the model gets trained, the validation accuracy also increases with every epoch. The loss curves show a decreasing trend. This denotes that the model has a comparable performance on the training and validation sets and does not overfit on the data.

4.3.5 Layer ablation study and visualizations

To show the effectiveness of our proposed method, we perform an ablation study and provide some visualizations by showing the activation maximization plots and colormaps for some sample images from the datasets.

Layer ablation study In this section, we perform layer ablation study to quantify the effect of some vital layers of the RA-DLNet architecture for VSA. We re-train our model by ablating the following components on

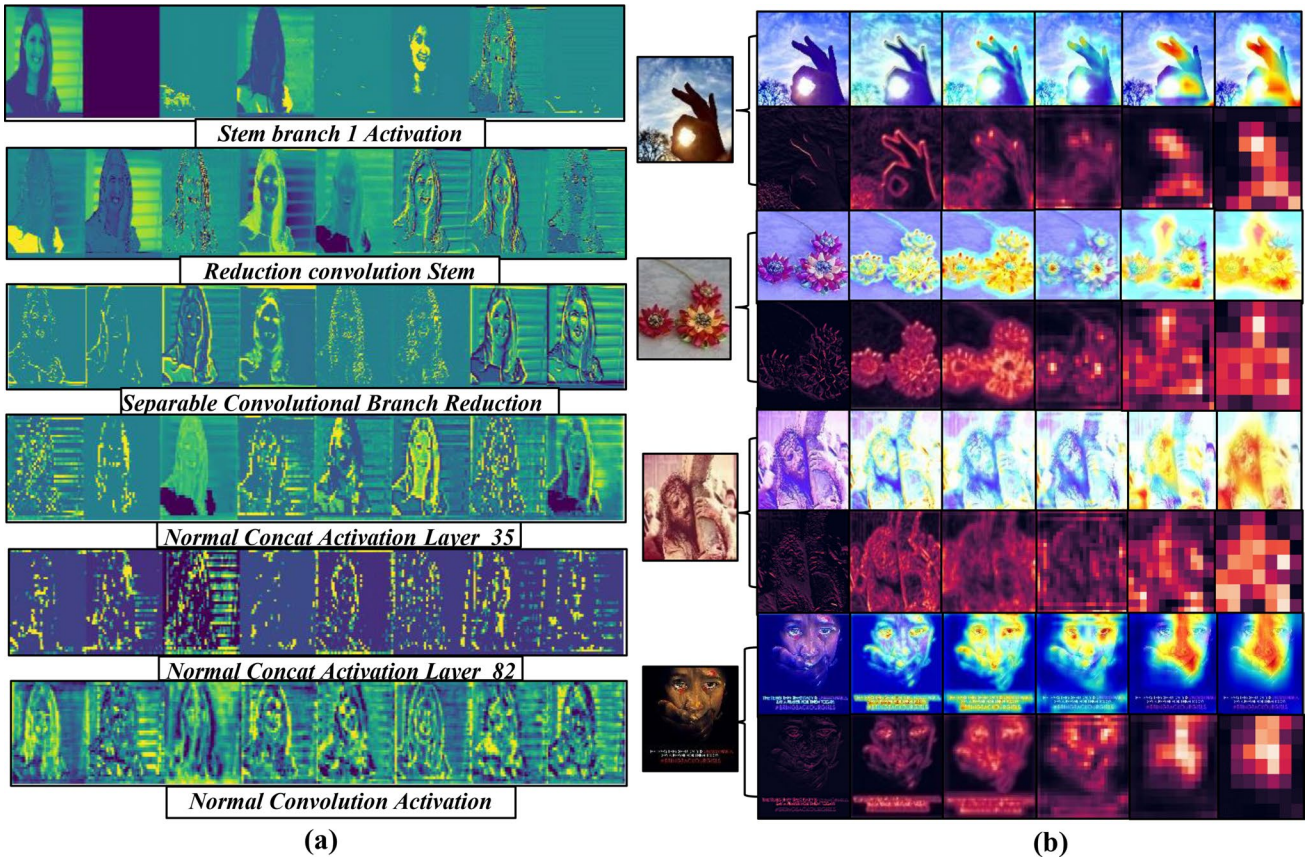


Fig. 9 **a** Activation maximation plots to visualize each channel generated by different layers of the RA-DLNet architecture. **b** Colormaps and corresponding heatmaps of some images to visualize the crucial regions of example images

the Flickr sentiment dataset. The results are reported in Table 11 and the highest ACC is mentioned in bold.

We conclude several important observations from the study. When we removed the trunk branch and soft mask branch of our attention module and extracted the spatial features from the images, we achieved only 45.1% accuracy. This verifies the effectiveness of our attention module for extracting the visual sentiments and proves that the attention module is necessary for more accurate sentiment prediction. However, if only the soft mask branch is ablated, we observe that the model reaches the lowest accuracy. The reason behind this behavior is that when the network is trained with noisy data, the wrong gradients computed from noisy labels might update the truck branch parameters, which further deteriorate the learning of our model. Hence, the soft mask branch was added to prevent the truck branch from getting updated with gradients caused by label error.

When *Dense_1* (2048) and *Dense_2 (1024)* layers are ablated, accuracy decreases to 71.3% and 76.6% respectively. Since the decrease in the number of parameters from the pooling layer to *Dense_2 (1024)* is sudden, the model's efficiency is profoundly affected. However, the reduction in the number of parameters from *Dense_1 (2048)* to *Dense_3 (2)* is quite smooth, which has slightly less impact on the accuracy. The intuition behind this is that softer dimensionality reduction proves beneficial for the architecture which was provided by intermediate fully connected layers. The evolving accuracy at each layer shows how the features are learned effectively along with the network.

Visualizations To visualize what information is captured by different layers of the proposed model, we show the *activation maximation plots* in Fig. 9a. From the figure, we observe that as we go deeper and deeper in the layers, the input image becomes less visually interpretable. This is because, initially, all the filters of the layer are not activated, and hence, most of the information of the image is retained. But as we go deeper into the layers, more and more filters get activated to get a higher representation of the image. These representations carry less information about the visuals of the image and more details on the class of the image. Finally, a point comes when the layers are not activating any filters, and the network has nothing more to learn. At this point, the training of the model gets stopped. Hence, these plots explain how the particular filters are activated in each layer from the input pattern.

To provide more visualizations on what the CNN learns from the visual patterns of an image, we show the color maps and the corresponding heatmaps of some images in Fig. 9b. These maps highlight the crucial regions of the

Table 12 Summary of CNN architectures

Model	Input size	Top-1 accuracy (%)	Top-5 accuracy (%)	Parameters (millions)	Depth	Size (MB)
VGG-16	224×224	71.3	90.1	138	23	528
VGG-19	224×224	71.3	90.0	143	26	549
ResNet-50	224×224	74.9	92.1	25	–	98
Inception-V3	299×299	78.8	94.4	23	159	92
InceptionResnet-V2	299×299	80.1	95.1	55	572	215
Xception	299×299	79.0	94.5	22	126	88
NASNet	331×331	82.7	96.2	88	–	343

Table 13 Five cross-validation results (mean ± std) on other popular architectures

Dataset	VGG-16	VGG-19	ResNet – 50	Inception-V3	Inception-Resnet V2	Xception	Nasnet
3-Agree	67.6 ± 0.67	67.2 ± 0.42	52.3 ± 2.12	75.7 ± 3.07	76.8 ± 3.86	77.4 ± 0.54	81.3 ± 0.70
4-Agree	72.4 ± 0.82	70.0 ± 2.92	38.2 ± 2.78	76.6 ± 1.45	78.1 ± 0.72	78.8 ± 3.84	83.2 ± 0.84
5-Agree	77.4 ± 0.02	75.2 ± 0.75	34.4 ± 1.85	79.6 ± 1.29	84.5 ± 2.22	84.8 ± 3.13	89.1 ± 0.01
Twitter II	76.1 ± 3.16	75.5 ± 2.59	77.9 ± 2.07	75.4 ± 0.24	77.9 ± 0.94	74.3 ± 2.75	81.2 ± 1.18
ArtPhoto	43.1 ± 2.85	45.3 ± 2.16	31.6 ± 2.41	40.2 ± 2.82	63.4 ± 8.29	53.1 ± 2.16	71.8 ± 0.22
CMU-MOSI	50.7 ± 1.99	51.7 ± 0.87	39.0 ± 2.77	41.5 ± 1.77	49.8 ± 0.77	61.0 ± 1.11	67.7 ± 1.87
CMU-MOSEI	53.9 ± 2.74	54.0 ± 1.68	44.4 ± 2.01	52.6 ± 0.74	67.7 ± 0.04	64.6 ± 0.22	77.2 ± 2.60
Flickr	70.4 ± 1.88	69.3 ± 2.74	50.9 ± 3.11	68.8 ± 1.12	76.1 ± 0.94	80.7 ± 1.76	83.4 ± 1.23
Instagram	65.5 ± 1.01	60.0 ± 1.34	45.0 ± 2.00	67.4 ± 1.26	72.3 ± 0.78	78.0 ± 2.44	82.7 ± 0.91
F&I	56.7 ± 1.58	61.7 ± 1.11	53.4 ± 2.78	60.1 ± 0.57	63.8 ± 1.44	68.9 ± 0.94	72.8 ± 0.06

Table 14 Training efficiency of all architectures (time/epoch) (s)

Dataset	VGG-16	VGG-19	ResNet-50	Inception-V3	Inception-Resnet-V2	Xception	NASNet
3-Agree	516	533	485	742	796	759	1031
4-Agree	501	566	483	723	795	770	1041
5-Agree	544	500	488	710	767	764	1046
Twitter II	498	512	511	718	787	1030	1098
Artphoto	336	418	195	224	411	409	1100
CMU-MOSI	490	471	544	658	704	958	1177
CMU-MOSEI	511	504	497	610	621	744	1244
Flickr	710	729	702	854	942	1024	1300
Instagram	628	714	647	904	1011	1297	1845
F&I	604	611	694	684	710	984	1475

image which helps CNN to make the final prediction. This shows what features CNN relies on by focusing on that region. As seen in the figure, the first input image corresponds to a human hand. The initial layers are not exactly concentrating on the hand, but as we reach towards the final layer, the network starts focusing on the hand region of the input image. Accordingly, in the corresponding heatmap of the image, the pixels of the hand regions are getting activated towards the final layer. This shows that the model is learning to localize the crucial visual features of the image.

4.3.6 Comparison of NASNet with other CNN architectures

To further examine the competence of the proposed architecture, we have compared the performance of the NASNet with six popular CNN architectures: VGG-16 [71], VGG-19 [71], Inception-Resnet-V2 [72], Inception-V3 [73], ResNet-50 [74], and Xception [75] by replacing the NASNet layer in our proposed RA-DLNet architecture with the above six architectures and training the model end-to-end for each architecture. The experimental settings for every architecture are the same as described in Sect. 4.2.1.

Table 12 shows the complete summary of these architectures. The top-1 accuracy and top-5 accuracy denotes the performance of the model on the ImageNet dataset. Table 13 shows a comparison between different models based on fivefold cross-validation accuracy on all the visual sentiment datasets. To reduce the variance, each experiment is performed five times. The mean $\pm$ standard deviation of the accuracies is reported in Table 13.

As evident from the above table, the top-performing CNN for the VSA task is NASNet. However, in Table 12, we see that ResNet-50 has shown better results than VGG-16 and VGG-19 architectures on the ImageNet dataset. Surprisingly, the results obtained for VSA does not convey the same. Despite being more sophisticated than VGG-16 and VGG-19 architectures, ResNet-50 has shown the worst performance on visual sentiment datasets. In Twitter I dataset, the maximum accuracy achieved by ResNet-50 is around 53%, whereas the minimum accuracy on VGG-16 and VGG-19 is 68%. The same behavior is observed on CMU-MOSI, Flickr, Instagram, and F&I datasets. However, in ArtPhoto and CMU-MOSEI dataset, VGG-16 and VGG-19 perform better than Inception-V3 as that latter overfits the data and gives less accurate results. For further analysis, we have compared the training efficiency of different architectures on all the datasets. Table 14 shows the results in seconds. NASNet has taken the highest training time on every dataset and ResNet-50 takes the minimum time, whereas in F&I emotion dataset, InceptionV3 takes the minimum time. However, as seen in Table 13, ResNet-50 has performed poorly for visual sentiment classification. Thus, it is crucial to study the trade-off between accuracy and training time for selecting the best architecture.

The experimental results suggest that the deep learning-based methods can outperform low-level and mid-level based approaches for VSA. The advantage of using a pre-trained model is that the simplicity of the model makes it easy to learn and train according to the requirements. Moreover, by combining the local and global features in the images, the final discriminative features are captured by the RA-DLNet architecture which helps in visual sentiment analysis.

5 Conclusion

This work presents a RA-DLNet for VSA. The proposed framework can extract the sentiments and emotions from the uploaded photos on various social media platforms like Twitter, Flickr, and Instagram. Since the uploaded images are based on the real-life experience of the users, the RA-DLNet architecture is validated on these images to show the effectiveness of the model in capturing the real sentiments expressed by the users. This is accomplished by fine-tuning a pre-trained model to fetch the features and pass them to the residual attention network for capturing the most informative

regions of the image. The results prove that the RA-DLNet architecture overpass the state-of-the-art results on the six publicly available, real-world challenging datasets for binary sentiment classification and on two benchmark datasets for multi-class emotion classification. Furthermore, we have provided a comparative analysis of the NASNet architecture used in our model with the six popular CNN architectures to show the efficacy of the RA-DLNet architecture.

In future, the proposed RA-DLNet architecture can be applied for movie genre classification by developing an emotion theory that will study the impact of evoked emotions on humans from the movie trailers. Furthermore, we will extend our work to incorporate multimodal data in the form of speech and text to analyze the impact of each modality for multimodal sentiment classification.

References

1. You, Q., Luo, J., Jin, H., Yang, J.: Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 381–388. USA (2015)
2. Ohn-bar, E., Trivedi, M.M.: Multi-scale volumes for deep object detection and localization. Pattern Recogn. **61**, 557–572 (2016)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 142–158 (2016)
4. Oquab, M., Bottou, L.: Learning and transferring mid-level image representations using convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition Learning, pp. 1717–1724. Columbus, OH (2014)
5. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. Florida (2009)
6. Chu, B., Madhavan, V., Beijbom, O., Hoffman, J., Darrell, T.: Best practices for fine-tuning visual classifiers to new domains. In: Hua, G., Jégou, H. (eds.) European Conference on Computer Vision, pp. 435–442. Springer, Amsterdam (2016)
7. Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.-F.: Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: 21st ACM International Conference on Multimedia, pp. 223–232 (2013)
8. Siersdorfer, S., Minack, E., Deng, F., Hare, J.: Analyzing and predicting sentiment of images on the social web. In: 18th ACM International Conference on Multimedia, pp. 715–718 (2010)
9. Vonikakis, V., Winkler, S.: Emotion-based sequence of family photos. In: Proceedings of the 20th ACM International conference on Multimedia, pp. 1371–1372 (2012)
10. Jia, J., Wu, S., Wang, X., Hu, P., Cai, L., Tang, J.: Can we understand van gogh's mood? Learning to infer affects from images in social networks. In: 20th ACM International Conference on Multimedia, pp. 857–860 (2012)
11. Li, B., Feng, S., Xiong, W., Hu, W.: Scaring or pleasing: exploit emotional impact of an image. In: 20th ACM International Conference on Multimedia, pp. 1365–1366 (2012)
12. Wang, S., Wang, J., Wang, Z., Ji, Q.: Multiple emotion tagging for multimedia data by exploiting high-order dependencies among emotions. IEEE Trans. Multimedia **17**(12), 2185–2197 (2015)
13. Yuan, J., You, Q., Mcdonough, S., Luo, J.: Sentribute: image sentiment analysis from a mid-level perspective. In: Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, pp. 1–8. Chicago (2013)
14. Zhao, S., Gao, Y., Jiang, X., Yao, H., Chua, T., Sun, X.: Exploring principles-of-art features for image emotion recognition. In: 22nd ACM International Conference on Multimedia, pp. 47–56. Florida (2014)
15. Chen, Y., Chen, T., Liu, T., Liao, H.Y.M., Chang, S.: Assistive image comment robot—a novel mid-level concept-based representation. IEEE Trans. Affect. Comput. **6**(3), 298–311 (2015)
16. Chen, F., Ji, R., Su, J., Cao, D., Gao, Y.: Predicting microblog sentiments via weakly supervised multimodal deep learning. IEEE Trans. Multimedia **20**(4), 997–1007 (2018)
17. Yang, J., She, D., Sun, M., Cheng, M.-M., Rosin, P.L., Wang, L.: Visual sentiment prediction based on automatic discovery of affective regions. IEEE Trans. Multimedia **20**, 2513–2525 (2018)
18. Xiong, H., Liu, Q., Song, S., Cai, Y.: Region-based convolutional neural network using group sparse regularization for image sentiment classification. EURASIP J. Image Video Process. **30**, 1–9 (2019)
19. Zhao, B., Wu, X., Feng, J., Peng, Q., Yan, S.: Diversified visual attention networks for fine-grained object classification. IEEE Trans. Multimedia **19**(6), 1245–1256 (2017)
20. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 8697–8710. Utah (2018)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations. California (2015)
22. Szegedy, C., Vanhoucke, V., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2015)
23. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: International Conference on Learning Representations. California (2015)
24. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3156–3164 (2017)
25. Campos, V., Salvador, A., Jou, B., Giró-i-nieto, X.: Diving deep into sentiment: understanding fine-tuned CNNs for visual sentiment prediction. In: 1st International Workshop on Affect & Sentiment in Multimedia, pp. 57–62 (2015)
26. Wang, J., Fu, J., Xu, Y., Mei, T.: Beyond object recognition: visual sentiment analysis with deep coupled adjective and noun neural networks. In: Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 3484–3490. New York (2016)
27. Song, K., Yao, T., Ling, Q., Mei, T.: Boosting image sentiment analysis with visual attention. Neurocomputing **312**, 218–228 (2018)
28. Islam, J., Zhang, Y.: Visual sentiment analysis for social images using transfer learning approach. In: IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom), pp. 124–130. Atlanta (2016)
29. Fan, S., Jiang, M., Shen, Z., Koenig, B.L., Kankanhalli, M.S., Zhao, Q.: The role of visual attention in sentiment prediction. In: 25th ACM International Conference on Multimedia, pp. 217–225. California (2017)
30. Sharma, R., Tan, L.N., Sadat, F.: Multimodal sentiment analysis using deep learning. In: 17th IEEE International Conference on Machine Learning and Applications, pp. 1475–1478 (2018)

31. Li, Z., Jiao, Y., Yang, X., Zhang, T., Huang, S.: 3D attention-based deep ranking model for video highlight detection. IEEE Trans. Multimedia **20**(10), 2693–2705 (2018)
32. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI Conference on Artificial Intelligence, pp. 4278–4284. Arizona (2016)
33. Li, Z., Fan, Y., Liu, W., Wang, F.: Image sentiment prediction based on textual descriptions with adjective noun pairs. Multimedia Tools Appl. **77**(1), 1115–1132 (2017)
34. Yang, H., Yuan, C., Li, B., Du, Y., Xing, J.: Asymmetric 3D convolutional neural networks for action recognition. Pattern Recogn. **85**, 1–12 (2019)
35. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258. Honolulu (2017)
36. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. Las Vegas (2016)
37. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1249–1258. Las Vegas (2016)
38. Toshev, A., Szegedy, C.: DeepPose: human pose estimation via deep neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, pp. 1653–1660. Ohio (2014)
39. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587. Columbus, Ohio (2014)
40. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms (2017). arXiv preprint arXiv:1707.06347
41. Machajdik, J., Hanbury, A.: Affective image classification using features inspired by psychology and art theory. In: ACM International Conference on Multimedia, pp. 83–92 (2010)
42. Wang, X., Jia, J., Yin, J., Cai, L.: Interpretable aesthetic features for affective image classification. In: IEEE International Conference on Image Processing, pp. 3230–3234 (2013)
43. Rao, T., Xu, M., Liu, H., Wang, J., Burnett, I.: Multi-scale blocks based image emotion classification using multiple instance learning. In: IEEE International Conference on Image Processing (ICIP), pp. 634–638. Arizona (2016)
44. Rao, T., Xu, M., Liu, H.: Generating affective maps for images. Multimedia Tools Appl. **77**(13), 17247–17267 (2018)
45. Liu, X., Li, N., Xia, Y.: Affective image classification by jointly using interpretable art features. J. Vis. Commun. Image Represent. **58**, 576–588 (2019)
46. Campos, V., Jou, B., Giro-i-Nieto, X.: From pixels to sentiment: fine-tuning CNNs for visual sentiment prediction. Image Vis. Comput. **65**, 15–22 (2017)
47. Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. In: International Conference on Learning representations (2017)
48. Yadav, A., Vishwakarma, D.K.: Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, pp. 1–51 (2019)
49. She, D., Yang, J., Cheng, M.M., Lai, Y.K., Rosin, P.L., Wang, L.: WSCNet: weakly supervised coupled networks for visual sentiment classification and detection. IEEE Trans. Multimedia. **22**(5), 1358–1371 (2019)
50. Fan, S., Jiang, M., Koenig, B.L., Xu, J., Kankanhalli, M.S., Zhao, Q.: Emotional attention: a study of image sentiment and visual attention. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7521–7531. Salt Lake (2018)
51. Lee, J., Kim, S., Kim, S., Park, J., Sohn, K.: Context-aware emotion recognition networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10143–10152. Seoul (2019)
52. Bawa, V.S., Kumar, V.: "Emotional sentiment analysis for a group of people based on transfer learning with a multi-modal system. Neural Comput. Appl. **31**(12), 9061–9072 (2018)
53. Yang, J., She, D., Sun, M.: Joint image emotion classification and distribution learning via deep convolutional neural network. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), pp. 3266–3272 (2017)
54. Zhu, X., Li, L., Zhang, W., Rao, T., Xu, M., Huang, Q., Xu, D.: Dependency exploitation: a unified CNN-RNN approach for visual emotion recognition. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pp. 3595–3601 (2017)
55. Yang, J., She, D., Lai, Y.K., Yang, M.H.: Retrieving and classifying affective images via deep metric learning. In: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), pp. 491–498. Louisiana (2018)
56. Zhao, S., Lin, C., Xu, P., Zhao, S., Guo, Y., Krishna, R., Ding, G., Keutzer, K.: CycleEmotionGAN: emotional semantic consistency preserved CycleGAN for adapting image emotions. In: The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), pp. 2620–2627. Hawaii (2019)
57. Zhang, W., He, X., Lu, W.: Exploring discriminative representations for image emotion recognition with CNNs. IEEE Trans. Multimed. **22**(2), 515–523 (2019)
58. Chen, T., Borth, D., Darrell, T., Chang, S.F.: Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks (2014). arXiv preprint arXiv:1410.8586
59. Katsurai, M., Satoh, S.: Image sentiment analysis using latent correlations among visual, textual, and sentiment views. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2837–2841 (2016)
60. He, X., Zhang, H., Li, N., Feng, L., Zheng, F.: A multi-attentive pyramidal model for visual sentiment analysis. In: International Joint Conference on Neural Networks, pp. 1–8 (2019)
61. Yang, J., She, D., Lai, Y.K., Rosin, P.L., Yang, M.H.: Weakly supervised coupled networks for visual sentiment analysis. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7584–7592. Salt Lake City (2018)
62. Zadeh, A., Zellers, R., Pincus, E., Morency, L.P.: MOSI: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos (2016). arXiv preprint arXiv:1606.06259
63. Zadeh, A., Liang, P.P., Vanbriesen, J., Poria, S., Tong, E., Cambria, E., Chen, M., Morency, L.P.: Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 2236–2246 (2018)
64. You, Q., Luo, J., Jin, H., Yang, J.: Building a large scale dataset for image emotion recognition: the fine print and the benchmark. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), pp. 308–314. Arizona (2016)
65. Dumpala, S.H., Sheikh, I., Chakraborty, R., Kopparapu, S.K.: Sentiment classification on erroneous ASR transcripts: a multi view learning approach. In: IEEE Spoken Language Technology Workshop (SLT 2018), pp. 807–814. Greece (2018)
66. Dumpala, S.H., Sheikh, I., Chakraborty, R., Kopparapu, S.K.: Audio-visual fusion for sentiment classification using cross-modal autoencoder. In: 32nd Conference on Neural Information Processing Systems (NIPS 2018), pp. 1–4. Canada (2018)
67. Chauhan, D.S., Akhtar, M.S., Ekbal, A., Bhattacharyya, P.: Context-aware interactive attention for multi-modal sentiment and emotion analysis. In: Proceedings of the 2019 Conference on

Empirical Methods in Natural Language Processing, pp. 5646–5656 (2019)
68. Akhtar, M.S., Chauhan, D.S., Ghosal, D., Poria, S., Ekbal, A., Bhattacharyya, P.: Multi-task learning for multi-modal emotion recognition and sentiment analysis. In: Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 370–379. Minnesota (2019)
69. Sun, Z., Sarma, P.K., Sethares, W.A., Liang, Y.: Learning relationships between text, audio, and video via deep canonical correlation for multimodal language analysis. In: AAAI Conference on Artificial Intelligence (AAAI) (2019)
70. Zadeh, A., Chen, M., Poria, S., Cambria, E., Morency, L.P.: Tensor fusion network for multimodal sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1103–1114 (2017)
71. Chen, M., Wang,S., Liang, P.P., Baltrušaitis, T., Zadeh, A., Morency, L.P.: Multimodal sentiment analysis with word-level fusion and reinforcement learning. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI), pp. 163–171 (2017)
72. Li, H., Xu, H.: Video-based sentiment analysis with hvnLBP-TOP feature and bi-LSTM. In: The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), pp. 9963–9964. Hawaii (2019)
73. Zadeh, A., Liang, P.P., Poria, S., Vij, P., Cambria, E., Morency, L.P.: Multi-attention recurrent network for human communication comprehension. In: Thirty-Second AAAI Conference on Artificial Intelligence, pp. 5642–5649. Louisiana (2018)
74. Yadav, A., Vishwakarma, D.K.: A comparative study on bio-inspired algorithms for sentiment analysis. Cluster Comput. (2020). https://doi.org/10.1007/s10586-020-03062-w
75. Sun, Z., Sarma, P.K., Sethares, W., Bucy, E.P.: Multi-modal sentiment analysis using deep canonical correlation analysis. In: The 20th Annual Conference of the International Speech Communication Association (INTERSPEECH), pp. 1323–1327 (2019)

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

Neural Computing and Applications
https://doi.org/10.1007/s00521-020-05018-y

ORIGINAL ARTICLE

A deeply coupled ConvNet for human activity recognition using dynamic and RGB images

Tej Singh[1] · Dinesh Kumar Vishwakarma[2]

Received: 6 May 2019 / Accepted: 7 May 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract
This work is motivated by the tremendous achievement of deep learning models for computer vision tasks, particularly for human activity recognition. It is gaining more attention due to the numerous applications in real life, for example smart surveillance system, human–computer interaction, sports action analysis, elderly healthcare, etc. Recent days, the acquisition and interface of multimodal data are straightforward due to the invention of low-cost depth devices. Several approaches have been developed based on RGB-D (depth) evidence at the cost of additional equipment's setup and high complexity. Contrarily, the methods that utilize RGB frames provide inferior performance due to the absence of depth evidence and these approaches require to less hardware, simple and easy to generalize using only color cameras. In this work, a deeply coupled ConvNet for human activity recognition proposed that utilizes the RGB frames at the top layer with bi-directional long short-term memory (Bi-LSTM). At the bottom layer, the CNN model is trained with a single dynamic motion image. For the RGB frames, the CNN-Bi-LSTM model is trained end-to-end learning to refine the feature of the pre-trained CNN, while dynamic images stream is fine-tuned with the top layers of the pre-trained model to extract temporal information in videos. The features obtained from both the data streams are fused at decision level after the softmax layer with different late fusion techniques and achieved high accuracy with max fusion. The performance accuracy of the model is assessed using four standard single as well as multiple person activities RGB-D (depth) datasets. The highest classification accuracies achieved on human action datasets are compared with similar state of the art and found significantly higher margin such as 2% on SBU Interaction, 4% on MIVIA Action, 1% on MSR Action Pair, and 4% on MSR Daily Activity.

Keywords Bi-LSTM · Deep learning · Dynamic motion image · Human activity recognition

1 Introduction

Recently, the vision-based understanding of event entices numerous real-life applications such as gaming, robotics, patients monitoring, content-based retrieval, video surveillance, and security [1–4]. Over the decade, many efforts made to recognize the human activity in videos, but still, it is a challenging task due to intra-class action similarity, occlusions, view variations, and environmental conditions [5, 6]. Since the introduction of inexpensive depth Kinect sensor, various solutions proposed based on multiple modalities using RGB, depth, and 3D coordinate data and their multiple combinations. However, such approaches face limitations due to positions of joints and coordinates of upper or lower body parts used to represent activity. An analysis of existing challenges in RGB and RGB-D datasets of human activity is discussed in [7, 8]. It is observed that most of the real-life applications for activity recognition use RGB only cameras such as CCTVs installed on public places for monitoring purposes. However, the solutions based on features extraction from RGB only frames are not sufficient enough to identify suspicious human activity due to the lack of temporal information and

✉ Dinesh Kumar Vishwakarma
dinesh@dtu.ac.in

Tej Singh
tejsingh_2k16phdec05@dtu.ac.in

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

[2] Department of Information Technology, Delhi Technological University, New Delhi 110042, India

Published online: 21 May 2020

existing environmental conditions. Formerly, the optical flow estimation is most popular motion representation method in video sequences. However, optical flow-based solutions are less effective in challenging conditions such as view variations and occlusions present in videos. Therefore, we utilized the concept of a compact single dynamic motion image (DMI) obtained from the whole action video using an approximate rank pooling mechanism [9]. The DMIs captured the long-term dynamics and motion cues to understand temporal patterns in videos automatically. In this work, we utilized RGB only frames to learn the features that make our model less complex, robust, and efficiently apply for daily life activities.

In the proposed hybrid model, we utilized the latest pre-trained Inception-v3 deep architecture [10] that outperforms and shows excellent accuracy results in ILSVR 2012 image classification competition as compared to other pre-trained deep models such as AlexNet [11] and VGGNet [12]. Further, the Bi-LSTM model is used to deal with sequential data that increases the overall efficiency of the proposed architecture. Our deep model outperforms the state-of-the-art methods on standard human activity RGB-D datasets by utilizing RGB only frames of video.

The remaining sections of the proposed work are as follows: Sect. 2 consists of related work of various handcrafted and deeply learned approaches for human activity analysis. Section 3 consists of the underlying architecture of proposed deeply coupled ConvNet. The experimental setup, training details, results, and a discussion on standard video benchmarks are explained in Sect. 4. Finally, the work is concluded in Sect. 5.

2 Related work

In the last decade, many handcrafted and automatically learned feature-based approaches developed for human action recognition in the videos. Earlier human activity recognition approaches are based on handcrafted features mainly focused on simple atomic actions that seem to be somewhat less useful to practical applications [13–18]. The main drawback of these approaches is data pre-processing and difficult to generalize in real life despite gaining a high accuracy model. Later on, after the success of convolutional neural networks (CNNs) on text and image classification, various spatiotemporal approaches for video activity analysis were proposed that can automatically learn the features and classify from raw RGB video only [19–21]. However, such approaches could not achieve higher accuracy due to data dependency for training the CNN models and the lack of hardware resources [22–26].

Initially, a global representation-based features approaches are proposed for human activity analysis. Bobick and Davis [27] introduced the average energy-based approach motion history image (MHI) and motion energy image (MEI) to recognize human activities in controlled environmental conditions. Blank et al. [28] addressed the activity classification problem using space–time 3D template MEI based on extracted human silhouettes. However, such holistic solutions are less sensitive to measure the possible variation in the videos, e.g., view variation and occlusions [29]. To overcome the limitations of global approaches, Laptev [30] proposed an efficient local feature extractor space–time interest points (STIPs) for action recognition by utilizing the extension of Harris detector. Matikainen et al. [31] introduced a trajectory-based local feature extractor to recognize human action in the video sequence. In order to recognize the actions of different length and timescales in real time, a method based on the string kernel is proposed by Brun et al. [32]. They represented an action with the help of a string called "aclet," and similarity between these "aclets" is described based on Gaussian kernel. Charletti et al. [33] proposed a hybrid feature extractors approach to recognize human activity using depth videos. The dimensionality of obtained feature vectors is reduced with the help of LDA and PCA techniques. Finally, the activities are classified using GMM classier. However, they claimed superior accuracy, but their modal is sensitive to view variations and noises. Ji et al. [23] introduced a soft regression-based transition maps approach for early detection of human activities using depth frames only. They divided human action into different patterns and evaluated the temporal coherence between action sequences. Seggesi et al. [34] proposed an automatic configurable trained feature extractor for the representation of skeleton pose from training samples data. However, their modal showed promising results on static poses, but the response is found slow for real-time action recognition. To recognize the online action in a complex background using depth cameras, Ji et al. [22] proposed a hybrid approach by embedding skeleton coordinates into depth frames and extracted features using a spatiotemporal pyramid on a partitioned set of action sequences. However, these methods are showing satisfactory performance but less competent to tackle the environmental changes such as camera jitters, occlusion, and illumination variations [35–38].

The first step toward the automatic features learning from raw RGB frames was introduced by Ji et al. [20]. This approach is based on 3D CNN for human action recognition in videos. However, due to rigid architecture for accepting fixed input frames, such a model is not so efficient for video of different lengths. Simoyan and Zisserman [21] proposed a two-stream deep model that accepts raw RGB frames and optical flow as an input to the pre-trained deep learned model. Baccouche et al. [19] introduced a

hybrid model for action recognition that extracts spatial features using CNN and LSTM for motion cues. Currently, a combination of both in-depth and local features are popular for activity recognition [21, 39–41]. Feng et al. [26] proposed a geometrical relational features approach based on multilayer LSTM network for recognition of human activities using skeleton joints information. Keçeli et al. [42] presented an approach to recognize the dyadic activity from depth sequence that is the combination of 3D and 2D CNN architectures. They extract the temporal features through 3D CNN trained 3D depth volume, while 2D CNN is fine-tuned on weighted sum depth sequences. The obtained features are ranked using relief algorithm and classify using an SVM classifier. Ijjina and Chalavadi [43] proposed a multimodal action recognition model based on feature extraction from RGB and depth videos using CNN architecture. An ELM classifier is used to recognize the human activities from these fused features architecture. Jing et al. [44] presented a spatiotemporal-based hybrid neural network which characterizes human activity in RGB-D video based on a two-stream neural network. Further, they claimed to improvement in the classification accuracy by utilizing joint loss function to exploit the spatial and temporal features of videos. To address the human activity, Srihari et al. [45] introduced a four-stream CNN network consisted of two RGB-D video data and two temporal motion optical flow streams. Elboushaki et al. [46] proposed a multi-dimensional CNN that learned high-level features for gesture recognition in RGB-D videos in conjunction with LSTM network. They investigated various fusion schemes at different layers of a deep network for the classification task.

Based on the above literature review, it can be observed that shape and motion information is correlated with in-depth sequences. It is challenging to record both of these information simultaneously. Furthermore, the 3D skeleton joint coordinates are not capable of discriminating some activities due to noise and occlusion errors such as self-occlusion with body parts, etc. There are some action examples such as 'eating' and 'drinking' having the same motion pattern which cannot be distinguished by using 3D skeleton joints coordinates. Hence, motivated from the earlier state of the art, we have designed a tightly coupled deep ConvNet for human activity recognition using RGB frames of the videos.

3 The proposed ConvNet architecture

The improved image recognition approaches extend and motivate automatic human action recognition in video sequences. Most of the video action recognition approaches are based on shallow higher-dimensional spatiotemporal features extraction from stacked raw video frames. Human action can be disintegrated into spatial and temporal variations in a video. The spatial features contain the appearance information about an object in each sequence of a given video. On the other hand, temporal features represented in the form object moving across the video sequences. The proposed human action recognition model accordingly divided into two streams model as depicted in Fig. 1. Hence, in this work, a hybrid two-stream deep architecture based on two different data streams (spatial and temporal) is proposed, which are then fused by late fusion techniques. The spatial features are extracted from RGB frames fed at regular interval to the input of first or upper stream as shown in Fig. 1, and at the same time, the temporal or second or bottom stream has input as DMI, which captures the full temporal dynamics of human action. Both the streams are trained using pre-trained Inception-v3 deep architecture.

The first or upper stream of the network is trained end-to-end learning through a pre-trained ConvNet and followed by the Bi-LSTM network for the acquisition of additional sequential information. The fully connected layer flattened the output of Bi-LSTM, and softmax layer gives the probabilistic score, which is going to fuse with the temporal score obtained by the second or bottom stream.

The second or bottom stream of the network is fine-tuned on the pre-trained ConvNet, and in this stream, a temporal enrich images called DMIs are inputted to the network, which extracts temporal features. These obtained features are passed to a fully connected layer, which flattened the features, and further softmax layer gives the probabilistic temporal score.

These streams of networks are connected parallelly, as shown in Fig. 1. Further, the scores obtained by these two streams of the network are fused with late fusion techniques at the decision level to enhance the classification accuracy of the proposed model.

3.1 Features extraction with pre-trained Inception-v3 architecture

In this section, we briefly discussed the underlying architecture of deep Inception-v3. Figure 2 shows the deep Inception-v3 architecture consisted of one input block, three blocks of Inception Modules A, B and C, two blocks of grid size reduction, one auxiliary classifier block and one output block. Inception-v3 network enriched with some advanced features such as RMSProp optimizer, batch normalization, label smoothing to reduce overfitting and add loss function as compared with the previous version of inception architectures. This network consists of 42 deep layers that accept input data $299 \times 299 \times 3$ of spatial

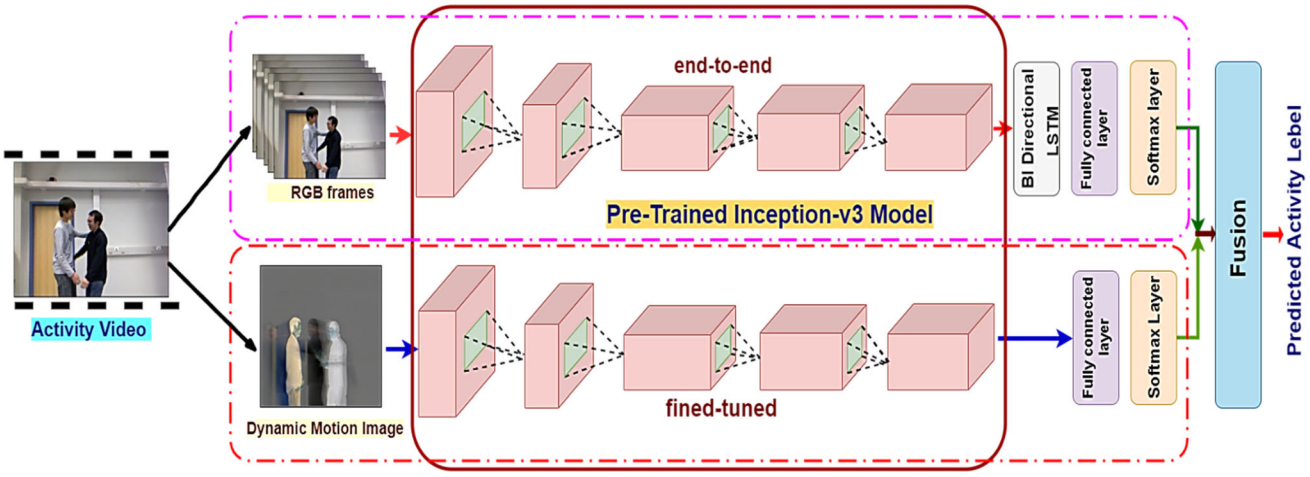


Fig. 1 The proposed ConvNet architecture for human activity recognition

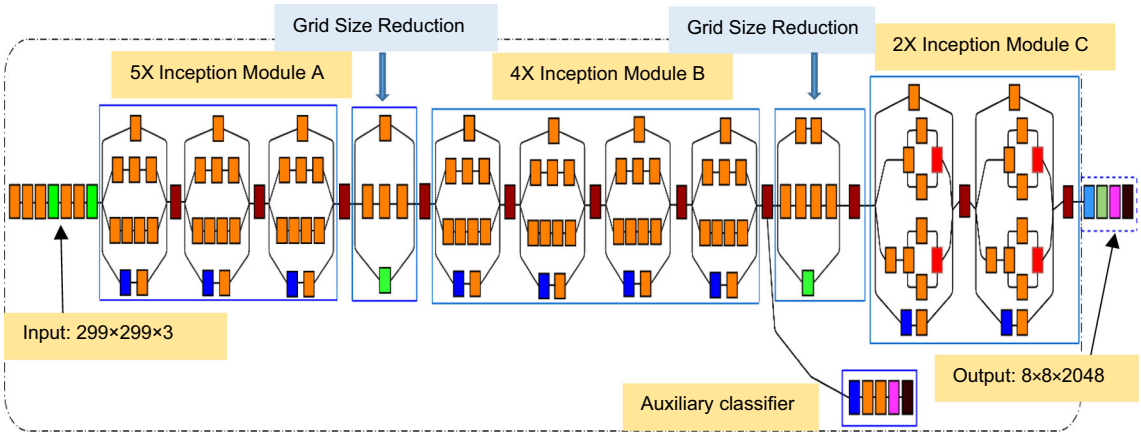


Fig. 2 Block diagram of Inception-v3 improved deep architecture

dimension at input block. The Inception Module A is used for smaller factorization convolutional (Conv), and a 5×5 (Conv) filter is converted into two 3×3 Conv filters which results in reducing parameters between the layers without decreasing the efficiency of the network. The Inception Module B utilized for asymmetric spatial factorization convolutions which converts a 3×3 Conv filter into 1×3 Conv followed by 3×1 Conv filters. It can be observed that a $n \times n$ Conv can be represented by $1 \times n$ Conv followed by a $n \times 1$ Conv saving the computational cost and reduced the overall parameters. The Inception Module C is introduced for stimulating the high-dimensional representations similarly working as Module B in this network. The grid size reduction block is used for downsizing the feature map such as in deep AlexNet or VGGNet models. The main difference between traditional models and Inception-v3 is that a $m \times m$ grid with n filters is divided into $m/2 \times m/2$ grid with $2n$ filters. Thus, the overall computational cost is decreased by using convolutional operation followed by pooling operation. This network contains one auxiliary classifier on the top of the last 17×17 layer, which is used as a regularizer for enhancing the convergence of the deep network. In the following subsection, we briefly discussed the underlying architecture of LSTM and Bi-LSTM for sequential data. In the proposed work, the deep Inception-v3 network is trained in an end-to-end manner followed by the Bi-LSTM architecture, which represents one of the action descriptors for the two-stream HAR model.

3.2 The Bi-Directional LSTM (Bi-LSTM)

It is observed that the traditional feedforward network and convolutional neural network are inefficient to deal with sequential data such as video analysis. These networks take input of fixed length. To overcome the problem of fixed size, input padding is used, but the performance of such approaches is not comparable with RNNs [47] and LSTM network [48]. The RNN deep network has compatibility with the help of the chain and loop structure to deal with

sequential data. However, RNNs are inefficient as compared to long short-term memory for longer duration sequential input because of the vanishing gradient in the backpropagation process. To overcome the problem of vanishing gradient, the LSTM architecture was proposed, which utilized two parallel RNNs for enhancing the long-term dependencies and training for bigger input networks. The gated structure of the LSTM network helped to overcome the many sequential learning problems in a very effective way. Figure 3 illustrates the underlying blocks architecture of LSTM network. It can be observed from Fig. 3 that LSTM network has a similar chain-like structure like RNN. It consisted of four neural layers, while RNN is having only one neural sigmoid layer (tan$h$). A horizontal line at the top of Fig. 3 is called the "cell state" of LSTM which works like a conveyor belt. This cell state has the capabilities to remove and add information. The LSTM consists of three gates: input (i_t), forget and output (o_t) gates to control and protect the cell states. These gates have sigmoid neural layer and capable of processing the information through pointwise multiplication units.

In the beginning, the "forget gate" (f_t) decides what information between (h_{t-1}, x_t) is passed through the cell states and defined as Eq. (1):

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

where the activation function is denoted by the $\sigma(\bullet)$ and x_t, h_{t-1}, W_f, and b_f represent the input, output at previous LSTM block, weight, and bias at the forget gate layer, respectively, at time t. On the next step, the input gate (i_t), decides what new information is to store in the cell state. The input gate equation is given as Eq. (2):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

where W_i, and b_i represent the weight and bias at the input gate layer, respectively.

For the update, a new state in the cell state, a sigmoid function creates a new vector $\hat{C}_t$ as defined by Eq. (3). Later on, Eqs. (2) and (3) are combined for updating a new state:

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

In order to update the old cell (C_{t-1}) into new (C_t), the old state is multiplied with forget gate and add other parameters as shown in Eq. (4):

$$C_t = f_t * C_{t-1} + C_{t-1} + i_t * \hat{C}_t \tag{4}$$

Finally, the output o_t is given the cell state based on the output of the sigmoid of output gates and defined by:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

where W_o, and b_o represent the weight and bias at the output layer, respectively.

It is observed that unidirectional LSTM network stores the information data from the past direction only. On the other hand, a Bi-LSTM network preserves the information data both direction from past to future and the future to past. Further, the Bi-LSTM network performed much better than LSTM for sequential data applications. The basic architecture of a Bi-LSTM network is depicted in Fig. 4. The output sequence in the forward layer $\vec{h}$ is computed for given inputs in positive sequence from $x_{t-1}\ldots\ldots x_n$. The backward output sequence $\overleftarrow{h}$ is computed for reversed inputs sequences. The final layer output is evaluated with the help of Eqs. (1)–(6) for the backward and forward layer outputs. The bi-directional LSTM layer yields an output vector as: $Y_T = y_{t-1}, y_t, \ldots y_{t+n}$ in which each $y_t = \sigma\left(\vec{h}, \overleftarrow{h}\right)$,

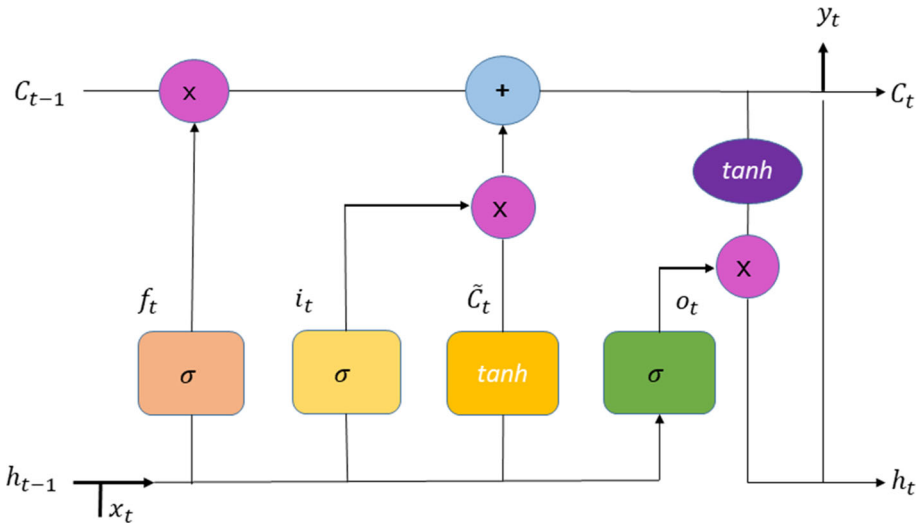


Fig. 3 Basic LSTM architecture

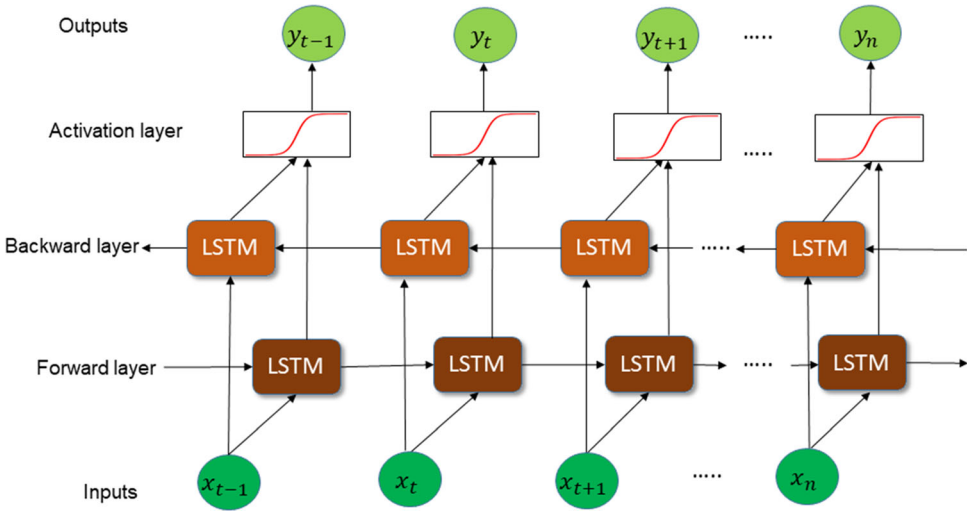


Fig. 4 The bi-directional LSTM architecture

where σ is called a summation function or concatenating function.

It is observed that fine-tuning the all sequential frames related to action labels is not a good idea for learning useful discriminative features. Therefore, in this work, RGB video is sampled out ten frames at a fixed interval of to the pre-trained CNN architecture and trained, excluding the fully connected layer. The upper half of the proposed ConvNet is end-to-end trained with only the last five layers. For each frame, a features vector of dimension $10 \times 4 \times 4 \times 2048$ is obtained from CNN architecture. Further, these extracted features are given to average pooling layer and got a vector of size $10 \times 1 \times 2048$ for each layer. Finally, the feature vectors acquired from the pre-trained model are further fed to the Bi-LSTM layer for better feature extractions. It helps for better generalization in terms of learning features from frames.

In the following subsection, the concept of dynamic motion images and fine-tuning with CNN for temporal extraction features is explained in detail.

3.3 Dynamic motion image (DMI) from video sequences

In this section, we emphasized understanding the long-term temporal dynamics in terms of a single dynamic image. Later on, these images are fed as an input for fine-tuned Inception-v3 architecture for features extraction. It is a paramount task to understand the content of videos precisely on a large scale because videos are consisted of a sequence of still images. Therefore, the summarization of the whole video sequence into a single still dynamic image using standard CNN architecture is introduced in work [9]. A rank pooling mechanism is adopted that utilized the work of Fernando et al. [24] for obtaining a dynamic image from whole video sequences. This technique encrypts the temporal variation cues of the video sequences into one single image. A rank pooling function directly applied to raw RGB frames to produces a single dynamic image for each activity video.

The idea behind the creation of dynamic images depends on the ranking function [24] that ranks its each video frames $(I_1, I_2, I_3, \ldots\ldots, I_T)$ according to the time axis. Let a feature vector represented by $(I_t) \in \mathbb{R}^{\mathbf{d}}, t \in [1, T]$, where $\varphi(\cdot)$ is showed the rank for each frame I_t at time instance t. A score function $\mathcal{S}(t|\mathbf{d}) = \langle \mathbf{d}, \mathcal{V}_t \rangle$, is associated with ranking function, where $\mathbf{d} \in \mathbb{R}^{\mathbf{d}}$ is a vector of parameters and $\mathcal{V}_t = \frac{1}{t}\sum_{\tau=1}^{t} \varphi(I_t)$ is the time average of these features at this instant. According to the RankSVM [49], the learning of vector $\mathbf{d}$ is modelled as a convex optimization problem and given by Eq. (7),

$$\mathbf{d}^* = \rho(I_1, I_2, I_3, \ldots\ldots, I_T; \varphi) = \underbrace{\operatorname{argmin}}_{d} E(\mathbf{d}) \tag{7}$$

$$E(\mathbf{d}) = \frac{\delta}{2} \parallel \mathbf{d} \parallel^2 + \frac{2}{T(T-1)} \times \sum_{q>t} \max\{0, 1 - \mathcal{S}(q|\mathbf{d}) - \mathcal{S}(t|\mathbf{d})\} \tag{8}$$

where $\rho(I_1, I_2, I_3, \ldots\ldots, I_T; \varphi)$ maps a sequence of T number of video frames into a single vector $\mathbf{d}^*$ is called a rank pooling function that aggregates information from all video frames. The first term $\frac{\delta}{2} \parallel \mathbf{d} \parallel^2$ of Eq. (8) is a quadratic regularized function used for SVM. The second term related to soft counting loss that calculates how many pairs are not correctly ranked for $q > t$ by the ranking function. The score function is calculated for video frames based on ranking function, and a pair of the frame is chosen for which score having unit margin i.e. $\mathcal{S}(q|\mathbf{d}) > S(t|\mathbf{d}) + 1$.

The dynamic motion image obtained using approximate rank pooling function [9] is fifty times faster than rank pooling function for similar performance. The approximate

rank pooling mechanism is based on a gradient optimization algorithm and derived using Eqs. (9)–(12) as follows:

For

$$\mathbf{d} = 0,\ \mathbf{d}^* = \vec{0} - \eta \nabla E(\mathbf{d})\Big|_{\mathbf{d}=\vec{0}} \propto -\nabla E(\mathbf{d})|_{\mathbf{d}=\vec{0}}$$

for all $\eta > 0$ where

$$\begin{aligned} \nabla E\left(\vec{0}\right) &\propto \sum_{q>t} \max\{0, 1 - \mathcal{S}(q|\mathbf{d}) - \mathcal{S}(t|\mathbf{d})\}\Big|_{\mathbf{d}=\vec{0}} \\ &= \sum_{q>t} \nabla \mathbf{d}, \\ \mathcal{V}_t - \mathcal{V}_q &= \sum_{q>t} \mathcal{V}_t - \mathcal{V}_q \end{aligned} \tag{9}$$

Further, the function $\mathbf{d}^*$ used as video descriptor because it aggregates information from all stacks frame and defined as:

$$d^* \propto \sum_{q>t} \mathcal{V}_t - \mathcal{V}_q = \sum_{q>t} \left[\frac{1}{q}\sum_{i=1}^{q} \varphi_i - \frac{1}{t}\sum_{j=1}^{t} \varphi_j\right] = \sum_{t=1}^{T} \Omega_t \varphi_t \tag{10}$$

where the Ω_t is denoted as:

$$\Omega_t = 2(T - t + 1) - (T + 1)(\boldsymbol{H}_T - \boldsymbol{H}_{t-1}) \tag{11}$$

where $H_t = \sum_{i=1}^{T} \frac{1}{t}$ is the ith Harmonic number and $H_0 = 0$.

Therefore, the updated approximate rank pooling function is defined as the weighted sum of adjacent video frames as:

$$\hat{\rho}(I_1, I_2, I_3, \ldots\ldots, I_T; \varphi) = \sum_{t=1}^{T} \Omega_t \varphi_t \tag{12}$$

The weights parameter $\Omega_t \in (1, T)$ is calculated for fixed video length accordingly, as shown in Fig. 5.

This demonstration showed that Ω_t depends on the consecutive video frames belongs to frames (1, T). The weight parameter Ω_T is the combination of the sum of all weight parameters obtained from each frame $\frac{2*i-T-1}{i}$, where $i \in (t, T)$. It is observed that DMIs computation is limited only to by pre-multiplying function Ω_t for all video frames. Therefore, approximate rank pooling function does not require to compute the intermediate average features vector $\mathcal{V}_t = \frac{1}{t}\sum_{\tau=1}^{t} \varphi(I_t)$.

Instead, it can be directly calculated by using individual frame feature $\varphi(I_t)$ and $\Omega_t = 2t - T - 1$, which is a linear function of time t. The example of dynamic image formation using the approximate rank pooling technique from 'approaching' activity video is illustrated in Fig. 6. It can be noted that each video frames is multiplied with the corresponding weight, i.e. frame I_1 is multiplied by Ω_1 for each channel individually. The final obtained DMI is the weighted sum of each Red, Green, and Blue channels of each frame and having the same size as the frames. It is observed that the temporal action modelling pattern can be easily seen from a dynamic image when a person is approaching another person irrespective to background and illumination conditions. The so obtained DMIs are fed to pre-trained ConvNet to extracts the temporal features to recognize the action in a video.

3.4 Late fusion

The late fusion techniques for two-stream networks are categorized as Sum Fusion, Maximum fusion and Concatenation Fusion [50]. The main objective of fusing the two-stream network is that features extracted from the same pixel's location through the different channel are combining for better prediction. For example, to differentiate between the activities of 'talking on mobile phone' and 'drinking water through glass' can be recognized easily through the recognition of hand movement pattern by the temporal network at some spatial location. In contrast, the spatial network can identify the location of the head and their combination help to increase the overall prediction accuracy. In the proposed work, these techniques are defined using the scores of a decision level of Inception-Bi-LSTM stream (a) and DMI stream (b).

A general function f for fusing the two feature maps a and $b$at a given time t is denoted by Eq. (13).

$$f : x_t^a, x_t^b \rightarrow y_t \tag{13}$$

where $x_t^a \in \mathbb{R}^{H^a \times W^a \times D^a}$ and $x_t^a \in \mathbb{R}^{H^b \times W^b \times D^b}$ are two different features maps. The output feature map is denoted as: $y_t \in \mathbb{R}^{H \times W \times D}$, where W, H and D are represented width, height and number of the channel of respective feature maps. For simplicity, we assume that $W^a = W^b = W, H^a = H^b = H, D^a = D^b = D$.

The late fused score obtained by these approaches are denoted as y^{sum} y^{sum} (Sum), $y_{\max}$ (Maximum), and y^{cat} (Concatenation).

Sum fusion It calculates the sum $y^{\text{sum}} = f^{\text{sum}}\left(x^a, x^b\right)$ of two features maps in the feature channels d, at the same spatial location (i, j) and expressed using in Eq. (14).

$$y_{i,j,d}^{\text{sum}} = x_{i,j,d}^a + x_{i,j,d}^b \tag{14}$$

where $1 \le i \le H, 1 \le j \le W, 1 \le d \le D$ and $x^a, x^b, y \in \mathbb{R}^{HxWxD}$.

The sum fusion scores y^{sum} show a random correlation between the network layers.

Maximum fusion In this technique, $y^{\max} = f^{\max}\left(x^a, x^b\right)$ the maximum score is selected between the two feature maps. It defined using Eq. (15).

$$
\begin{array}{lccccccc}
 & I_1 & \oplus & I_2 & \oplus & I_3 & \cdots\cdots\cdots & I_{T-1} \oplus I_T \\
\Omega_1 \rightarrow & \frac{2*\mathbf{1}-T-1}{\mathbf{1}} & & \frac{2*\mathbf{2}-T-1}{\mathbf{2}} & & \frac{2*\mathbf{3}-T-1}{\mathbf{3}} & \cdots\cdots\cdots & \frac{2*(\boldsymbol{T}-\mathbf{1})-T-1}{\boldsymbol{T}-\mathbf{1}} \\
 & & & & & \frac{2*(\boldsymbol{T})-T-1}{\boldsymbol{T}} & & \\
\Omega_2 \rightarrow & & & \frac{2*\mathbf{2}-T-1}{\mathbf{2}} & & \frac{2*\mathbf{3}-T-1}{\mathbf{3}} & \cdots\cdots\cdots & \frac{2*(\boldsymbol{T}-\mathbf{1})-T-1}{(\boldsymbol{T}-\mathbf{1})} \\
 & & & & & \frac{2*(\boldsymbol{T})-T-1}{\boldsymbol{T}} & & \\
\Omega_3 \rightarrow & & & & & \frac{2*\mathbf{3}-T-1}{\mathbf{3}} & \cdots\cdots\cdots & \frac{2*(\boldsymbol{T}-\mathbf{1})-T-1}{(\boldsymbol{T}-\mathbf{1})} \\
 & & & & & \frac{2*(\boldsymbol{T})-T-1}{\boldsymbol{T}} & & \\
\vdots & & & & & & & \\
\Omega_{T-1} \rightarrow & & & & & & & \frac{2*(\boldsymbol{T}-\mathbf{1})-T-1}{(\boldsymbol{T}-\mathbf{1})} \\
 & & & & & \frac{2*(\boldsymbol{T})-T-1}{\boldsymbol{T}} & & \\
\Omega_T \rightarrow & & & & & & & \frac{2*(\boldsymbol{T})-T-1}{\boldsymbol{T}}
\end{array}
$$

Fig. 5 The process of calculation of parameter Ω_T for finite length video sequences T. The bold part shows the dependency of parameter Ω_T on consecutive video frames $\in (1, T)$

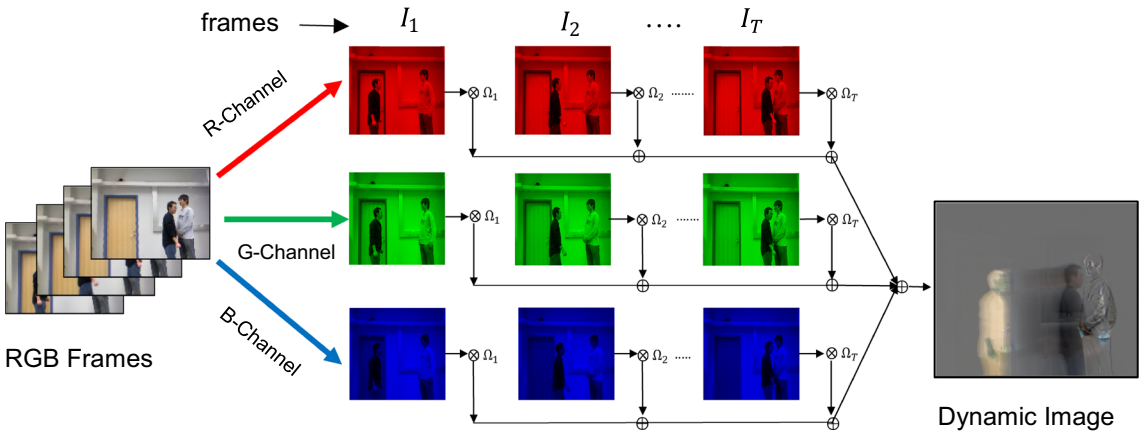


Fig. 6 The formation of dynamic motion image using approximate rank pooling mechanism from red, green and blue channel of each video frame (color figure online)

$$y_{i,j,d}^{\max} = \max\left\{x_{i,j,d}^{a} + x_{i,j,d}^{b}\right\} \tag{15}$$

where all the parameter representations are similar as in Eq. (14).

It increases accuracy as it incorporates best the predictions of both the models and this approach is used in our model.

Concatenation fusion In this fusion method, $y^{\text{cat}} = f^{\text{cat}}\left(x^{a}, x^{b}\right)$ features extracted from both streams are stacked across the feature channel d as:

$$y_{i,j,2d}^{\text{cat}} = x_{i,j,d}^{a} \quad y_{i,j,2d-1}^{\text{cat}} = x_{i,j,d}^{b} \tag{16}$$

where $y_t \in \mathbb{R}^{H\times W\times 2D}$, and the concatenation fusion does not show any correlation between the two feature maps.

4 Implementation details

Keras API deep learning library is used for the implementation of two streams of convolutional networks. This deep network architecture is trained on one NVIDIA GTX 2 GB graphic with 8 GB RAM, GPU machine. In the following section, model descriptions, training, results and comparison with similar state-of-the-art approaches on four standard RGB-D datasets are discussed in detail.

4.1 Datasets

To evaluate the performance of our two-stream network, we use the four standard datasets one is focused on a dyadic activity as SBU Interaction [51], single and human–object interaction activity datasets as MIVIA Action [33], MSR Daily Activity [52] and MSR Action Pairs [53]. The detailed description of these datasets such as a number of actions, actors, and challenges are explained in the following subsection.

4.1.1 The SBU Interaction Dataset

Yun et al. [51] introduced this interaction activity dataset. It is recorded with three different modalities RGB, depth frames, and 3D coordinates with the help of Kinect Sensor. It consists of seven subjects performing eight human–human interaction activities: departing (S1), approaching (S2), hugging (S3), Pushing (S4), kicking (S5), punching (S6), exchanging object (S7), and shaking hands (S8). The sample images from this dataset are depicted in Fig. 7a.

4.1.2 MIVIA Action

Carletti et al. [33] proposed this dataset intending to recognize human activities and human–object interaction in an indoor lab environment. This dataset is recorded with two different modalities RGB and depth frames with the of Kinect sensors. It contains 14 actors (7 females and 7 males) performing 7 activities such as: "drinking" (M1), "sleeping" (M2), "opening a jar" (M3), "sitting" (M4), "interacting with a table" (M5), "stopping" (M6), and "random motion" (M7). The sample images from this dataset are depicted in Fig. 7b.

4.1.3 MSR Daily Activity 3D Dataset

Wang et al. [52] proposed this dataset intending to recognize the daily use of human activity in an indoor room environment. It is recorded with two different modalities: RGB video and depth frames by a Kinect sensor. It contains 10 subjects doing 16 different daily activities: tossing paper (MD1), playing game (MD2), stand up (MD3), playing guitar (MD4), walking (MD5), using laptop (MD6), cheer up (MD7), using vacuum cleaner (MD8), calling on cellphone (MD9), sit still (MD10, drinking (MD11), lay down on sofa (MD12), reading book (MD13), writing on a paper (MD14), sitting down (MD15), and eating (MD16). All such activities are repeated twice for the sitting and standing position. The sample images from this dataset are shown in Fig. 7c.

4.1.4 MSR Action Pairs 3D

Oreifej and Liu [53] introduced this dataset consists of pairs of action videos. It is a challenging dataset because similar activities have the same shape and motion cues such as 'Put down' and 'Put up'. Ten subjects are performing 6 actions pairs: "Pushing and Pulling a chair" (MA1), "Putting and Taking off a backpack" (MA2), "Sticking and Removing a poster" (MA3), "Wearing and Taking off a hat" (MA4), "Lifting and Placing a box" (MA5), "Picking up and Putting down a box" (MA6). Each action was repeated three times by subjects in which 5 subjects are used for testing and 5 for training purposes. The sample image from this dataset is depicted in Fig. 7d.

4.2 Model parameter description and training settings

The proposed hybrid ConvNet is trained for two different data streams, i.e. RGB frames and dynamic image independently. The overfitting problem that occurs due to smaller training datasets in LSTM is compensated with the

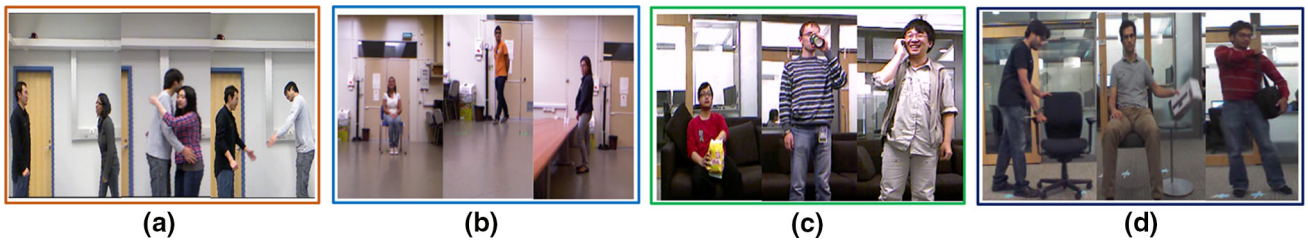


Fig. 7 Sample RGB frames from: **a** SBU Interaction Dataset **b** MIVIA Action Dataset **c** MSR Daily Activity 3D Dataset **d** MSR Action Pairs 3D Dataset

implementation of L2 regularization and dropout mechanism. The CNN-LSTM stream extracted the features from RGB frames that are fed with a batch size of 7 videos. The RGB stream is trained in an end to end fashion up to 150 epochs to refine the features of the pre-trained CNN. Initially, the learning rate of 10^{-4}, and a momentum constant equal to 0.9 is used for training the SGD optimizer. A recurrent dropout of 0.6 is used and added with each Bi-LSTM layer. In the SGD optimizer, the Minimum Square Error (MSE) loss function is selected for calculated the loss during training and test process.

Alternatively, the dynamic motion images are fine-tuned with last fully connected layers on the pre-trained CNN model. This DMIs stream is trained on CNN with a batch size of 8 videos up to 150 epochs in Adam Optimizer [54]. The initial learning rate and various parameters in Adam optimizer are used as: 10^{-4}, $\beta_1 = 0.9, \beta_2 = 0.999$, and $\varepsilon = 10^{-8}$, respectively, and 0.8 of dropout added after every fully connected layer. We used Adam optimizer for DMIs because it is easy to implement, computationally efficient, requires less memory as compared with other optimizers. It requires less tuning and suitable for non-static patterns objective. Figure 8 illustrates the training and test loss curves using MSE loss function for all four datasets. It is observed from the loss curves that our model attained a lower value of square error near 147 epochs.

All the datasets used in the evaluation of the algorithm is multiclass, and hence, we adopted a multiclass classification cross-validation scheme. The leave one out fivefold cross-validation multiclass classification scheme is applied for SBU dataset, where the number of test samples predicted as true class samples are defined as true positives and the test samples predicted as any of other negative classes of the action datasets is considered as false negatives or true negatives. For MIVIA Action Dataset, leave-one-subject-out (LOSO)-CV is applied. There is a total of 14 actors performing the 7 human activities. In this evaluation protocol, 13 actors are used for training and the remaining one actor is for testing. This process is repeating 14 times, always leaving another actor's data for testing. The evaluation method adopted for MSR Daily Activity and MSR Action pairs datasets in which half of the subject used for training and half of the subject for testing. The scores generated by both the softmax layers are fused using some late fusion techniques for prediction the final label activity.

4.3 Results analysis and comparisons

The proposed ConvNet model is tested on four datasets. The evaluation performance is measured in terms of average recognition accuracy (ARA) per class for many classes $(\mathcal{C}_i)$, which is calculated as in Eq. (17),

$$\text{ARA} = \sum_{i=1}^{k} \frac{\frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{k} \tag{17}$$

where tp_i true positive of $(\mathcal{C}_i)$, tn_i true negative of $(\mathcal{C}_i)$, fp_i false positive of $(\mathcal{C}_i)$, fn_i false negative of $(\mathcal{C}_i)$.

4.3.1 The Mann–Whitney *U* test (Wilcoxon Rank Sum Test)

In order to have a better understanding of classification accuracy, a Wilcoxon rank-sum test is used to analyze the result. It is a nonparametric test [55] which is used to test whether two samples are similar in a distribution or not. The null hypothesis that comes from the same observation samples (i.e. have the equal median) or alternatively, whether observation in one sample tends to be greater than observed in the other samples. In the proposed work, we have tested the accuracies of both streams i.e. RGB frames and DMI using nonparametric test. Because both streams extracted the features from the same input video sequences independently. The null and alternative hypotheses for the test are stated:

H_0 = the two independent samples accuracy are the same verses,
H_1 = the two independent samples' accuracy is not the same.

This nonparametric test is conducted as a two-tailed test and observed that populations are not the same as opposed to specifying directionality. The test statistic for Wilcoxon rank-sum test is represented as U and is chosen from a minimum of U_{RGB} and U_{DMI} given by the following equations.

$$U_{\text{RGB}} = n_1 n_2 - \frac{n_1(n_1+1)}{2} - \sum R_1 \tag{18}$$

$$U_{\text{DMI}} = n_1 n_2 - \frac{n_2(n_2+1)}{2} - \sum R_2 \tag{19}$$

where $\sum R_1$, and $\sum R_2$ are the sum of the ranks for RGB Frame samples and DMI samples, respectively, n_1 is the total samples of RGB frame accuracy and n_2 is the total samples of DMI accuracy as illustrated in Tables 1, 2, 3 and 4.

Each time we tested and observed the value of U (U_{Stat}) whether it supports the null or alternative hypothesis like parametric testing. Further, we have determined the critical value ($U_{\text{Cri}(0.05 \text{ or } 0.01)}$) and compared with a minimum value of U_{Stat}. If the critical value is higher, then we reject the null hypothesis H_0 and if the U_{Stat} value is higher than the critical value we reject the alternate hypothesis H_1 i.e.

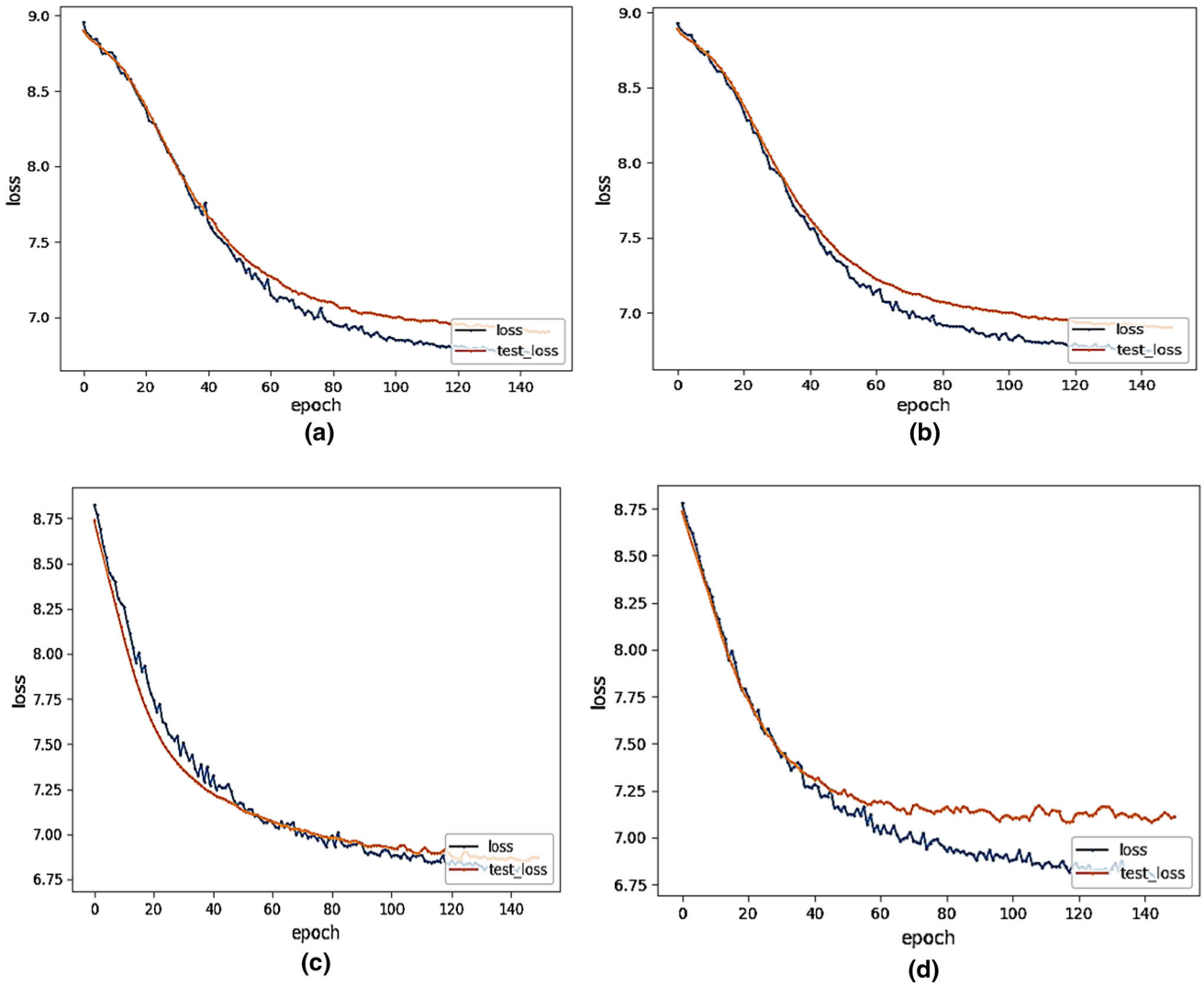


Fig. 8 Shows the training minimum squared (MSE) loss and test loss for activity datasets: **a** SBU Interaction **b** Mivia Action **c** MSR Action Pairs **d** MSR Daily Activity

Table 1 Activity-wise results of RGB frames and DMI streams on SBU Interaction Dataset

Activity	S1	S2	S3	S4	S5	S6	S7	S8
Accuracy (%)								
RGB	82	84	72	71	72	73	76	83
DMI	98	86	95	96	92	94	82	88

Table 2 Activity-wise results of RGB frames and DMI streams on MIVIA Action Dataset

Activity	M1	M2	M3	M4	M5	M6	M7
Accuracy (%)							
RGB	82	74	92	85	76	83	79
DMI	95	94	96	89	94	97	84

$U_{\text{Stat}} = \min(U_{\text{RGB}}, U_{\text{DMI}})$; and if $U_{\text{Stat}} < U_{\text{Cri}(0.05 \text{ or } 0.01)}$, then H_0 (Rejected), and (Accepted).

It can be observed from Table 5 that for each dataset the null hypothesis H_0 is rejected and alternate hypothesis H_1 is accepted because in the value of $U_{\text{Stat}} < U_{\text{Cri}(0.05 \text{ or } 0.01)}$. Therefore, our observation on the samples from independent input streams such as RGB and DMI is not accepted in the same way according to a two-tailed analysis of Wilcoxon rank-sum hypothesis.

The result obtained by various late fusion techniques is compared to four datasets in Table 6. It can be observed that max fusion and sum fusion techniques give high scores as compared to average and concatenation fusion technique. The max fusion yields highest recognition accuracy

Table 3 Activity-wise results of RGB frames and DMI streams on MSR Action Pairs Dataset

Activity	MA1	MA2	MA3	MA4	MA5	MA6
Accuracy (%)						
RGB	91	88	90	87	84	86
DMI	97	96	98	95	87	96

at softmax layers because it selects the maximum probability from both of the softmax prediction scores and assigns a label activity in correspondence to that probability. The highest accuracy is highlighted with a bold letter.

The comparison of average recognition accuracy (ARA) achieved on these datasets with two different input data stream as RGB frames, and single dynamic motion image (DMI) and the max fusion scores of RGB with DMI is shown in Fig. 9. It is clear from Fig. 9, the max fusion of both RGB with DMI gives the best results as compared with independent input data streams: RGB and single dynamic motion image. It can be seen from work Bilen et al. [9] that the dynamic image obtained with approximate rank pooling with CNN showed excellent results in many indoor/outdoor activities recognition tasks in videos. Further, features extracted from RGB only frames are not sufficient to represent the complex activities of human activities. Therefore, the proposed approach fused the discriminative features extracted from DMI and RGB frames to represents spatiotemporal variation in activity video. The proposed model shows excellent results on all four video benchmarks. The results obtained on the MSR Daily activity dataset show somewhat less accuracy as compared with SBU Interaction, MIVIA Action and MSR Action pairs due to intra-class similarity exists between activity classes and complex background.

Table 4 Activity-wise results of RGB frames and DMI streams on MSR Daily Activity

Activity	MD1	MD2	MD3	MD4	MD5	MD6	MD7	MD8
Accuracy (%)								
RGB	81	67	82	80	71	66	82	84
DMI	88	72	90	87	79	72	94	90
Activity	MD9	M10	MD11	MD12	MD13	MD14	MD15	MD16
Accuracy (%)								
RGB	72	73	78	84	80	67	80	79
DMI	80	81	86	91	89	72	88	84

Table 5 Wilcoxon rank-sum test (two-tailed test) research hypothesis results on activity datasets

Dataset	Samples	U_{RGB}	U_{DMI}	U_{Stat}	$U_{Critical}$	
	$n_1(RGB) = n_2(DMI)$				$\alpha = 0.05$	$\alpha = 0.01$
SBU Interaction	8	61.50	2.5	2.5	13	7
Mivia Action	7	46	3	3	8	4
MSR Action Pairs	6	32.5	3.5	3.5	5	2
MSR Daily Activity	16	201	55	55	75	60

Table 6 Comparison of accuracy (%) of different fusion techniques on human activity datasets

Dataset	RGB + DMI late fusion scores			
	Sum fusion	Average fusion	Concatenation fusion	Max fusion
SBU	98.10	96.45	96.50	**98.70**
MIVIA	98.40	97.20	95.75	**99.41**
MSR Action Pair	96.90	94.60	94.80	**98.30**
MSR Daily Activity	**95.36**	90.40	91.60	94.37

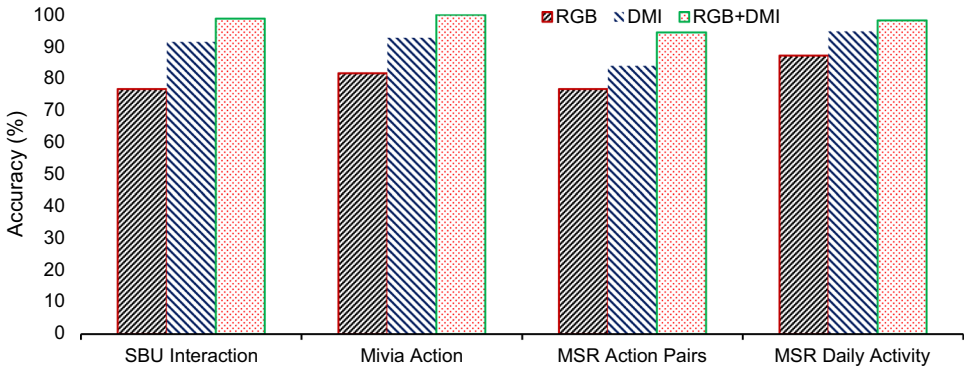


Fig. 9 Results on four datasets with different data inputs: only RGB frames, dynamic motion image (DMI) and RGB + DMI. The accuracy obtained using a max fusion of RGB + DMI outperforms over RGB and DMI inputs

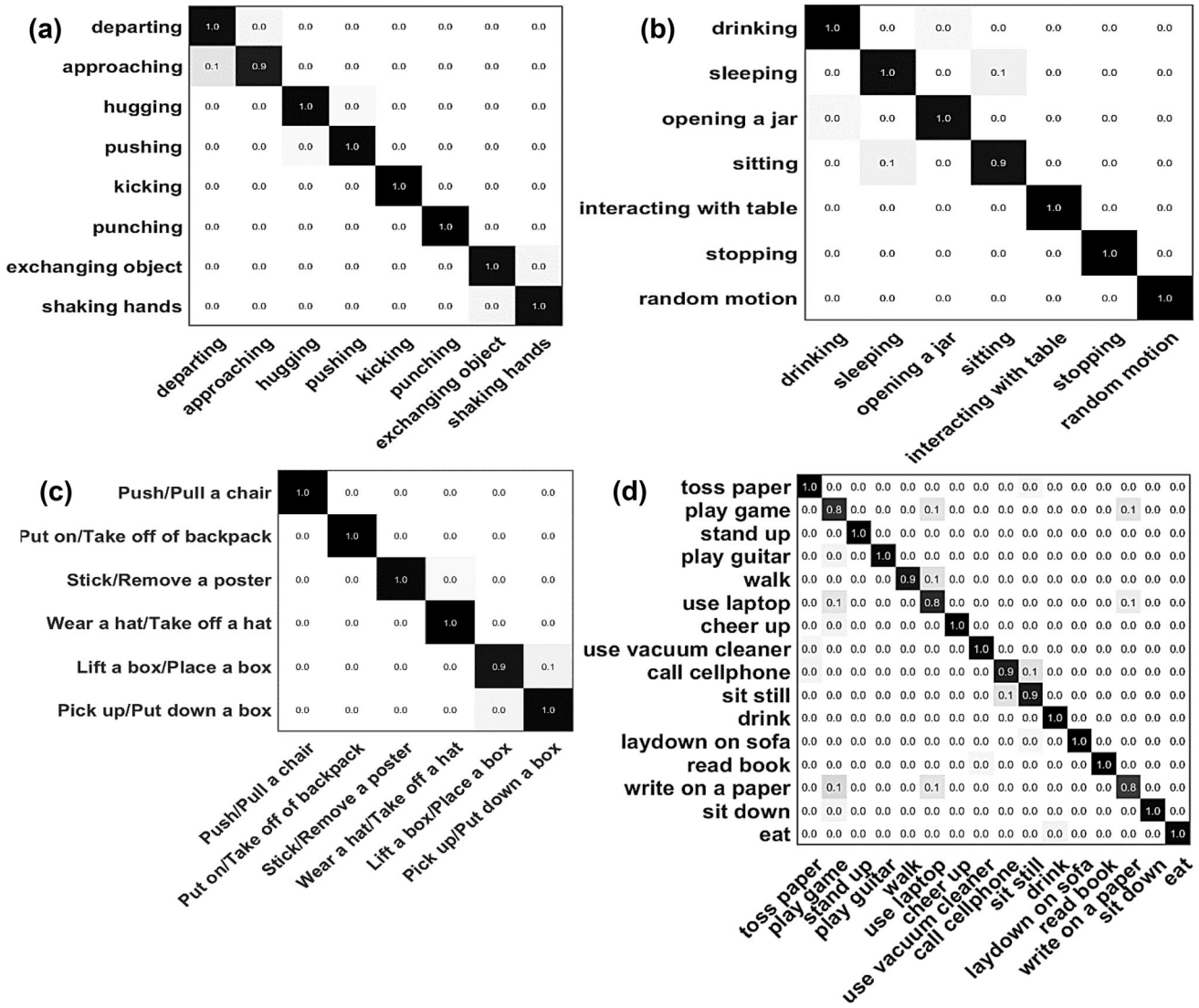


Fig. 10 The confusion matrix of datasets: **a** SBU Interaction, **b** MIVIA Action, **c** MSR Action Pairs, **d** MSR Daily Activity

The classification result on these datasets is shown in the form of a confusion matrix is as shown in Fig. 10. The confusion matrix of SBU dataset is shown in Fig. 10a. It is a challenging dataset because of the similarity of action and the same motion cues in video frames. It is clear from the confusion matrix that the main confusion exists with two similar activities such as "approaching" and "departing" and "handshaking" and "exchanging object." The

Table 7 State-of-the-art comparison of SBU Interaction Dataset

Works	Methods	Accuracy (%)
Yun et al. [51]	SVM and MIL Boost	87.30
Feng et al. [26]	BiVector and LSTM	82.70
Ijjina and Chalavadi [43]	ConvNet + ELM classifier	86.58
Keçeli et al. [42]	2D and 3D ConvNet	94.70
Elboushaki et al. [46]	MultiD-CNN	97.51
Proposed approach (RGB)	Inception-v3 + Bi-LSTM	76.60
Proposed approach (DMI)		91.40
Proposed approach (RGB + DMI) max fusion		**98.70**

Table 8 State-of-the-art comparison of MIVIA Action Dataset

Works	Method	Accuracy (%)
Carletti et al. [33]	Reject	79.80
Foggia et al. [56]	Deep learning	84.70
Brun et al. [57]	Edit distance	85.20
Ijjina and Chalavadi [43]	ConNnets + ELM classifier	93.37
Saggese et al. [34]	Skeleton feature	95.00
Brun et al. [32]	Aclets sequences	95.40
Proposed approach (RGB)	Inception-v3 + Bi-LSTM	81.59
Proposed approach (DMI)		92.71
Proposed approach (RGB + DMI) max fusion		**99.41**

state-of-the-art comparison of similar works on SBU dataset is listed in Table 7. Our deep model is evaluated on SBU dataset with fivefold cross-validation similar in the work [51]. The recognition accuracy of the proposed approach with two-stream fusion shows the best accuracy (bold text) on this dataset as shown in Table 7.

The experimental result of MIVIA action and comparison with similar approaches are shown in Table 8 and the accuracy achieved by proposed approach is highlighted in bold text. The LOSO cross-validation evaluation protocol is followed for training and testing the proposed model similar in [33]. It is observed from the confusion matrix shown in Fig. 10b that actions having no motion in the video found confusion for recognize for example "sitting" and "sleeping" activities. The spatial features for both the activities are the same but lack of temporal cue due to no movement of actors. In this situation, DMI stream features extraction is not so useful for activity prediction. However, the proposed ConvNet model obtained the best results with the fusion of CNN-LSTM and dynamic images as compared with existing methods.

The result comparison of MSR Action Pairs is shown in Table 9 and the accuracy achieved by proposed approach is highlighted in bold text. The evaluation protocol is followed as in similar works by Wang et al. [52]. Out of total performer, half of the actors are used for training the model and half of the actors for testing the action classes. It is observed that from the confusion matrix in Fig. 10c that all the six action pairs are correctly recognized by the proposed approach and no occurrence of intra-class pair confusion. There is slight confusion between "lift a box" and "put down a box" action pairs but our hybrid model

Table 9 State-of-the-art comparison of MSR Action Pairs Dataset

Works	Method	Accuracy (%)
Wang et al. [52]	LOP	82.22
Jia et al. [58]	LTTL	91.40
Oreifej and Liu [53]	HON4D	93.33
Vemulapalli and Chellapa [59]	FTP representation	94.67
Ji et al. [23]	One-shot learning	95.10
Ji et al. [22]	Skeleton embedded feature	97.70
Proposed approach (RGB)	Inception-v3 + Bi-LSTM	87.70
Proposed approach (DMI)		94.76
Proposed approach (RGB + DMI) max fusion		**98.30**

Table 10 State-of-the-art comparison of MSR Daily Activity Dataset

Works	Method	Accuracy (%)
Amor et al. [25]	Rate-invariant analysis	70.00
Seidenari et al. [60]	NBNN bag-of-poses	70.00
Cai et al. [61]	Markov random field	78.20
Zhang and Parker [62]	BIPOD	79.70
Ji et al. [22]	Skeleton embedded feature	81.30
Jing et al. [44]	Joint loss function	88.00
Srihari et al. [45]	4-stream CNN	89.05
Huynh-The et al. [63]	PAM + pose-transition	90.63
Proposed approach (RGB)	Inception-v3 + Bi-LSTM	76.64
Proposed approach (DMI)		83.90
Proposed approach (RGB + DMI) max fusion		**94.37**

shows good performance to recognized each action pairs activity.

The confusion matrix for MSR Daily Activity Dataset is shown in Fig. 10d. It is seen from the confusion matrix that the main confusion occurs in similar activities such as play game using a laptop, and write on the paper. Most of the activities are correctly classified with a high confidence level. As illustrated in Table 10, our proposed model performed well and achieved superior accuracy with similar state-of-the-art approaches, which is highlighted in bold text.

5 Conclusion

In this paper, a robust two-stream deep ConvNet model is developed for the recognition of single, multi-person and human–object interaction (HOI) activities in the video sequence. This model uses two deep learned architectures: First is the pre-trained CNN and Bi-LSTM to extract the discriminative features from the given RGB frames, and second is a pre-trained CNN fine-tuned with fully connected layers fed with dynamic motion image as an input. It is observed that the activities which do not have motion, the CNN-Bi-LSTM combination, classify the activity classes with better recognition accuracy. On the other hand, the activities which have high motion, the dynamic images are used to boost the prediction with the CNN-LSTM stream after late fusion at softmax layer. The prediction accuracy is computed on four publically available video benchmarks such as SBU (98.70%), MIVIA (99.41%), MSR Action Pairs (98.30%), and MSR Daily Activity (94.37%). The comparisons with other state of the art are outlined for the proposed deep architecture proving the dominance of the framework in terms of accuracy.

In future work, depth frames along with 3D skeleton coordinate information and multi-view actions classes may be used to make the action prediction more dynamic to intra-class-life applications. It may also be applied for real-time detection of human activities, and other useful applications such as crowd anomaly detection, sports actions classification, and development of intelligent surveillance system, etc.

Acknowledgements The authors would like to acknowledge the computation of the work was supported by Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, New Delhi, India.

Compliance with ethical standards

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

References

1. Aggarwal JK, Xia L (2013) Human activity recognition from 3D data—a review. Pattern Recognit Lett 48:70–80
2. Dhiman C, Vishwakarma DK (2018) A review of state-of-the-art techniques for abnormal human activity recognition. Eng Appl Artif Intell 77:21–45
3. Suto J, Oniga S, Lung C, Orha I (2018) Comparison of offline and real-time human activity recognition results using machine learning techniques. Neural Comput Appl 1–14
4. Vishwakarma DK, Kapoor R, Maheshwari R, Kapoor V, Raman S (2015) Recognition of abnormal human activity using the changes in orientation of silhouette in key frames. In: IEEE international conference on computing for sustainable global development (INDIACom), New Delhi
5. Schuldt C, Laptev I, Caputo B (2004) Recognizing human actions: a local SVM approach. In: 17th International conference on pattern recognition
6. Vishwakarma DK, Kapoor R (2015) Integrated approach for human action recognition using edge spatial distribution, direction pixel, and R-transform. Adv Robot 29(23):1551–1561
7. Singh T, Vishwakarma DK (2018) Video benchmarks of human action datasets: a review. Artif Intell Rev 52(2):1107–1154
8. Zhang J, Li W, Ogunbona PO, Wang P, Tang C (2016) RGB-D based action recognition datasets: a survey. Pattern Recognit 60:86–105

9. Bilen H, Fernando B, Gavves E, Vedaldi A, Gould S (2016) Dynamic image networks for action recognition. In: IEEE international conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, pp 3034–3042
10. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2015) Rethinking the inception architecture for computer vision, arXiv:1512.00567 [cs.CV]
11. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 1097–1105
12. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
13. Herath S, Harandi M, Porikli F (2017) Going deeper into action recognition: a survey. Image Vis Comput 60:4–21
14. Ladjailia A, Bouchrika I, Merouani H, Harrati N, Mahfouf Z (2019) Human activity recognition via optical flow: decomposing activities into basic actions. Neural Comput Appl 1–14
15. Wang H, Klaeser A, Schmid C, Liu C-L (2013) Dense trajectories and motion boundary descriptors for action recognition. IJCV 103:60–79
16. Liu J, Luo J, Shah M (2009) Recognizing realistic actions from videos "in the Wild". In: IEEE international conference on computer vision and pattern recognition (CVPR)
17. Vishwakarma DK, Singh K (2016) Human activity recognition based on spatial distribution of gradients at sub-levels of average energy silhouette images. IEEE Trans Cogn Dev Syst 99:1
18. Dhiman C, Vishwakarma DK (2019) A robust framework for abnormal human action recognition using R-transform and Zernike moments in depth videos. IEEE Sens J 19(13):5195–5203
19. Baccouche M, Mamalet F, Wolf C, Garcia C, Baskurt A (2011) Sequential deep learning for human action recognition. In: Proceedings of the second international conference on human behavior understanding
20. Ji S, Xu W, Yang M, Yu K (2013) 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231
21. Simonyan K, Zisserman A (2014) Two-stream convolutional networks for action recognition in videos. In: Proceedings of the advances in neural information processing systems
22. Ji X, Cheng J, Feng W, Tao D (2017) Skeleton embedded motion body partition for human action recognition using depth sequences. Sig Process 143:56–68
23. Ji Y, Yang Y, Xu X, Shen HT (2018) One-shot learning based pattern transition map for action early recognition. Sig Process 143:364–370
24. Fernando B, Gavves E, Oramas M, Ghodrati A, Tuytelaars T (2015) Modeling video evolution for action recognition. In: IEEE international conference on computer vision and pattern recognition (CVPR)
25. Amor BB, Su J, Srivastava A (2016) Action recognition using rate-invariant analysis of skeletal shape trajectories. IEEE Trans Pattern Anal Mach Intell 38(1):1–13
26. Feng J, Zhang S, Xiao J (2017) Explorations of skeleton features for LSTM-based action recognition. Multimed Tools Appl 78:591–603
27. Bobick AF, Davis JW (2001) The recognition of human movement using temporal templates. IEEE Trans Pattern Anal Mach Intell 23(3):257–267
28. Blank M, Gorelick L, Shechtman E, Irani M, Basri R (2005) Actions as space-time shapes. In: Tenth IEEE international conference on computer vision (ICCV'05), Beijing
29. Gorelick L, Blank M, Shechtman E, Irani M, Basri R (2007) Actions as space-time shapes. Trans Pattern Anal Mach Intell 29:2247–2253
30. Laptev I (2005) On space-time interest points. Int J Comput Vision 64(2–3):107–123
31. Matikainen P, Hebert M, Sukthankar R (2009) Trajectons: action recognition through the motion analysis of tracked features. In: IEEE 12th international conference on computer vision
32. Brun L, Percannella G, Saggesea A, Vento M (2016) Action recognition by using kernels on aclets sequences. Comput Vis Image Underst 144:3–13
33. Carletti V, Foggia P, Percannella G, Saggese A, Vento M (2013) Recognition of human actions from RGB-D videos using a reject option. In: International workshop on social behaviour analysis
34. Saggese A, Strisciuglio N, Vento M, Petkov N (2018) Learning skeleton representations for human action recognition. Pattern Recognit Lett 118:23–31
35. Dalal N, Triggs B, Schmid C (2006) Human detection using oriented histograms of flow and appearance. In: Proceedings of the European conference on computer vision
36. Laptev I, Lindeberg T (2004) Local descriptors for spatio-temporal recognition. In: ECCV workshop on spatial coherence for visual motion analysis
37. Rodriguez MD, Ahmed J, Shah M (2008) Action MACH: a spatio-temporal maximum average correlation height filter for action recognition. In: IEEE conference on computer vision and pattern recognition, Anchorage, AK
38. Al-Nawashi M, Al-Hazaimeh O, Saraee M (2017) A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. Neural Comput Appl 28:565–572
39. Wang H, Schmid C (2013) Action recognition with improved trajectories. In: Proceedings of the international conference on computer vision (ICCV)
40. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L, (2014) Large-scale video classification with convolutional neural networks. In: IEEE conference on computer vision and pattern recognition, Columbus, OH
41. Peng X, Zou C, Qiao Y, Peng Q (2014) Action recognition with stacked fisher vectors. In: ECCV
42. Keçeli AS, Kaya A, Can AB (2018) Combining 2D and 3D deep models for action recognition with depth information. SIViP 12:1197–1205
43. Ijjina EP, Chalavadi KM (2017) Human action recognition in RGB-D videos using motion sequence information and deep learning. Pattern Recognit 72:504–516
44. Jing C, Wei P, Sun H, Zheng N (2019) Spatiotemporal neural networks for action recognition based on joint loss. Neural Comput Appl 32:4293–4302
45. Srihari D, Kishore PVV, Kumar EK, Kumar A, Kumar MTK, Prasad MVD, Prasad CR (2020) A four-stream ConvNet based on spatial and depth flow for human action classification using RGB-D data. Multimed Tools Appl 79:11723–11746. https://doi.org/10.1007/s11042-019-08588-9
46. Elboushaki A, Hannane R, Afdel K, Koutti L (2020) MultiD-CNN: a multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2019.112829
47. Williams RJ, Hinton GE, Rumelhart DE (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536
48. Hochreiter S, Schnidhuber J (1997) Long short-term memory. Neural Comput 9(1997):1735–1780
49. Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222
50. Feichtenhofer C, Pinz A, Zisserman A (2016) Convolutional two-stream network fusion for video action recognition, arXiv:1604.06573v2 [cs.CV]
51. Yun K, Honorio J, Chattopadhyay D, Berg TL, Samaras D (2012) Two-person interaction detection using body-pose features and

multiple instance learning. In: IEEE international conference computer vision and pattern recognition workshops (CVPRW), Rhode Island
52. Wang J, Liu Z, Wu Y, Yuan J (2012) Mining Actionlet ensemble for action recognition with depth cameras. In: IEEE conference on computer vision and pattern recognition, Rhode Island
53. Oreifej O, Liu Z (2013) HON4D: histogram of oriented 4D normals for activity recognition from depth sequences. In: IEEE international conference on computer vision and pattern recognition (CVPR), Portland, OR
54. Kingma PD, Ba JL (2015) ADAM: a method for stochastic optimization. In: International conference on learning representations, San Diego
55. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18(1):50–60
56. Foggia P, Saggese A, Strisciuglio N, Vento M (2014) Exploiting the deep learning paradigm for recognizing human actions. In: IEEE AVSS
57. Brun L, Foggia P, Saggese A, Vento M (2015) Recognition of human actions using edit distance on aclet strings. In: VISAPP
58. Jia C, Kong Y, Ding Z, Fu Y (2014) Latent tensor transfer learning for RGB-D action recognition. In: Proceedings of the 22nd ACM international conference on multimedia, Orlando, FL, USA
59. Vemulapalli R, Chellapa R (2016) Rolling rotations for recognizing human actions from 3D skeletal data. In: IEEE international conference on computer vision and pattern recognition (CVPR)
60. Seidenari L, Varano V, Berretti S, Bimbo AD, Pala P (2013) Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. In: IEEE international conference on computer vision and pattern recognition (CVPR), Portland
61. Cai X, Zhou W, Wu L, Luo J, Li H (2016) Effective active skeleton representation for low latency human action recognition. IEEE Trans Multimed 18(2):141–154
62. Zhang H, Parker LE (2015) Bio-inspired predictive orientation decomposition of skeleton trajectories for real-time human activity prediction. In: IEEE international conference on robotics and automation (ICRA), Seattle, WA
63. Huynh T-T, Hua C-H, Tu NA, Hur T, Bang J, Kim D, Amin MB, Kang BH, Seung H, Shin S-Y, Kim E-S, Lee S (2018) Hierarchical topic modeling with pose-transition feature for action recognition using 3D skeleton data. Inf Sci 444:20–35

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

2020 IEEE 7th International Conference on Industrial Engineering and Applications

A Multi-Phase Ensemble Model for Long Term Hourly Load Forecasting

Kushagra Bhatia, Rajat Mittal, Nisha, Madan Mohan Tripathi
Department of Electrical Engineering
Delhi Technological University
Delhi, India
e-mail: {kushagra1198, mrajat308, chaharnisha98}@gmail.com, mmtripathi@dce.ac.in

***Abstract*—Long-term projection of electricity demand is necessary for strategizing production, transmission, distribution and grid expansion in power systems. In this work, we propose a model for forecasting hourly profile of load data which must be taken into consideration by power system planners to produce cost optimal and realizable solutions. The developed ensemble model is formulated in two phases, with the initial phase primarily centered on stacking of gradient and adaptive boosting regressors. In the subsequent phase, the variance is diminished by bagging Lasso LARS regressor on the stacked dataset. For implementation of the proposed model, we collect real-world data of the Germany electricity market for thirteen years spanning from 2006 to 2018. Electricity demand forecasts have been evaluated for the duration of five-years from 2014 to 2018 and are found to be extremely accurate as well as consistent. The presented model on comparison with five benchmark load forecasting models is observed to surpass all of them with a mean absolute percentage error of 1.59 on the test set. Furthermore, unlike neural network models, the proposed ensemble is computationally inexpensive with a training time of 110s.**

Keywords-long term load forecasting;extreme gradient boosting;adaptive boosting;bagging;lasso LARS;smart grid

I. INTRODUCTION

The continuous increase in the load demand of consumers and sophistication of power stations has made it extremely important to accurately know in advance the load for a range of commercial, technical and economic benefits. The critical decisions such as load switching, generating and purchasing electric power and infrastructure expansion are made by electric utilities using the load forecasting [1]. Unlike other commodities, the rate of production of electricity must be in balance with the rate of its consumption [2]. The increase in electrification and end-user flexibility has significantly changed the pattern of electricity consumption. The improved load demand forecasts may help regulators in the potential evaluation of flexibility, thereby allowing formation power grid tariffs and yields profitable long-term investments in the power system. However, load forecasting for longer-periods such as five years is a difficult task due to the complex nature of load which may vary depending on the seasons resulting in inaccurate predictions. The revenue of an electric utility industry is highly influenced by the over or under estimation of future load demand [3]. Over-estimation may result into building new power stations in order to meet the false demand while underestimation may produce a shortage in the required load demand making the system insecure and unreliable.Our major contributions in the work are demonstrated:

1. We develop a novel ensemble model for long-term load demand forecasting with the focus of combining the strengths of bagging and boosting.
2. The proposed model is used to forecast hourly load demand for a duration of five year and for each year the model significantly outperforms five state-of-the-art models with regards to accuracy.
3. The model developed in this paper has a computation time of 110s, making it suitable for retraining and is also proven to be consistent with a small value of confidence interval for the five-year period.

An extensive variety of techniques have been implemented for the short-term electric load demand forecasting namely, neural network [4-6], similar day approach [7], statistical learning algorithms [8], fuzzy logic [9] and various regression models varying from days to weeks. However, short-term load forecasting cannot be applied for determining economic location, size, and type of future power plant this is where long-term load demand forecasting [10] becomes essential. The paper [11] has proposed a long-term load forecasting technique for future planning using a fuzzy logic method with a MAPE of 6.9% for Mubi region in Adamawa state. AlRashidi M and El-Naggar K.[12] present a new method based on particle swarm optimization for long-term prediction of annual peak demand and also compares his work with the LES method. Moreover, encouraging and promising results were shown by models based on computational intelligence for instance neural networks, expert systems and regression models [13-15]. Although all the aforementioned models have shown significant and efficient performances, monthly or annular granularity has been considered, making them incapable of generating detailed information for precise long-term investments and power planning [16]. The paper [17] has reviewed current methodologies and future trends for forecasting long-term hourly load demand and recommended various aspects to be considered for forecasting in a rapidly changing power system. Similarly, the model [18] shows the importance of load profile with an hour resolution as an input to the power system models for the long-term estimation of aggregated load. The hourly electricity demand data provided bythe modern smart grids can be beneficial forhighly accurate and reliable long-term demand forecasts.

The primary objective of this work is to explore the ability of ensemble learning for long-term electricity demand

978-1-7281-6785-5/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:17:54 UTC from IEEE Xplore. Restrictions apply.

forecasting. The proposed model exploits the strengths of both boosting and bagging, providing higher stability and therefore, enhancing the model's overall performance. The first phase of the proposed model focuses on boosting where adaptive boosting and two different versions of extreme gradient boosting with decision tree as base estimators are implemented providing flexibility and reducing overfitting respectively. The independent outputs of the aforementioned models are then horizontally stacked, forming a new dataset, to be passed onto the next phase of the ensemble model. In the second phase, the new dataset is bootstrap aggregated usingLasso LARS regression reducing the variance and boosting the overall accuracy. Hyperparameter-tuning of the multi-phase ensemble model is carried out using Random grid search. The proposed model is trained and tested utilizing the real-world demand data of German electricity market via ENTSO-E data portal from the year 2006-2018.

The remainder of this paper has been organized in the following manner. Section II comprises of a detailed description of multi-phase ensemble approach along with the methods used in its formulation. In Section III, the implementation including feature engineering of real-world data, hyperparameter tuning of proposed model and evaluation metrics are explained. Analysis and visual depiction of the developed model along with its comparison to various benchmark models is illustrated in Section IV. The work with its key points is summarized in Section V along with potential future work.

II. METHODOLOGY

The proposed model is formulated in two phases. The first phase comprises of stacked generalization of adaptive boosted trees and different versions of extreme gradient boosted trees. In the subsequent phase, Lars implementation of Lasso regression is bootstrap aggregated on the improvised dataset computed in the previous phase. The load demand dataset $\mathbf{D}=[\mathbf{x},\mathbf{y}]$, where $\mathbf{x}$ is the input feature matrix represented as $\mathbf{x}=(\mathbf{x}_1,\mathbf{x}_2,....\mathbf{x}_N)^{\mathrm{T}}$ and $\mathbf{y}=(y_1,y_2,....,y_{\mathrm{N}})^{\mathrm{T}}$ is the year-ahead load target vector with N training samples.

A. First Phase

1) Adaptive Boosting (AdaBoost): AdaBoost algorithm is an iterative algorithm which focuses on adjusting the learning pattern in accordance to the errors given by weak learners. Initially, use of Adaboost algorithm was emphasized for classification applications [19] but in recent times, it has been commonly applied for forecasting [20]. This algorithm corrects its predecessor by boosting the weights of the underfitted training instances and incorporating this change while fitting subsequent week learners. We have used the AdaBoost.R2 boosting algorithm [21] in the first phase of the proposed ensemble model, where the weight updation step is given by:

$$w_i \rightarrow w_i \beta^{1-L_i} \tag{1}$$

here, L_i is the loss for i^{th} training sample and β is the confidence measure defined as:

$$\beta=\frac{\bar{L}}{1-\bar{L}} \tag{2}$$

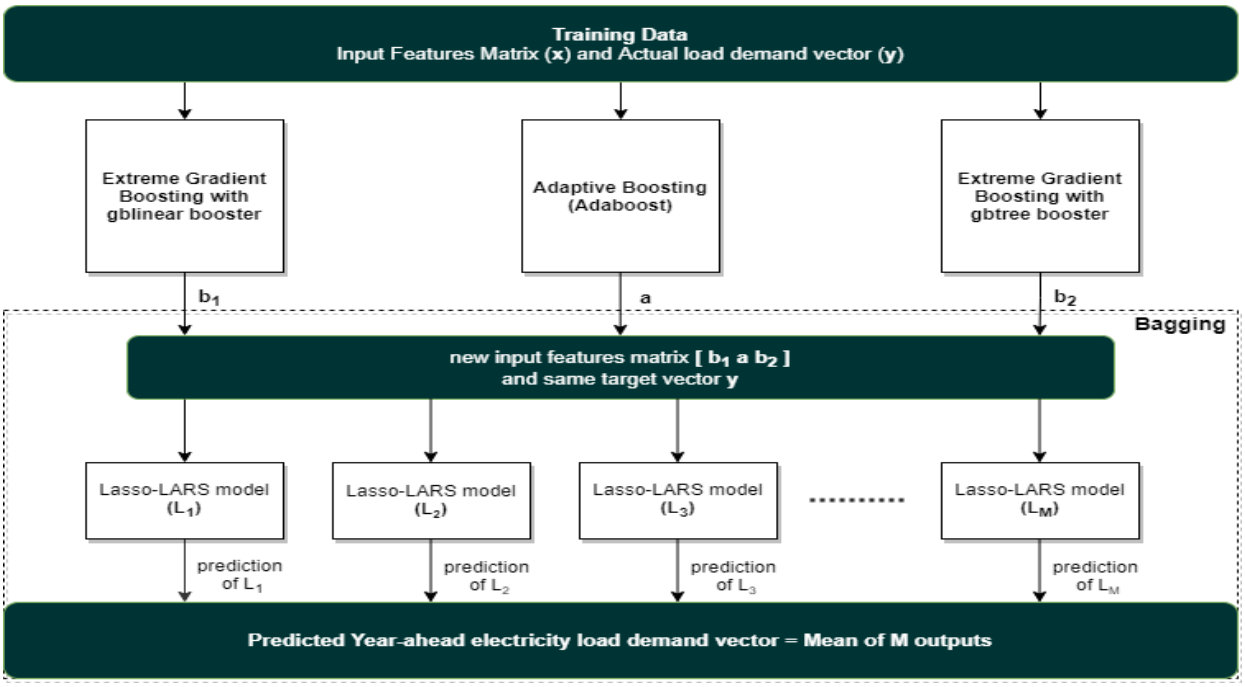


Figure 1.Proposed multi-phase ensemble.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:17:54 UTC from IEEE Xplore. Restrictions apply.

where, $\bar{L}$ is the average loss, if $\bar{L}$ is low, β is also low indicating a high confidence in predictor.The cumulative predictions for dataset $\mathbf{D}=[\mathbf{x},\mathbf{y}]$ is given by:

$$h_f = inf\left\{m \in \mathbf{y}: \sum_{t:h_t \le m} \log(1/\beta_t) \ge \frac{1}{2}\sum_t \log(1/\beta_t)\right\} \quad (3)$$

h_t represents the predictions of T week learners. The primary reason for including AdaBoost in the proposed model is its flexibility to be used in ensemble models, and low number of hyperparameter.

2) eXtreme Gradient Boosting (XGBoost): The efficient parallel and distributed computing capabilities allows XGBoost [22] to be fast and reliable model for long-term electricity demand forecasting. The core idea behind this algorithm is to learn the optimal variables of a particular objective function by utilizing various techniques Given the load forecasting dataset $\mathbf{D}=[\mathbf{x},\mathbf{y}]$, the system response for a data point is approximated using *R* additive functions as:

$$\hat{y}_i = \phi(x_i) = \sum_{r=1}^{R} f_r(\mathbf{x}_i), \quad f_r \in F \quad (4)$$

The space of regression trees in (4) may be defined as:

$$F = \{f(\mathbf{x}) = w_{q(\mathbf{x})}\} \quad (5)$$

here, each f_r corresponds to a tree structure q and leaf weights, represented by w. The objective function of XGBoost is minimized with the goal being error reduction and ensemble tree optimization, given as:

$$L = \sum_{\mathrm{i}=1}^{\mathrm{N}} l(\mathrm{y}_i, \hat{\mathrm{y}}_i) + \sum_{\mathrm{r}=1}^{\mathrm{R}} \Omega(f_{\mathrm{r}}) \quad (6)$$

$$\Omega(f_{\mathrm{r}}) = \gamma \mathrm{T} + \frac{1}{2}\lambda \sum_{\mathrm{j}=1}^{\mathrm{T}} w_{\mathrm{j}}^2 \quad (7)$$

where, l is the convex loss function between actual and predicted load demand values and Ω is the penalization term, defined in (7). The number of leaves in decision tree is given by T and the quadratic term in the (7) corresponds to the L_2 regularization imparted to the final weights. λ and γ are the two-tree complexity controlling parameters.

In our final model we have used two different versions of extreme gradient boosting. The XGBoost library provides different booster options, namely gbtree, gblinear and dart. The parallel combination of independent XGBoost models with gblinear and gbtree boosters, demonstrated a significant improvement as compared to individual XGBoost model in the first phase of the proposed ensemble model. The predictions of the two versions of XGBoost (**b**$_1$ and **b**$_2$) along with predictions of AdaBoost model (**a**) are stacked horizontally and the obtained new dataset called the stacked set S = [**b**$_1$ **a** **b**$_2$] is passed onto the next phase.

B. Second Phase

1) Lasso LARS: Lasso LARS is the implementation of lasso model using the LARS algorithm developed by Bradley Efron [23], where exact solution is yielded which is piecewise linear as a function of its coefficient's norm. The most basic shrinkage operating technique, LASSO (Least Absolute Shrinkage and Selection Operator) diminishes parameters which correlate to zero or close to zero, overcoming multicollinearity [24] and resulting into small variant estimators. LARS (Least Angle Regression) is a model selection algorithm which is less greedy than ordinary least squares but does not converge to a stable result when number of samples is large. The simple modification of removal of non-zero coefficient from active set of predictors when it hits zero and recomputation of the joint direction makes an efficient algorithm for estimating computational Lasso parameters. Lasso LARS is used as a meta-regressor for stacked generalization of the first phase models.

Bagging: The final year ahead load demand forecasts corresponding to the training set are computed by bagging Lasso LARS regression on the stacked set S = [**b**$_1$ **a** **b**$_2$], obtained in the first phase. The M Bootstrap sets constructed by randomly selecting N' sample points from the stacked set are used to fit independent models. Moreover, the points chosen to be included in a bootstrap set are replaced in the stacked set for subsequent random sampling, hence allowing a particular data point to be included in multiple times. In case, $N' = N$, each randomly sampled set is estimated to contain 63.2% of unique points from the stacked set. The final predictions are obtained by averaging the outputs of M Lasso regression models on M bootstrap sets [25]. Overfitting is a commonly observed drawback with stacked generalization ensembles; this limitation is handled by inculcating bagging into the second phase, effectively reducing the overall variance of the model [26]. Moreover, bootstrap aggregation significantly improves the accuracy of load demand predictions.

III. IMPLEMENTATION

A. Dataset and Feature Selection

The data for load forecasting, acquired from ENTSO-E Transparency platform consists of load demand, day-ahead LMP along with weather parameters including wind speed, dew point temperature, dry bulb temperature and precipitation. All the parameters have an hour granularity and are extracted for a period spamming thirteen year i.e. from 2006 to 2018. Feature engineering is used to formulate new features from existing parameters that substantially improve the forecasting model. The final set of selected features broadly consist of time pointers, which have been extracted from the UTC time stamp of each sample point of the dataset, as well as historical load demand features. Weather parameters collected from the platform have not

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:17:54 UTC from IEEE Xplore. Restrictions apply.

been selected as input features. For long-term load forecasting, when forecasts of individual weather parameters are used as input, stockpiling of error leads to deterioration of demand predictions and hence, are not recommended to be included. It must be specified that redundant features are identified on the basis of Extra Tree feature importance and are removed. The final set of input features as specified in Fig. 2 are: previous year same day hourly demand, difference in same day and previous day hourly load demand of previous year and pointers for hour, day, week and month.

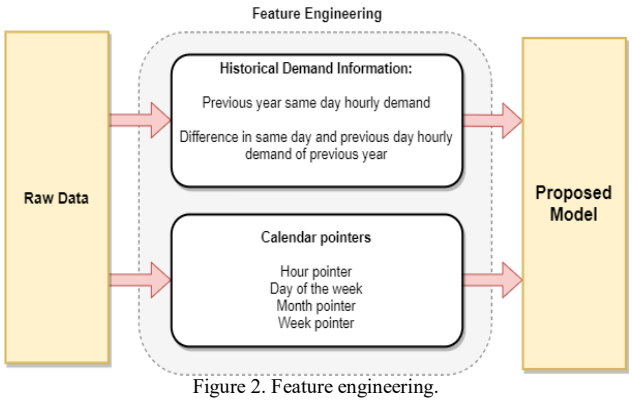


Figure 2. Feature engineering.

To forecast the load demand for the next five years, we use walk-forward prediction approach. In this rolling-based approach, data from 2006 to 2013 is used for training and, predictions are made for the year 2014, which are stored for subsequent use. In the next iteration the stored predicted load values are engineered and utilized for forecasting hourly load demand for year 2015. The same procedure is repeated to obtain the load demand predictions till year 2018.

B. *Hyperparameter Tuning and Feature Scaling*

Hyperparameter tuning has been carried out using random grid search and the tuned parameters are reported in Table I.

TABLE I. HYPERPARAMETERS

Phase	Method	Hyperparameter	Tuned Value
I	XGBoost (gblinear)	Max depth of trees	5
		Learning Rate	0.021
		L_1 regularization term	1e-3
		L_2 regularization term	0.1
I	AdaBoost	Learning Rate	0.12
		Number of estimators	50
I	XGBoost (gbtree)	Max depth of trees	7
		Learning Rate	0.013
		L_1 regularization term	Ie-4
		L_2 regularization term	0.1
II	Lasso LARS	Alpha	1.0
		Number of iterations	200
II	Bagging	Number of estimators	30

Min-Max feature scaling was carried out to scale the features selected in the previous sub-section to [-1, 1] range. The historical selected features have a large value compared to calendar pointers and hence scaling is essential for faster training.

C. *Metric for Evaluation of Models*

The measures used to quantify the accuracy of load demand forecasting models are Mean Absolute Percentage Error (MAPE), Root Mean Squared Percentage Error (RMSPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE). The prior metrics are given in terms of percentage while the latter two are absolute values. The measures are defined as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - p_t'}{y_t}\right| \times 100\% \tag{8}$$

$$\text{RMSPE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - p_t'}{y_t}\right|^2} \tag{9}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - p_t'| \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}|y_t - p_t'|^2} \tag{11}$$

In (8) to (11), n is the number of sample points in test set, y_t and $p_t^{'}$ are the actual and forecasted load demand values. Since the error term is squared in RMSPE and RMSE, a higher penalty is assigned to large inaccurate predictions, differentiating them from theothertwo error metrics.We calculate the confidence interval (CI) [27] of absolute percentage error. For a vector **c**, associated to the absolute percentage error between the actual and predicted demand values in the test set, the confidence interval range is given by:

$$CI = \left[\bar{\mathbf{c}} - z^*\frac{\sigma}{\sqrt{n}}, \bar{\mathbf{c}} + z^*\frac{\sigma}{\sqrt{n}}\right] \tag{12}$$

where, n, $\bar{\mathbf{c}}$ and σ denotes the length, mean and standard deviation of **c**and z^* represents the critical value for thestandard normal distribution.We compute the z^* value by looking up the z-table corresponding to the area under the standard normal curve given by:

$$2\times\left(1-\frac{(1-C)}{2}\right)$$

Here, C represents the specified confidence level. For a confidence of 95% the critical values (z^*) is 1. 96.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:17:54 UTC from IEEE Xplore. Restrictions apply.

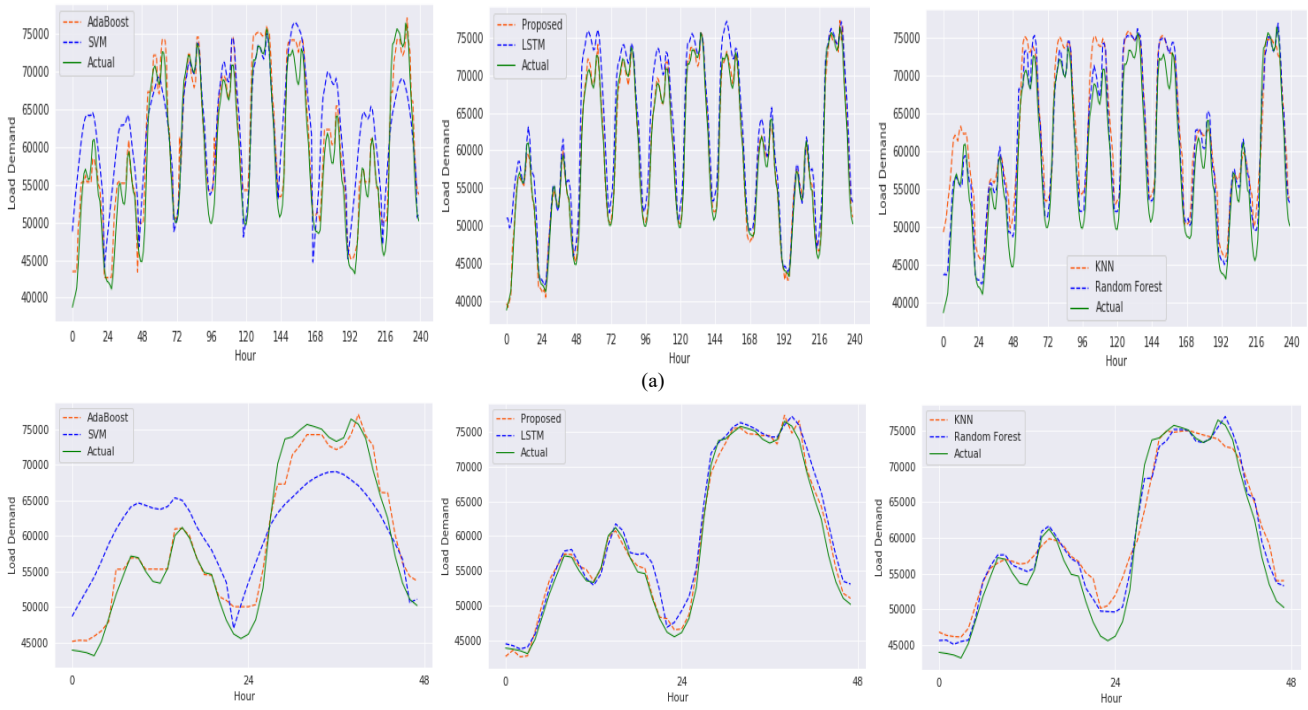


Figure 3. Predicted and actual load forecasts of all models for (a) 10-consecutive days (b) 2-consectutive days of year 2016.

IV. RESULTS

In this section, five of the state-of-the-art models for load demand forecasting including random forest, support vector machine (SVM), k-nearest neighbors (KNN), AdaBoost and LSTM-RNN are compared to the proposed multi-phase ensemble model. Each of the aforementioned models is trained using the same set of features, as explained in Section II. For the entire test set (all five years), MAPE and RMSPE, are calculated using (8) and (9), and are tabulated in Table II for all the models in this study. The superior performance of the proposedensemblemodel asshownin Fig.4, on both percentage metrics validates its suitability for accurate long-term hourly electricity demand forecasting. In order to quantify the accuracy and consistency of load forecasts for individual years, we report the MAPE, RMSPE, MAE and RMSE along with the 95% confidence interval of absolute percentage error for the five-year period, 2014 to 2018 in Table III. The confidence interval bounds indicate that the proposed model's absolute percentage error lies within the specified range for majority of the data points and thepercentage values quantify the precision of the forecasts. For better visualization and performance comparison of the proposed ensemble, we plot the actual and predicted load demand values of all the models for a randomly chosen set of 10 consecutive days from 2016 in Fig. 3(a). Additionally, Fig. 3(b) depicts the predicted vs actual demand values for two consecutive days from the same year. The developed model captures extremities more smoothly and consistently in addition to the periodic drop observed at the 24th hour of each day. Two randomly selected weeks from four seasons namely, winter, spring, summer and fall are used for seasonal evaluation. MAPE of each of the four seasons of 2018, for all the models are reported in Table IV. It is evident from Fig. 5 that the proposed ensemble performs dominantly with the lowest MAPE value for all season.

TABLE II. MAPE AND RMSPE COMPARISION

Model	MAPE	RMSPE
SVM [28]	6.94	9.01
AdaBoost	3.89	5.99
KNN [29]	4.31	6.45
Random Forest [30]	3.48	5.53
LSTM-RNN [16]	3.83	5.78
Proposed Model	1.59	2.54

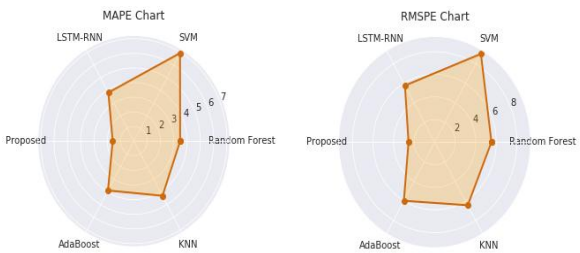


Figure 4.MAPE and RMSE Radar-Chart comparison.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:17:54 UTC from IEEE Xplore. Restrictions apply.

TABLE III. YEAR-WISE EVALUATION OF PROPOSED MODEL FORECASTS

Year	RMSE	RMSPE	MAE	MAPE	CI of absolute percentage error	
					%	***Bound***
2014	1324.3	2.45	879.2	1.54	2.57	[1.502,1.581]
2015	1311.6	2.45	852.6	1.50	2.70	[1.460,1.540]
2016	1305.2	2.43	871.9	1.54	2.55	[1.504,1.582]
2017	1473.1	2.78	956.3	1.69	2.72	[1.648,1.741]
2018	1400.1	2.57	956.4	1.68	2.42	[1.639,1.721]

TABLE IV. SEASON-WISE MAPE COMPARISION

Model	Winter	Spring	Summer	Fall
SVM[28]	6.71	7.71	6.13	6.23
AdaBoost	3.38	4.67	2.36	2.67
KNN[29]	4.09	5.64	2.82	3.23
Random Forest[30]	2.97	4.67	1.93	2.32
LSTM-RNN[16]	3.56	4.69	2.79	3.62
Proposed Model	1.08	2.32	1.32	1.47

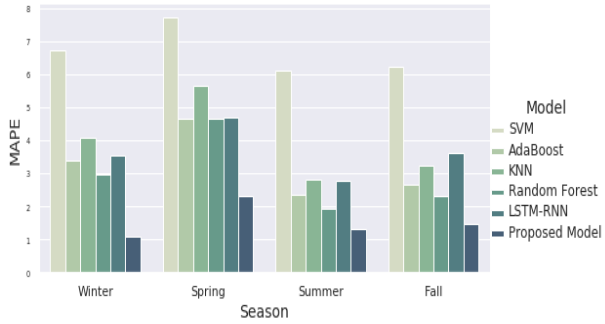


Figure 5.Season-wise MAPE Barplot.

V. CONCLUSION

The paper inquires into the effectiveness of using an ensemble model centered on boosting and bagging techniques. A novel ensemble model for long-term load demand forecasting has been proposed with hourly resolution. The presented model is formulated in two phases where the first phase focuses on boosted trees models, namely, Adaptive boosting and two distinct versions of Extreme gradient boosting. While Adaboost offers flexibility and high degree of precision, the XGBoost with gbtree and gblinear boosters, delivers high performance and accuracy for load forecasting. These boosted trees models were then stack generalized using Lasso-LARS regression which after bagging is used to determine the final prediction of electricity demand. The proposed model has been tested to forecast the load demand of five years for German electricity demand extracted from ENTSO-E Transparency platform. The hourly resolution of load demand predictions can be highly beneficial to the energy service providers, minimizing the risk for planning long-term investments and grid expansion. The over-estimation or under-estimation of future load demand critically affects the revenue of electric utility industry therefore, requiring load forecast models to be accurate and consistent. The proposed ensemble model was benchmarked against five state-of-the-art models including SVM, Lasso, Random Forest, AdaBoost and LSTM-RNN and is proven to be competently accurate and suitable for industrial use with a mean absolute percentage error of1.59%. The presented model is also found to be computationally inexpensive with a training time of 110s seconds. The performance of the developed model can be improved by taking into account the forecasted weather parameters, new technologies for EVs, effects of energy-efficiency measures and renewable energy sources. To utilize meteorological factors into the load forecasting model, it is essential to first develop a reliable method for accurate meteorological forecasts. The proposed ensemble model can also be applied for mid-term and long-term locational marginal price forecasting by including various non-linear factors affecting the electricity price for load demand scheduling in price-directed grids.

REFERENCES

[1] A. K. Singh, Ibraheem, S. Khatoon, M. Muazzam and D. K. Chaturvedi, "Load forecasting techniques and methodologies: A review," Power, Control and Embedded Systems (ICPCES), 2012 2nd International Conference on, Allahabad, 2012, pp. 1-10

[2] H. K. Alfares and M. Nazeeruddin, "Electric Load Forecasting : Literature Survey and Classification of Methods," , vol. 33, no. 1, 2002," Int. Journal of Systems Science, vol. 33, no. 1, pp. 1-7, 2002.

[3] Graven, M., Heyd-Metzuyanim, E.,"Mathematics identity research: the state of the art and future directions," ZDM Mathematics Education 51, 361–377 (2019)

[4] K. B. Sahay, N. Kumar and M. M. Tripathi, "Short-term load forecasting of Ontario Electricity Market by considering the effect of temperature,"2014 6th IEEE Power India International Conference (PIICON), Delhi, 2014, pp. 1-6.

[5] G. M. Din and A. K. Marnerides, "Short term power load forecasting using deep neural networks," in Proc. Int. Conf. Comput., Netw. Commun., 2017, pp. 594–598.

[6] A. Agarwal, A. Ojha, S,C, Tewari, and M.M. Tripathi, "Hourly load and price forecasting using ANN and fourier analysis", In Proceedings of the 2014 6th IEEE Power India International Conference (PIICON), 1-6.

[7] M. Barman, N. B. Dev Choudhury, S. Sutradhar, "A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India," Energy, in press.

[8] A. K. Srivastava, A. S. Pandey, and D. Singh, "Short-term load forecasting methods: A review," in Proc. Int. Conf. Emerg. Trends

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:17:54 UTC from IEEE Xplore. Restrictions apply.

Elect. Electron. Sustain. Energy Syst., Sultanpur, India, 2016, pp. 130–138.

[9] Tehlan, A., Kumar, V., Tehlan, A., et al.: 'Fuzzy logic based load forecasting', Int. J. Appl. Eng. Res., 2016, 11, (9), pp. 6625–6626

[10] Devarajan, Sandhiya and S. Chitra. "Load Forecasting Model for Energy Management System using Elman Neural Network." (2019)

[11] Ali, D.; Yohanna, M.; Puwu, M.I.; Garkida, B.M. "Long-term load forecast modelling using a fuzzy logic approach,". Pac. Sci. Rev. A 2016, 18, 123–127.

[12] AlRashidi M, El-Naggar K.,"Long term electric load forecasting based on particle swarm optimization," Appl Energy 2010;87:320–6

[13] M. S. Kandil, S. M. El-Debeiky, and N. E. Hasanien, "Long-term load forecasting for fast developing utility using a knowledge-based expert system," IEEE Trans. Power Syst., vol. 17, no. 2, pp. 491–496, May 2002

[14] K. Amarasinghe, D. L. Marino, and M. Manic, "Deep neural networks for energy load forecasting," in The 26th IEEE International Symposiumon Industrial Electronics (ISIE), Edinburgh, Scotland, UK, Jun. 19–21 2017

[15] Nalcaci, G.; Ozmen, A.; Weber, G.W. "Long-term load forecasting: Models based on MARS, ANN and LR methods," Cent. Eur. J. Oper. Res. 2019, 27, 1033–1049.

[16] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, "Long term load forecasting with hourly predictions based on long-short-term-memory networks," in IEEE Texas Power and Energy Conference (TPEC). IEEE, 2018, pp. 1–6.

[17] K. B. Lindberg, P. Seljom, H. Madsen, D. Fischer, and M. Korpås, "Longterm electricity load forecasting: Current and future trends," UtilitiesPolicy, vol. 58, pp. 102-119, 2019.

[18] F. Andersen, H. Larsen, T. Boomsma, "Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers," Energy Conversion and Management 68 (2013) 244–252.

[19] P. Viola, M.J. Jones, "Robust real-time object detection," International Journal of Computer Vision 57 (2) (2001) 34-47.

[20] L. Gao, P. Kou, F. Gao, X.H. Guan, "AdaBoost regression algorithm based on classification-type loss," in: 8th World Congresson Intelligent Control and Automation (WCICA), IEEE, 2010, pp. 682-687

[21] Drucker, H. (1997), "Improving regressors using boosting techniques," Proceedings of Fourteenth International Conference on Machine Learning (D. Fisher, Jr., ed.) 107–115. MorganKaufmann, San Francisco.

[22] Tianqi Chen and Carlos Guestrin, "Xgboost: A scalable tree boosting system;" In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785–794. ACM, 2016.

[23] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," Ann. Statist., vol. 32, no. 2, pp. 407–499, 2004.

[24] T. M. Adelheid Januaviani , N. Gusriani , K. Joebaedi , S. Supian , S. Subiyanto, "The Best Model of LASSO With The LARS (Least Angle Regression and Shrinkage) Algorithm Using Mallow's Cp," World Scientific News 2019, pp. 245-252 Available: http://psjd.icm.edu.pl/psjd/element/bwmeta1.element.psjd-bce3d866732e-4ee4-972d-ba9d366b6d0d

[25] J.A. Aslam, R.A. Popa, and R.L. Rivest, "On Estimating the Size and Confidence of a Statistical Audit," Proc. Usenix/Accurate Electronic Voting Technology on Usenix/ Accurate Electronic Voting Technology Workshop (EVT 07), Usenix Assoc., 2007, p. 8.

[26] Bühlmann P, Yu B, "Analyzing Bagging," The Annals of Statistics 2002, 30:927-961.

[27] A. da Silva and L. Moulin, "Confidence intervals for neural network based short-term load forecasting," IEEE Trans. Power Syst., vol. 15, no. 4, pp. 1191–1196, Nov. 2000.

[28] R. Achnata, "Long Term Electric Load Forecasting using Neural Networks and Support Vector Machines," IJCST, Vol. 3, No. 1, 2012.

[29] Mihir R Patel, Ravi V Patel, Dharmesh Dabhi, Jignesh S Patel, "Long Term Electrical Load Forecasting considering temperature effect using Multi-Layer Perceptron Neural Network and k-Nearest Neighbor algorithms", IJRECE, APRIL- JUNE 2019.

[30] Ahmad, T. and H. Chen, "Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems," Sustainable Cities and Society, 2019. 45: p. 460-473.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:17:54 UTC from IEEE Xplore. Restrictions apply.

Applied Intelligence
https://doi.org/10.1007/s10489-020-01649-9

A novel approach for multi-cue feature fusion for robust object tracking

Ashish Kumar[1] · Gurjit Singh Walia[2] · Kapil Sharma[1]

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
Object tracking is a significant problem of computer vision due to challenging environmental variations. Single cue appearance model is not sufficient to handle the variations. To this end, we propose a multi-cue tracking framework in which complementary cues namely, LBP and HOG were exploited to develop a robust appearance model. The proposed feature fusion captures the high-level relationship between the features and diminishes the low-level relationship. Transductive reliability is also integrated at each frame to make tracker adaptive with the changing environment. In addition, K-Means based classifier creates clear and concise boundary between positive and negative fragments which are further used to update the reference dictionary. This adaptation strategy prevents the erroneous updation of the proposed tracker during background clutters, occlusion, and fast motion. Qualitative and quantitative analysis on challenging video sequences from OTB-100 dataset, VOT dataset and UAV123 reveal that the proposed tracker performs favorably against 13 others state-of-the-art trackers.

Keywords Object tracking · Feature fusion · Classifier · Reliability

1 Introduction

Object tracking is an imperative field of computer vision with wide range of applications in video surveillance, human-computer interaction, augmented reality, robotics, and motion analysis. Tracking is detecting the object in the first frame and predicting its state in each subsequent frame of a video sequence. In recent years, a lot of methods have been proposed in this direction but it is still a challenging problem due to dynamic environmental conditions that include pose variations, scale variations, illumination variations, full or partial occlusion, and background clutters. Due to such dynamic environmental variations, single cue is not sufficient to handle the target appearance variations. Hence, multi-cues are necessary for building a robust appearance model.

In object tracking, appearance model can be categorized into generative model and discriminative model. Trackers such as MTT [1], ASLA [2] and FRAG [3] are categorized under the generative model. These trackers track the target by searching for the most similar region. On the other hand, discriminative model considers tracking as a binary classification problem and separates the foreground from the background by training a classifier. WMIL [4], discriminative reverse sparse tracker [5], and CT [6] are categorized as discriminative trackers. Generally, Trackers under these methods combine the various extracted features either at score level or at feature level. Score level fusion combines the classifier scores for different features and exploits these scores to segment the target from the background. Multi-stage tracker [7], CSPF [8] and key-patch based sparse representation tracker [9] considered score level fusion. On the other hand, feature level fusion concatenates individual features vectors obtained from each cue into the final feature vector. Under this, many approaches have been proposed [10, 11]. It has been well argued that feature level fusion preserves more feature level information than score level fusion [12, 13]. However, the obtained final single vector is of high dimension and processing this vector may impact the computational efficiency of the tracker. Also, concatenating the various heterogeneous features extracted from the target into a final robust feature is still a challenge.

In multi-cue object tracking, feature level fusion has been extensively explored. In [14], authors adaptively fused color and depth information along with spatio-temporal consistency constraints for tracking the object. Kang et al. [10] exploited

✉ Kapil Sharma
kapil@ieee.org

[1] Delhi Technological University, Delhi, India

[2] SAG, Defense Research and Development Organization, New Delhi 110042, India

Published online: 07 May 2020

intensity, edge and texture features of the target and fused them using multi-feature fusion through using a Fisher discrimination criterion. Wang et al. [15] fused location, color, and LBP features of the target. In [16], authors considered the thermal profile and grayscale features of the target and fused them using collaborative sparse representation. In [17], authors fused information of two channels and epipolar geometry for target estimation. Meng et al. [18] exploited Hue, Saturation and Value Histogram (HSV) features to cater occlusion and background clutters. In [19], authors integrated color histogram, HSV color, texture, and motion features for real-time tracking. Wang et al. [20] extracted color and HOG features from superpixels of the target and used a random forest regressor to learn the superpixel features. In [21], authors proposed a weighted local sparse model considering the context information and reliability of each local patch. The information was fused using an adaptive update template strategy. Sun et al. [22] proposed a complementary measurement matrix by inducing color and texture features. The ensemble classifier was used to classify the previous and current samples in order to avoid tracking failures. In [23], authors extracted optical flow and gabor features based on contours for the object's appearance model. On the other hand, Walia et al. [11] considered optical flow for coarse estimation of the target. After this, intensity, texture and edge features were exploited in the appearance model for the fine estimation. These features were fused using cross-diffusion mechanism to generate a unified robust feature. In sum, many robust multi-cue tracking solutions considering feature level fusion has been proposed. However, developing a robust non-linear unified feature for an adaptive appearance model catering the environmental challenges can be further explored.

In this paper, we propose a multi-stage tracking framework based on the adaptive fusion of multi-cues. The first stage of target rough localization improves the accuracy of tracker during precise localization. In the appearance model complementary cues namely, LBP and HOG are considered to handle illumination variations and occlusion. K-Means based classifier mechanism and fragment based appearance model are proposed to improve the tracker's accuracy during background clutters and fast motion. Robust unified feature vector is generated by integrating the LBP-HOG and discriminative classifier is to classify the candidate fragments into positive samples and negative samples. Transductive reliability is determined for each of the candidate fragments in order to update the reference dictionary. The main contributions of the paper are as follows:

- We propose a multi-stage tracking framework which includes the coarse to fine estimation of the target. This multi-stage estimation reduces the computational complexity of the tracker. For this, the target is initially roughly localized using optical flow. Optical flow estimates the motion cue for each pixel of the target and hence, improves the accuracy of tracker during precise localization.
- During precise estimation of the target complementary features namely, LBP and HOG are exploited in the appearance model to handle illumination variation, background clutters, and full or partial occlusion. Scale-invariant LBP feature determines local pattern information while HOG feature calculates edge information of the target and hence, is complementary to each other.
- Strength of scale invariant LBP feature vector and HOG feature vector are fused by determining the Euclidean distance between the feature vectors. The distance is determined between the feature descriptors and the samples from the reference dictionary. The proposed fusion not only analyzes the high order relationship between the features but also diminishes the weak relationship among them. At each step, the tracker is made adaptive by integrating transductive reliability for each fragment sample. This reliability ensures that the tracker is adaptive with the change in the environment.
- Discriminative classifier mechanism based on K-Means is proposed to classify the positive samples from the negative samples. Classifier creates a clear decision boundary between the positive fragment samples and negative fragment samples. This process maintains the tracker's accuracy during background clutters and occlusion by preventing the false updation of the tracker.

Further, the paper is organized as follows: Section 2 review the closely related work in the domain. Architecture of the proposed tracker is described in Section 3. In Section 4, we perform the experimental validation of the proposed tracker with qualitative and quantitative comparison against 13 others state-of-the-art. Finally, the conclusion and future directions are detailed in Section 5.

2 Related work

Multi-cue object tracking has been explored a lot over the last decade. Most of the work under multi-cue object tracking framework focus to develop a robust appearance model exploiting either feature level fusion or score level fusion. In this direction, we have reviewed the most relevant appearance model in the area. For more comprehensive details, the interested readers may refer to [24, 25].

Generative model search for the region most analogous to the target. Under this, appearance models were proposed viz. subspace learning [2], patch-based object histogram representation [3], and temporal low-rank structure based [26]. In this direction, Jia [2] et al. proposed a local sparse appearance model which exploits both local and spatial information of the target. Similarity scores were obtained by the pooling

alignment method to handle tracker's drift during occlusion. He et al. [9] represented appearance model of image patches generated by key patch sparse representation. The location of patches was determined by a contribution factor which provided a weighted score to each patch to handle occlusion. Bai et al. [27] developed a sparse representation based online dictionary with an elastic-net and sparsity consistency constraint to construct the target appearance. Robust similarity metric (RSM) was proposed to determine the similarity score between the observed sample and the learned sample. Yang et al. [5] used discriminative scores in an observation model based on sample fragmentation. An adaptive multitask weighted learning scheme was used for building relationships among positive and reliable samples. Despite that, the tracker tended to drift when association between samples was remarkably different. In [26], authors proposed reverse low rank learning algorithm using low rank structure with temporal consistency of the target. Max pooling feature based adaptive weighted scheme was used to localize target effectively. In [13], authors exploited color, texture and shape of the target for the object's appearance model. Sparse representation was used to remove the unreliable features and a kernelized framework was used for feature fusion in the update strategy. Xiao et al. [14] proposed range invariant depth model based on clustered decision tree to adaptively fuse the multi-cue features. In [15], authors exploited covariance matrices for multi-feature fusion. Generally, generative trackers directly fused the multi-cue features or score obtained from the foreground information only and hence, not able to handle background clutters and occlusion.

But, discriminative trackers consider both foreground as well as background features for tracking the target. Discriminative multi-cue information can be fused using score level fusion or feature level fusion. In this direction, Fu et al. [28] proposed structural similarity in order to determine the sample importance with a bag probability function. Classifier scores were used to obtain the final fused vector. Feng et al. [29] constructed weighted map for each frame by incorporating target saliency. Weight map was updated dynamically using the level-set algorithm. In [30], authors proposed low rank regularization with online discriminative learning to differentiate between the target and the background. Update model included the multiple sub-dictionaries to localize the target accurately. Fan et al. [31] proposed active feature selection for discriminating the target from the background. Update strategy includes multi-task learning with selected features to address the tracker's drift during occlusion. In [32], authors proposed structured prior regularizer for tracking object using patch based target representation. The similarity in target's appearance was preserved by a random forest classifier. In [10], authors proposed low-dimensional regional covariance matrices for multi-feature fusion. Extended cam-shift with adaptive kernel bandwidth was used for precise and accurate estimation of the target during environmental challenges. Wu et al. [33] utilized geometric structure of the object using landmark based label propagation (LLP). This geometric structure was explicitly considered by diffusing graph Laplacian regularizer wherein each sample was represented as an undirected graph. However, label prediction of object using landmarks was computationally complex. In [34], multiple visual cues were integrated by feature relearning fusion model. Discriminative feature specific information extracted from visual cues of reliable samples was fused into a robust fused feature. In [35], tensor pooling of local cues were used to build a global representation by capturing the intrinsic structure of the targeted object. Incremental learning of these tensors was performed to build the appearance model. Walia et al. [11] fused the multi-cue features using cross-diffusion strategy. Robust feature vector was generated by diffusing the sparse and dense features iteratively. In sum, feature-level fusion incorporate more high-level fusion information in the unified feature under discriminative multi-cue object tracking.

In another line of research, many deep learning based trackers have been proposed to address the dynamic changes in target appearance. For this, Danelljan et al. [36] proposed to integrate deep RGB and deep motion information with HOG feature by training a CNN model. Evaluation on multiple datasets has shown the superior performance of the proposed method. Cao et al. [37] proposed learning hierarchical features viz. image intensity and CNN features in a spatio-temporal context. Dynamic confidence map were exploited in the update model in a pre-trained deep neural network. Zeng et al. [38] utilized local and holistic information as gray scale and CNN features. Global confidence generated through sparse coding and local similarity through CNN network were integrated in a collaborative network for state estimation. On the other hand, orientation and velocity of the target extracted through motion model and hierarchical features through learned by CNN network [39]. Dynamic weighted motion model was exploited for object's trajectory prediction in a RNN network. However, the requirement of large pre-training data and low frame rate restricted application of deep learning based trackers in real-time.

Recent research indicates the strength of feature level fusion in developing the tracker's appearance model in multi-cue object tracking. Hence, current research should focus on developing a robust appearance model by exploiting the merits of non-linear feature level fusion that can handle the various environmental challenges. The next section details about the methodology of the proposed tracker.

3 Proposed tracker architecture

The proposed multi-stage tracker considers both motion cue and visual cue for localization of target in video sequences. At

stage I, the target is roughly localized by exploiting the optical flow. This provides a much better feature sampling for the next stage of precise localization. During precise localization, candidate fragments are generated around the approximated centroid of the previous stage and complementary multi-cue features namely, LBP (V_e) and HOG (V_h) are extracted. For each candidate fragment, Euclidean distance (E^f) is calculated between each feature descriptor and the corresponding feature of the positive sample ($\mathcal{P}$) and negative samples ($\mathcal{N}$) from the dictionary. The robust unified feature vector ($\mathcal{R}^f$) is generated by averaging the obtained distances which ensures that the high level features are captured well and low-level features are suppressed. Further, the unified feature is averaged to determine the confidence score (μ_f) for each fragment. Figure 1 represents the architecture of the proposed tracker. Initially, a set of positive samples ($\mathcal{P}$) and negative samples ($\mathcal{N}$) are stored in the reference dictionary ($\mathcal{D}_k$). These samples are updated with the reliable samples in the subsequent frames. Next, the fragment obtained by the proposed multi-cue feature fusion are passed through a K-Means classifier to classify them into the positive samples ($C^{\mathcal{P}}$) and the negative samples ($C^{\mathcal{N}}$). At each time step t, transductive reliability ($r(C^f)$) is calculated to assure the consistent updation of the tracker with reliable fragments catering the dynamic environment. The final state (S_t) of the target is obtained by the mean of the high confidence and high reliability fragments. Confidence score and tranductive reliability ensures that the tracker is updated with high confidence score reliable samples only. Also, the reference dictionary ($\mathcal{D}_k$) is updated with the selective replacement of the positive samples and the negative samples. The next section details the first stage of the proposed tracker.

3.1 Stage I: rough localization

In the first stage of rough localization, the target is localized through optical flow. Optical flow considers the rate of movement of the pixel in the respective directions between the consecutive frames. In the proposed method, Horn-Schunck method (HS) is used to determine the optical flow [40]. Optical flow determines the change in pixel intensity $\mathcal{I}(x, y, t)$ at point (x,y) at time t assuming the change in brightness between the consecutive frames is constant and the same has been depicted using (1)

$$\mathcal{I}(x, y, t) = \mathcal{I}_x v_x + \mathcal{I}_y v_y + \mathcal{I}_t \tag{1}$$

where, $v_x = \frac{dx}{dt}$ and $v_y = \frac{dy}{dt}$ represents the change in pixel displacement in x and y direction w.r.t. change in dt respectively. Smoothness constraint is determined by minimizing the error given by (2).

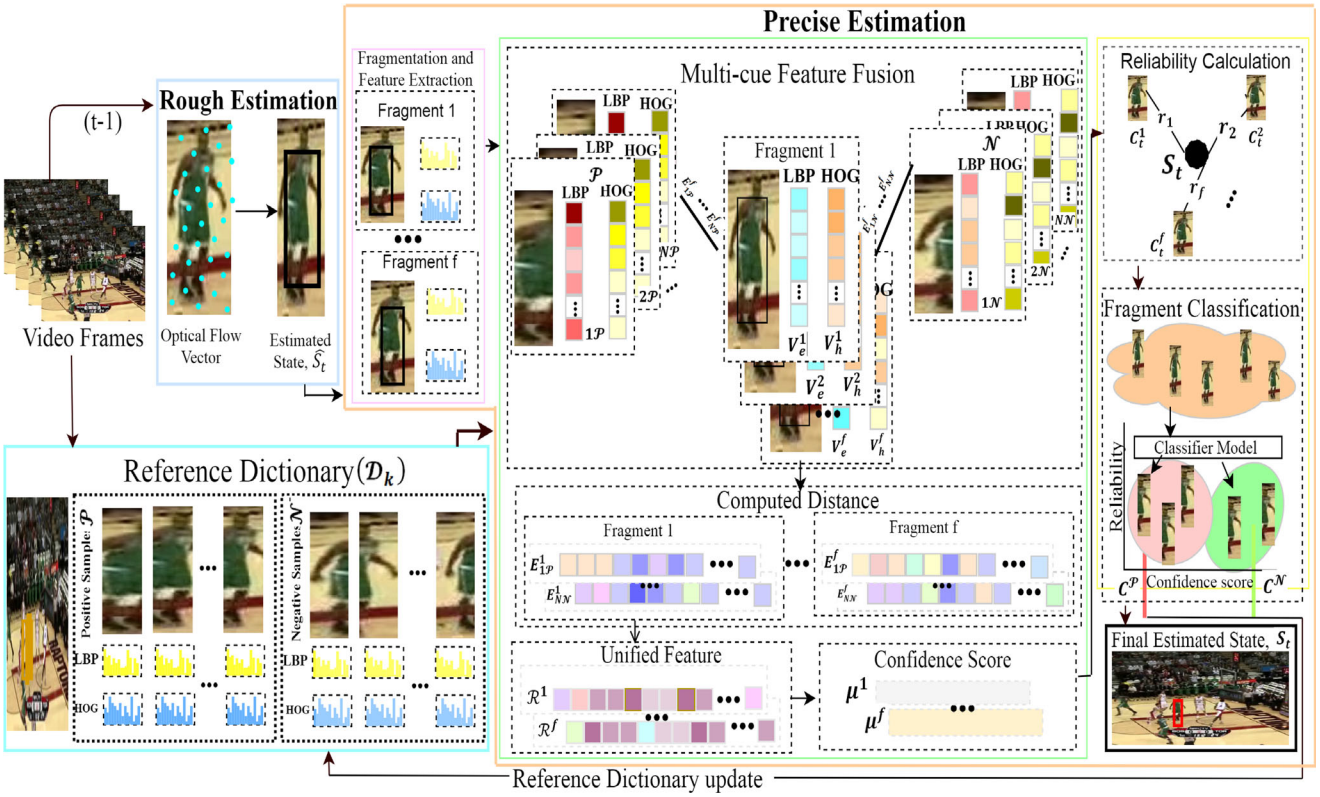


Fig. 1 Overview of the proposed tracker. At $t = 1$, tracker is initialized with positive and negative fragments in the reference dictionary. Stage I, target is roughly localized through optical flow. Then, fragments are generated around the localized target and LBP-HOG features are extracted. Unified feature is generated and transductive reliability is calculated for each localized fragment. Tracker is made adaptive by classifier and transductive reliability

$$\epsilon^2 = \iint \left(\mathcal{I}(x,y,t)^2 + \alpha^2 \left(\left(\frac{\partial v_x}{\partial x}\right)^2 + \left(\frac{\partial v_x}{\partial y}\right)^2 + \left(\frac{\partial v_y}{\partial x}\right)^2 + \left(\frac{\partial v_y}{\partial y}\right)^2 \right) dxdy \quad (2)$$

Where, α^2 scales the second term which is smoothness constraint. If we select N flow vectors around the precise localized centroid $S_{t-1,(x,y)}$ in the previous frame, then the centroid $\widehat{S}_t$ in the current frame is roughly localized and depicted using the (3).

$$\hat{S}_{t,(x,y)} = \left(S_{t-1,x} + \frac{1}{N}\left(\sum_{i=1}^{N} v_x^i \right), S_{t-1,y} + \frac{1}{N}\left(\sum_{i=1}^{N} v_y^i \right) \right) \quad (3)$$

This roughly localized centroid is used for generating the candidate fragments for precise localization of the target. The next section will detail about the precise localization of the target.

3.2 Stage II: precise localization

Precise localization of the target is achieved through multi-cue feature fusion of each candidate fragment. For this, candidate fragments are generated by sampling the target around the centroid $\left(\widehat{S}_t\right)$ determined in stage I using random walk model. Random walk caters for object's scaling and rotational variations. For each fragment, state ($\widehat{S}_t$) is propagated through the random walk model using (4).

$$\hat{S}_{t+1} = \widehat{S}_t + \eta_t \quad (4)$$

Where, $\widehat{S}_t = (c,\tau)^m$ in which c represents the scaling factor and τ is angular displacement for each fragment. m is a covariance matrix used to describe the uncertainty in state vector as $m = diag(\sigma_c,\sigma_\tau)$. η_t is zero mean Gaussian noise through a random walk model. Each localized fragment is scaled and rotated using the random walk model and multi-cue features are then extracted. In the proposed tracker, we have considered complementary cues viz. Texture and Histogram of Gradient(HOG) and their extraction details are as follows.

Texture cue extracts the low level image features with spatial arrangement of pixels in the image plane. Texture cue is robust to scaling and rotational variations. For each fragment, texture cue is extracted using scale-invariant Local binary pattern (LBP) [41].Uniformity LBP is determined considering central pixel (p_c,q_c) and neighborhood pixel (p_o,q_o) in radius γ using (5).

$$\begin{aligned} U\left(LBP_{\gamma,N_b}(p_c,q_c)\right) = & \left|\beta\left(\mathcal{I}_{j-1}\left(p_{j-1},q_{j-1}\right) - \mathcal{I}_c(p_c,q_c)\right)\right. \\ & \left. -\beta(\mathcal{I}_o(p_o,q_o) - \mathcal{I}_c(p_c,q_c))\right| \\ & + \sum_{j=0}^{N_b-1} \left|\beta\left(\mathcal{I}_j\left(p_j,q_j\right) - \mathcal{I}_c(p_c,q_c)\right)\right. \\ & \left. -\beta\left(\mathcal{I}_{j-1}\left(p_{j-1},q_{j-1}\right) - \mathcal{I}_c(p_c,q_c)\right)\right| \end{aligned} \quad (5)$$

$$LBP_{\gamma,N_b}(p_c,q_c) \quad (6)$$

$$= \begin{cases} \sum_{j=0}^{N_b-1} |\beta((\mathcal{I}_j\left(p_j,q_j\right) - \mathcal{I}_c(p_c,q_c))|, & U\left(LBP_{\gamma,N_b}(p_c,q_c)\right) \le 2 \\ j+1, & otherwise \end{cases}$$

Where, $\mathcal{I}_c$ is gray value of central pixel (p_c,q_c) in radius γ, $\mathcal{I}_j$ is gray value of j^{th} pixel (p_j,q_j) and $\mathcal{I}_o$ is gray value of N_b equally spaced neighboring pixels (p_o,q_o). These values are used to determine the bit wise transition in the uniform image pattern for calculating texture uniformity LBP.

Further, the normalized histograms are obtained by mapping the uniformity measure of LBP for each pixel. For each candidate fragment, these histograms are concatenated to obtain a single feature vector as V_e^f.

HOG feature is used to represent the edge information of the target. HOG captures the edge directions w.r.t. distribution of horizontal and vertical gradient intensities. The distribution of gradient intensities are determined by filtering the image with the kernels given by (7).

$$[-1,0,1] \quad \& \quad [-1,0,1]^T \quad (7)$$

If $\mathcal{G}_x(p,q)$ and $\mathcal{G}_y(p,q)$ represents the gradient values in the horizontal and vertical directions respectively. then, the magnitude ($\widehat{M}$) and orientation(θ) for each pixel is calculated using (8) and (9).

$$\widehat{M}_{(p,q)} = \sqrt{(\mathcal{G}_x(p,q))^2 + (\mathcal{G}_y(p,q))^2} \quad (8)$$

$$\theta_{(p,q)} = arctan\left(\frac{\mathcal{G}_y(p,q)}{\mathcal{G}_x(p,q)}\right) \quad (9)$$

Further, ROI(Region of Interest) of the image is divided into 9 rectangular cells and for each pixel in a cell a weighted vote for each edge orientation is calculated. Votes are bilinearly interpolated between the magnitude and the orientation for each bin. Here, cell histogram for each bin is calculated and bin (γ) for each cell is determined using (10).

$$G_h(\gamma) = \sum_{l=1}^{N_s} \left(\widehat{M}_{l,(p,q)} \delta\left(\theta'_{l,(p,q)} - \gamma\right)\right) \quad (10)$$

Where, $\delta(.)$ is Kronecker delta function, $\theta'(p,q)$ is quantized orientation and N_s is the number of pixels in each cell. To address the illumination variations and contrast gradient values are normalized using L2-norm and final concatenated feature vector is obtained by (11).

$$V_h^f = V_h(\gamma)/\sqrt{\left(\sum_{l=1}^{c\times c\times N_b} V_h(l)^2\right) + \aleph^2} \tag{11}$$

Here, c is the number of cells and N_b is the number of bins per cell and $\aleph$ is a constant. For each candidate fragment, HOG feature vector is determined as V_h^f.

3.2.1 Multi-cue feature fusion

For each candidate fragment f, the individual feature vector of LBP (V_e^f) and HOG (V_h^f) are unified to form a robust feature vector ($\mathcal{R}^f$). Initially, a set of positive samples ($\mathcal{P}$) and negative samples ($\mathcal{N}$) are stored in the reference dictionary ($\mathcal{D}_k$). In each frame, for each fragment the distance is calculated between the feature descriptor and the corresponding feature from the positive samples and the negative samples of the reference dictionary. Figure 2 illustrates the calculation of distance between the feature vector and the corresponding feature vector from the positive sample and the negative sample of the reference dictionary.

Further, the Euclidean distance between the i^{th} LBP feature point (a_i) and the corresponding feature point ($a_{i\mathcal{P}}$) of the positive sample and ($a_{i\mathcal{N}}$) the negative sample, respectively from the reference dictionary is depicted in Fig. 2a and calculated using (12), (13).

$$E^f_{e,N(i\mathcal{P})} = (a_i - a_{i\mathcal{P}}), \quad for \quad 1\le i\le m; 1\le\mathcal{P}\le N \tag{12}$$

$$E^f_{e,N(i\mathcal{N})} = (a_i - a_{i\mathcal{N}}), \quad for \quad 1\le i\le m; 1\le\mathcal{N}\le N \tag{13}$$

Here, m is the length of LBP feature vector and N represents the number of positive samples and the negative samples in the reference dictionary. Similarly, distance between j^{th} HOG feature point (b_j) and the corresponding feature point ($b_{j\mathcal{P}}$) of the positive sample and ($b_{j\mathcal{N}}$) the negative sample, respectively from the reference dictionary is depicted in Fig. 2b and calculated using (14), (15).

$$E^f_{h,N(j\mathcal{P})} = \left(b_j - b_{j\mathcal{P}}\right), \quad for \quad 1\le j\le n; 1\le\mathcal{P}\le N \tag{14}$$

$$E^f_{h,N(j\mathcal{N})} = \left(b_j - b_{j\mathcal{N}}\right), \quad for \quad 1\le j\le n; 1\le\mathcal{N}\le N \tag{15}$$

Here, n is the length of HOG feature vector and N represents the number of positive samples and the negative samples in the reference dictionary. Finally, the unified feature vector for f^{th} fragment between the features V_e and V_h is determined using (16).

$$\mathcal{R}^f = \sqrt{\left(E^f_{e,k}\right)^2 + \left(E^f_{h,k}\right)^2} \qquad k\in\mathcal{P},\mathcal{N} \tag{16}$$

The obtained unified feature vector ($\mathcal{R}^f$) will preserve the high-level relationship and suppress the low-level relationship among the feature vectors. This feature vector is averaged to obtain $\bar{\mathcal{R}}^f$. For each fragment, confidence score is calculated using (17).

$$\mu^f = \frac{1}{2\pi\sigma_f} e^{-\frac{(\bar{\mathcal{R}}^f)^2}{2\sigma_f^2}} \tag{17}$$

where, σ_f is the standard deviation of Gaussian noise in calculation process of the unified feature. Further, the obtained confidence scores of each fragments are subjected to K-Means classifier. The classifier provides the value [0,1] to the candidate fragments as per their resemblance into the positive samples and the negative samples during training. For f fragments, f scores are generated as ($\mu^1.\mu^2...\mu^f$), K-Means clustering is used to partition the fragments into K classes as $C_1,C_2...C_K$. If f^{th} fragment belongs to K^{th} cluster then $f \in C_K$ and $C_K = [C_\mathcal{P}, C_\mathcal{N}]$. If $w(C_K)$ is the measure of within-cluster variations then the objective function which is to be minimized is given by (18).

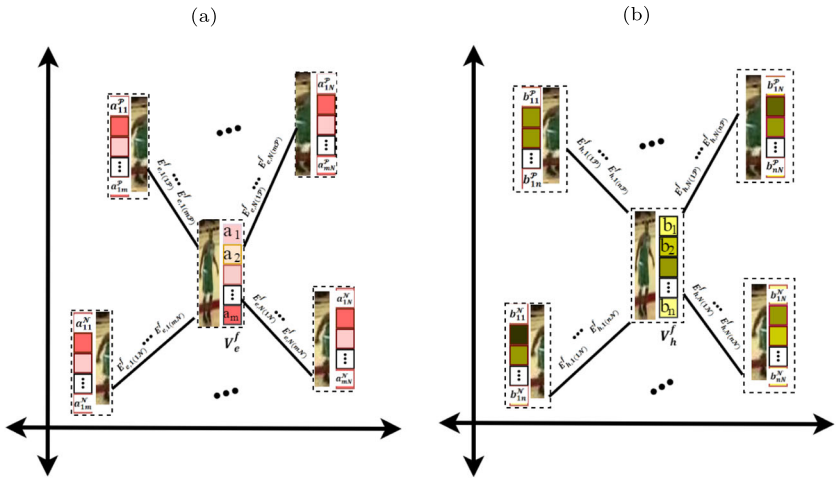


Fig. 2 Description of distance calculation for feature descriptor (**a**) LBP (**b**) HOG

$$argmin \sum_{i=1}^{K} \sum_{i \in C_K} ||w(C_K)|| \tag{18}$$

Here, the within-clusters variations $w(C_K)$ is determined using sum of squares method. For this, pairwise squared deviations of p points within the clusters is defined using (19).

$$w(C_K) = \frac{1}{|C_K|} \sum_{r=1}^{p} \sum_{a,b \in C_K} ||a_r - b_r||^2 \tag{19}$$

Where, $|C_K|$ represents the number of fragments in K^{th} cluster. Using (18) and (19), the K-means clustering is defined using (20), (21).

$$argmin \sum_{i=1}^{K} \frac{1}{|C_K|} \sum_{r=1}^{p} \sum_{a,b \in C_K} ||a_r - b_r||^2 \tag{20}$$

$$argmin \sum_{a_r \in C_K} \sum_{r=1}^{p} ||a_r - \phi_f||^2 = \sum_{a_r \neq b_r \in C_K} \sum_{r=1}^{p} (a_r - \phi_f)(\phi_f - b_r) \tag{21}$$

Where, ϕ_f is mean of p points in cluster C_K. Clusters centroid is determined as the mean of all the fragments score assigned to each cluster. K-Means algorithm determines the local cluster and hence, executed iteratively to determine the nearest centroid. Also, the transductive reliability is calculated for each candidate fragment. The final state is estimated by taking the mean of the fragments with high confidence and high reliability using (22).

$$S_{t,(x,y)} = \sum_{n=1}^{\widehat{N}} (C_n(x,y)) \tag{22}$$

Where, $C_{\widehat{N}}(x,y)$ is the centroid and $\widehat{N}$ is the number of candidate sample with high confidence score and high reliability, respectively. Reliability of each candidate fragment is calculated in order to ensure the tracker is adaptive with change in the environment. Reliability ($r(C^f)$) of the candidate fragment (C^f) can be calculated using (23).

$$r(C^f) = \frac{u\left(tanh\left(-\left(C^f - S_t + v\right)\right)\right)}{2} + w \tag{23}$$

Where, u,v and w are constant to outline the shape of *tanh* function. Further, to improve the performance during long term tracking a re-detection strategy has been employed in the proposed tracker. In case, the tracker drifted away from the target due to the sudden change in the environmental condition, the tracker is re-initialized to re-detect the target. Pseudocode of the proposed tracker is illustrated in Algorithm 1. The experimental validation of the proposed method will follow in turn.

Algorithm 1 The proposed method.

Function `Multi-cue Tracker():`
- Determine optical flow vector $\mathcal{I}(x, y, t)$ using (1)
- Calculate $\hat{S}_t$ using (3)
- **for** *At time step t* **do**
  - **if** $t == 1$ **then**
    - Initialise f fragments around the localised centroid $\hat{S}_t$
    - Initialise $\mathcal{D}_k$ as $\mathcal{P} = N, \mathcal{N} = N$
  - **else**
    - Extract feature vectors V_e, V_h.
    - **for** *I from 1 to N* **do**
      - **for** *i from 1 to m* **do**
        - Calculate $E^f_{e,N(i\mathcal{P})}$ and $E^f_{e,N(i\mathcal{N})}$ using (12), (13)
      - **end**
      - **for** *j from 1 to n* **do**
        - Calculate $E^f_{h,N(j\mathcal{P})}$ and $E^f_{h,N(j\mathcal{N})}$ using (14), (15)
      - **end**
      - Calculate $\mathcal{R}^f$ using (16)
    - **end**
    - **for** *J from 1 to j* **do**
      - Determine μ^f using (17)
    - **end**
    - **for** *Q from 1 to q* **do**
      - $[C_\mathcal{P}, C_\mathcal{N}] \leftarrow K - Means(C_K)$
    - **end**
    - **for** $C_K \in C^f$ **do**
      - Calculate reliablity $r(C_f)$ using (23)
    - **end**
  - **end**
  - Replace $C_\mathcal{P}$ in $\mathcal{P}$, $C_\mathcal{N}$ in $\mathcal{N}$
  - Update classifier with $C_\mathcal{P}$ and $C_\mathcal{N}$
  - Train tracker with $(\mathcal{P}, \mathcal{N})$
  - Update $D_k \leftarrow (C_\mathcal{P}, C_\mathcal{N})$
- **end**
- **return** $S_t, \mathcal{P}, \mathcal{N}$

4 Experimental validation

For rigorous performance analysis, the proposed tracker is evaluated on challenging video sequences taken from OTB-100 dataset [42], VOT dataset [43] and UAV123 dataset [44]. The considered video sequences include various environmental challenges viz. illumination variation, full or partial occlusion, background clutters, fast motion and deformation. The experimental results are compared against 13 others state-of-the-art trackers namely ASLA [2], STRUCK [28], MTT [1],

IDCT [45], DFT [46], L1-APG [47], CT [6], WMIL [4], fDSST [48], CSPF [8], BIT [49], STAPLE [50] and DCF-CA [51]. Publicly available author's release code of these trackers is used for fair performance comparison. The proposed tracker is implemented on MATLAB 2018 on a 2.4GHz processor machine with 6 GB RAM. Initially, the fragments are initialized on the target with minimal overlap. The number of fragments initialized will highly depend on the size of the target. Small and low resolution target requires fewer fragment during initialization in comparison to high resolution target. Also, The reference dictionary is updated with $\mathcal{P} = 20$ and $\mathcal{N} = 20$. Algorithm 1 will illustrates the pseudocode of the proposed tracker. To prove the robustness of our tracker during various environmental challenges the qualitative and quantitative analysis is described in the following sections.

4.1 Qualitative analysis

In this section, the qualitative analysis of the proposed tracker is performed. The details of video sequence and the considered challenge is as follows.

4.1.1 Illumination Variation

Sample frames under the illumination variation challenge are shown in Fig. 3. When there is sudden change in illumination at frame #48 for *Human8* sequence, the proposed approach and STAPLE successfully track the target. This is due to the HOG feature descriptor in the appearance model of the proposed model which compensate for the LBP feature which is susceptible to illumination variation. Ours, DFT, WMIL, BIT and STAPLE track the target with minimal error in

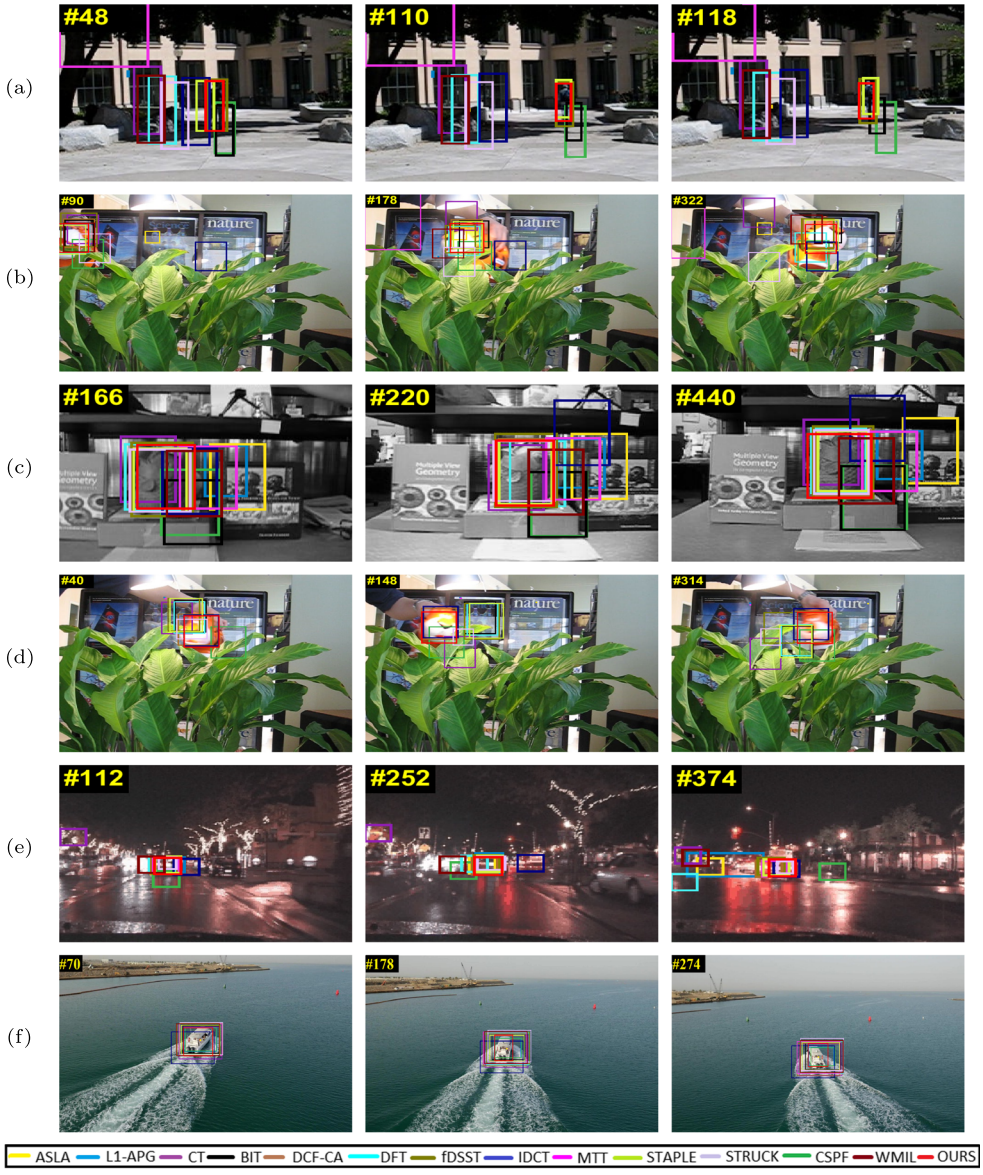


Fig. 3 Critical tracking frames from video sequence under Illumination variation challenge: **a** *Human8* **b** *Tiger* **c***Fish* **d** *Tiger1* **e** *Car2* **f** *boat5*

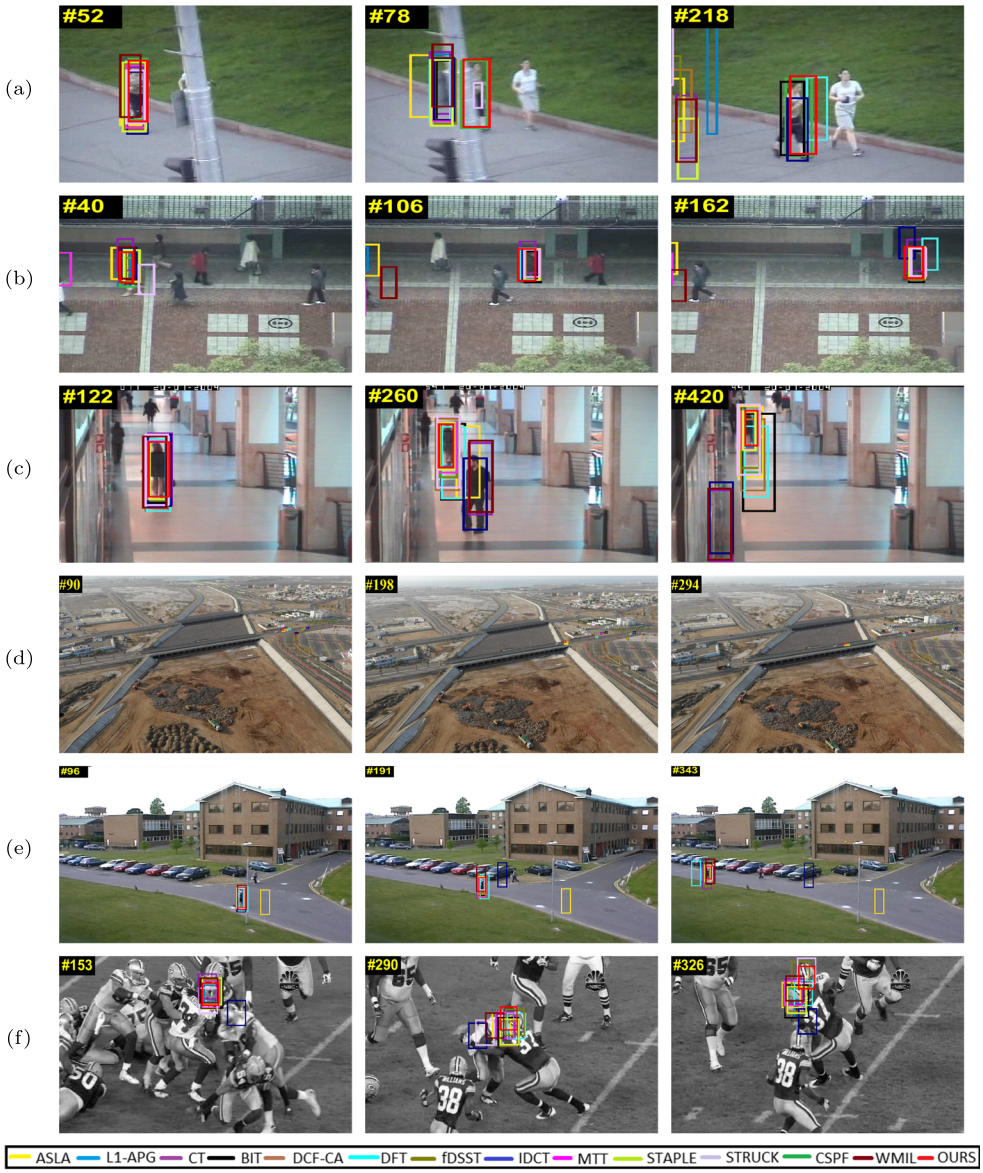


Fig. 4 Critical tracking frames from video sequence under Full or Partial Occlusion challenge: **a** *Jogging1* **b** *Subway***c** *Walking2* **d** *car15* **e** *Walking* **f** *Football*

comparison to other trackers as shown in Fig. 3b. Here, the transductive reliability with proposed fusion of multi-cue will suppress the unreliable features during tracking and prevent drift. In *Fish* sequence, OURS, STAPLE, STRUCK and DCF-CA keep track of the target successfully and same is depicted in Fig. 3c. In Fig. 3d, when there is illumination variations alongwith In-plane rotation in frames (#148 and #314), then only the proposed method with fDSST and STAPLE has shown relatively good performance. It is due to the exploited HOG and the rotation-invariant LBP features in the tracker which address the challenge gracefully. In *Car2* sequence, OURS, MTT and IDCT has performed substantially better in comparison to L1-APG, ASLA, WMIL, DFT and CSPF. Figure 3f depicts that ASLA, OURS, DCF-CA and STAPLE perform better in comparison to others. This is due to the robust unified feature that has been constructed by the fusion of multi-cue in the proposed tracker. Moreover, the complementary features LBP and HOG in the appearance model of the tracker improve the performance under the challenge.

4.1.2 Full or partial occlusion

Representative tracking results for *Jogging1* video sequence under full or partial occlusion challenge are shown in Fig. 4a. At frame #78, when the target is recovered from full occlusion by the pillar, then OURS, IDCT, BIT and CSPF locate the target successfully. This is due to the classifier mechanism which eliminate the non-reliable fragments from the tracker. In *Subway* sequence, target is occluded by various person several times. At frame #106, Generative trackers viz. ASLA, MTT and L1-APG lost the trackers as their appearance model lack drift alleviation mechanism. However, at frame

Fig. 5 Critical tracking frames from video sequence under Background Clutters challenge: **a** *Singer2* **b***MountainBike* **c** *Shaking* **d** *Bolt2* **e** *bike2* **f***Soccer1*

#162 trackers fDSST, STRUCK, BIT, STAPLE and OURS has perform considerably better. Here, the LBP feature compensates for the HOG feature for developing robust appearance model to handle the challenge. Some tracking frames for *Walking2* sequence are depicted in Fig. 4c. In this sequence, target is partially occluded by the similar object. At frame #260 IDCT, ASLA and STRUCK lose the target and locate the other object in the scene. While OURS, STRUCK, MTT and L1-APG keep track of the target till end. This is due to complementary LBP-HOG features are exploited in the appearance model. In addition, the proposed classifier mechanism also prevents the false updation of the reference dictionary. In sequence *Walking*, ASLA, IDCT and DFT lose the target while the rest of the trackers track the target with minimal error. In sequence *car15*, the target size is small, which OURS, STAPLE and STRUCK able to track while other trackers have deflected from the target. Tracking results for *Football* sequence are shown in Fig. 4f. In subsequent frames (#290 and #326), when target is occluded by other players OURS, fDSST and DFT shown relatively good performance in comparison to STRUCK, IDCT, L1-APG, STAPLE and DCF-CA. The performance of the proposed approach under the challenge is mainly attributed due to the generated unified robust feature. Also, the proposed classifier mechanism which classify the affected fragment during occlusion to prevent the drift of the tracker.

4.1.3 Background clutters

Tracking results of the proposed tracker under this challenge is depicted in Fig. 5. Sample frames for video sequence *Singer2* are shown in Fig. 5a. At frame #280, OURS with STAPLE and DCF-CA has handled the challenge efficiently in comparison to other trackers. Here, the exploited features descriptor in the appearance model supported the performance of the proposed approach under the challenge. In *MountainBike*

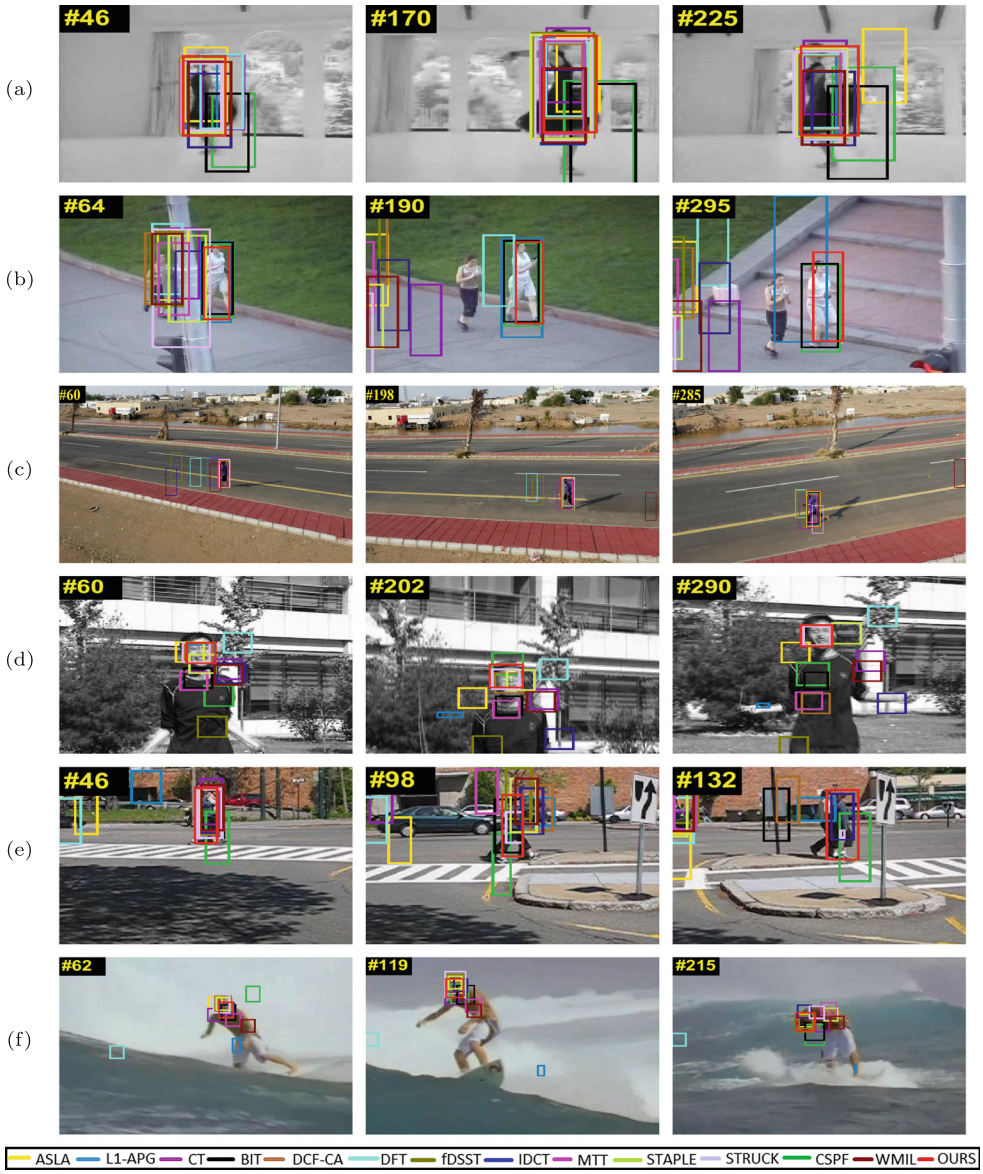


Fig. 6 Critical tracking frames from video sequence under Fast Motion and Deformation challenge: **a** *Dancer* **b***Jogging2* **c** *person23* **d** *Jumping* **e** *Pedestrian1* **f** *Surfer*

sequence, at frame #210 STRUCK, L1-APG and CT lose the target whilst OURS, BIT and MTT keep track of the target in the whole sequence. *Shaking* sequence has similar background with illumination variations in the whole sequence. CT, MTT, CSPF and DFT failed to handle the challenge and hence, drift away from the target. However, OURS, DCF-CA, BIT and WMIL perform considerably better in the sequence. Figure 5d, at frame # 270 OURS and STAPLE track the target while the other trackers locate the other similar object in the scene. Critical frames for sequence *bike2* are depicted in Fig. 5e, in which OURS has shown relatively good performance in comparison to other trackers. In *Soccer1* sequence, DCF-CA perform better with marginal less error in comparison to the proposed tracker. It may be due to false updation of the proposed tracker by the unreliable fragment during the challenge. In sum, the proposed approach able to handle the similar background as the proposed fusion of the multi-cue suppress the low level features and boost the high level features accompanied by the transductive reliability at each stage. In addition, the proposed classifier mechanism will replace the unreliable fragments and update the reference dictionary in order to maintain the performance of tracker during the challenge.

4.1.4 Fast motion and deformation

Critical tracking results illustrating the performance of the proposed approach under the challenge are shown in Fig. 6. In *Dancer* sequence, the deformation of the target due to fast motion which is accompanied by pose variations. ASLA lose the target but OURS with other trackers has shown better performance. Figure 6b presents the tracking results for

Table 1 Centre Location Error (in pixels)

Video sequence	ASLA [2]	STRUCK [28]	MTT [1]	DFT [46]	L1-APG [47]	STAPLE [50]	WMIL [4]	DCF-CA [51]	CT [6]	fDSST [48]	IDCT [45]	CSPF [8]	BIT [49]	OURS
Walking2	19.98	4.21	3.96	31.70	3.99	3.79	59.90	22.07	63.20	10.67	64.78	7.21	28.77	3.73
Jogging1	97.36	123.00	107.90	34.86	88.47	91.20	94.37	89.68	92.95	102.65	15.65	12.38	3.62	7.43
Car2	20.32	3.41	1.35	31.51	12.68	1.35	42.54	5.25	102.98	3.15	21.32	23.87	5.56	5.03
Jogging2	143.90	162.08	152.78	44.01	11.59	146.76	141.62	148.57	116.98	155.17	123.06	8.07	2.84	10.11
Shaking	19.98	107.36	188.94	103.06	122.16	137.60	11.01	8.01	162.88	7.57	134.62	47.91	7.33	5.67
Singer2	193.91	166.91	197.55	39.58	182.60	7.57	160.98	10.82	121.21	12.25	27.38	14.85	180.55	9.43
Tiger	62.61	66.83	204.21	11.57	119.75	11.57	28.31	13.27	70.27	11.04	84.41	56.7	24.9	11.53
Jumping	34.13	8.31	58.89	69.40	77.69	24.82	53.84	2.94	50.85	125.46	79.40	33.99	41.43	17.26
Human8	65.99	56.44	162.03	71.05	112.73	5.77	79.25	3.02	82.64	2.55	43.13	5.58	33.61	4.02
Dancer	23.11	9.37	7.94	12.35	9.04	6.34	17.63	7.29	23.21	6.75	21.21	71.2	66.8	5.64
Pedestrian1	96.23	7.61	55.39	119.69	22.92	34.65	39.78	20.11	79.30	40.01	23.63	15.22	18.72	7.58
Soccer1	121.81	110.59	36.84	112.41	61.76	65.80	84.34	21.65	80.62	11.45	87.96	40.88	7.12	5.70
Football	15.30	13.37	12.39	9.93	17.55	13.02	14.51	13.96	15.63	6.41	70.14	10.62	14.55	9.71
Subway	145.04	4.01	165.30	5.88	149.41	2.59	136.73	2.92	11.58	2.81	8.45	4.97	3.87	2.44
MountainBike	21.73	12.36	5.62	10.30	171.98	9.03	7.11	8.20	192.44	8.59	15.43	14.89	9.54	7.06
Fish	70.91	5.98	43.88	8.51	57.35	4.29	30.99	4.51	13.06	3.59	79.34	54.08	54.83	4.26
car15	-	3.53	-	-	-	1.93	217.7	-	15.51	211.83	-	8.49	221.57	6.34
Tiger1	96.71	76.84	110.77	73.56	119.51	56.34	-	13.56	55.07	62.97	61.90	65.41	54.99	17.99
Surfer	10.56	9.78	44.31	215.73	139.05	27.39	63.74	4.98	45.58	4.11	47.36	36.96	24.01	5.25
bike2	150.65	104.8	138.67	-	153.4	206.82	284	-	-	153.22	209.31	67.08	176.43	29.89
boat5	14.81	28.67	25.36	29.11	23.84	13.67	26.48	13.74	29.11	29.08	24.85	27.69	29.52	17.89
Walking	247.76	3.13	3.41	14.29	2.73	2.09	8.63	2.94	8.55	2.09	116.83	7.35	4.27	2.70
Bolt2	102.92	-	110.46	15.20	-	6.96	119.07	-	13.03	-	150.00	9.08	36.23	11.20
person23	357.73	52.26	6.82	91.98	15.88	8.06	410.14	8.73	41.15	153.83	85.69	6.12	7.71	5.15
Average	**92.76**	**49.60**	**80.21**	**52.53**	**76.19**	**37.06**	**92.72**	**20.30**	**64.69**	**49.01**	**69.38**	**27.11**	**44.12**	**8.88**

Best results are shown in green, and red, blue respectively

Table 2 F-Measure

Video sequence	ASLA [2]	STRUCK [28]	MTT [1]	DFT [46]	L1-APG [47]	STAPLE [50]	WMIL [4]	DCF-CA [51]	CT [6]	fDSST [48]	IDCT [45]	CSPF [8]	BIT [49]	OURS
Walking2	0.462	0.652	0.869	0.506	0.846	0.873	0.343	0.594	0.314	0.767	0.334	0.777	0.551	0.849
Jogging1	0.161	0.188	0.201	0.221	0.191	0.201	0.137	0.204	0.191	0.207	0.767	0.718	0.884	0.790
Car2	0.551	0.743	0.929	0.492	0.679	0.929	0.258	0.781	0.030	0.860	0.371	0.173	0.771	0.782
Jogging2	0.133	0.151	0.147	0.271	0.741	0.161	0.133	0.139	0.091	0.144	0.152	0.820	0.091	0.742
Shaking	0.323	0.081	-	0.236	-	0.057	0.781	0.821	0.051	0.848	0.020	0.292	0.051	0.903
Singer2	0.052	0.054	0.049	0.561	0.044	0.875	0.063	0.836	0.137	0.811	0.699	0.751	0.137	0.859
Tiger	0.127	0.262	-	0.813	-	0.805	0.587	0.786	0.191	0.807	0.251	0.249	0.588	0.809
Jumping	0.176	0.707	0.113	0.031	0.126	0.343	0.049	0.579	0.055	0.061	0.071	0.211	0.055	0.508
Human8	-	0.193	-	0.127	-	0.749	0.087	0.127	0.063	0.887	0.284	0.749	0.250	0.827
Dancer	0.631	0.811	0.814	0.745	0.788	0.872	0.732	0.773	0.662	0.874	0.703	0.227	0.250	0.881
Pedestrian1	0.055	0.601	0.334	0.020	0.360	0.596	0.526	0.527	0.236	0.530	0.434	0.542	0.524	0.738
Soccer1	0.059	0.156	-	0.200	-	0.260	0.182	0.494	0.197	0.663	0.302	0.213	0.722	0.766
Football	0.631	0.671	0.671	0.706	0.574	0.678	0.593	0.664	0.612	0.777	0.113	0.692	0.627	0.711
Subway	0.193	0.801	0.081	0.768	0.186	0.850	0.189	0.844	0.699	0.836	0.757	0.758	0.818	0.877
MountainBike	0.633	0.774	0.824	0.804	0.308	0.832	0.435	0.822	0.221	0.838	0.713	0.728	0.822	0.871
Fish	0.171	0.885	0.297	0.854	0.072	0.877	0.471	0.909	0.774	0.905	0.147	0.276	0.250	0.929
car15	-	0.643	0.066	0.062	-	0.676	0.092	0.059	0.418	0.072	0.043	0.292	0.032	0.519
Tiger1	-	0.213	-	0.303	-	0.303	-	0.820	0.125	0.268	0.732	0.221	0.311	0.736
Surfer	0.512	0.651	0.056	0.025	0.033	0.275	0.038	0.666	0.087	0.815	0.213	0.182	0.250	0.817
bike2	-	0.268	0.049	0.048	0.214	0.146	0.001	0.049	0.002	0.212	0.05	0.001	0.143	0.139
boat5	0.003	0.585	0.601	0.578	0.707	0.629	0.572	0.602	0.565	0.587	0.549	0.633	0.699	0.777
Walking	0.068	0.811	0.841	0.475	0.744	0.848	0.658	0.686	0.642	0.860	0.324	0.550	0.769	0.866
Bolt2	0.023	0.019	0.013	0.669	0.013	0.803	0.386	0.014	0.644	0.014	0.162	0.759	0.250	0.679
person23	-	0.528	0.808	0.047	0.664	0.844	0.035	0.776	0.272	0.06	0.453	0.862	0.836	0.875
Average	**0.261**	**0.477**	**0.409**	**0.398**	**0.405**	**0.603**	**0.319**	**0.566**	**0.303**	**0.571**	**0.360**	**0.487**	**0.445**	**0.760**

Best results are shown in green and red, blue respectively

Jogging2 sequence. At frame # 295, IDCT, ASLA, STRUCK, WMIL, STAPLE, DCF-CA and DFT has lost the target but OURS and L1-APG track the target with minimal error till end. *Jumping* sequences has high target's deformation with motion blur and fast motion. OURS, STRUCK, and DCF-CA catered this deformation to a great extent in comparison to other trackers. *person23* sequence has fast motion challenge which OURS, DCF-CA, STAPLE and CSPF handled gracefully. In *Pedestrian1* sequence, at frame #132 DCF-CA, STAPLE, L1-APG, DFT and BIT has lost the target due to inefficient appearance model which is not able to handle the target's deformation due to fast motion and out-of plane rotation. OURS, CSPF and STRUCK efficiently handle the deformation due to fast motion. It is due to rough localization of

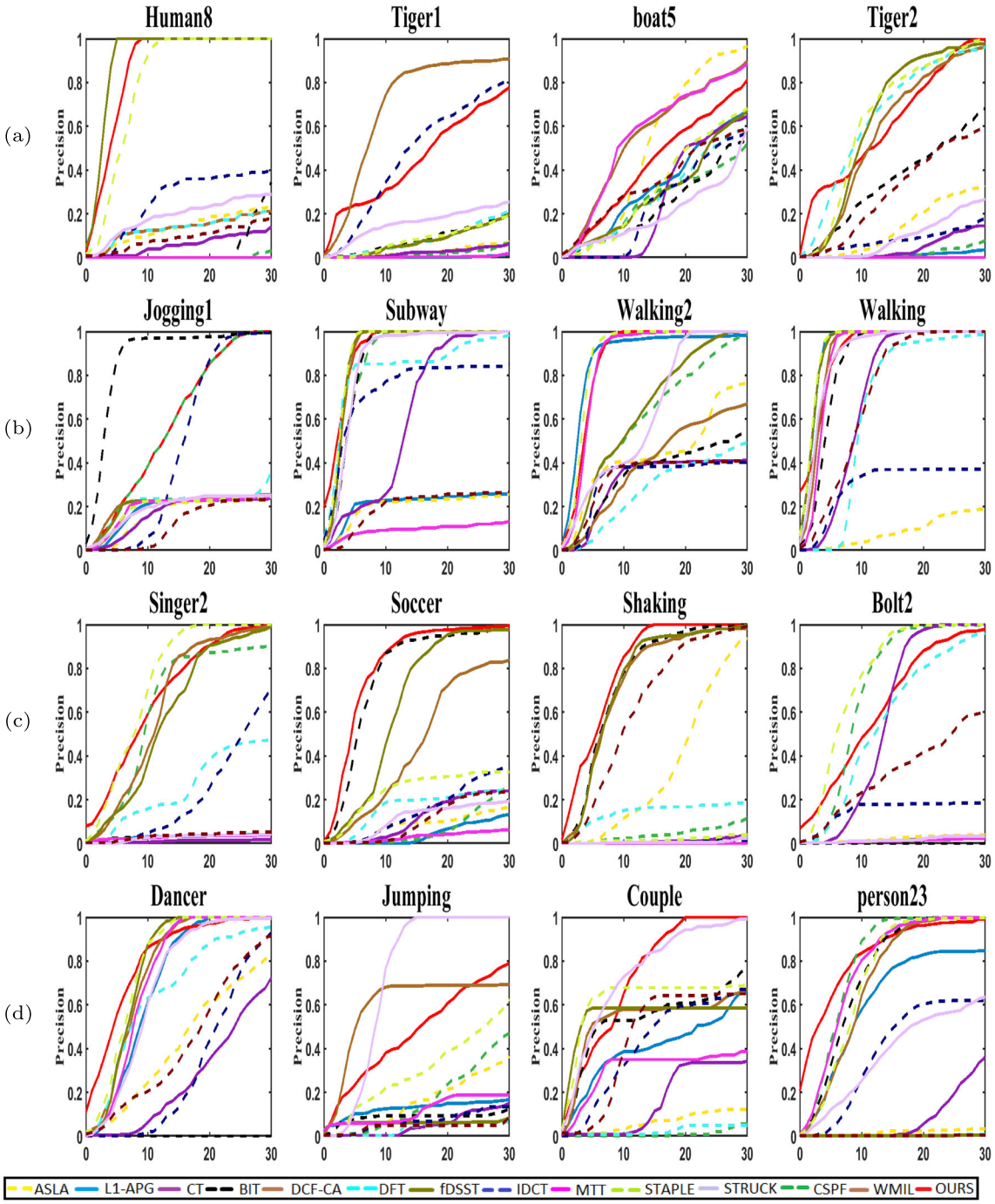


Fig. 7 Precision Plot *vs* CLE threshold (in pixels) under various visual tracking challenges: **a** Illumination Variations (*Human8*, *Tiger1*, *boat5*, *Tiger2*) **b** Full or Partial Occlusion (*Jogging1*, *Subway*, *Walking2*, *Walking*) **c** Background Clutters (*Singer2*, *Soccer*, *Shaking*, *Bolt2*) **d** Deformation and Fast Motion (*Dancer*, *Jumping*, *Couple*, *person23*)

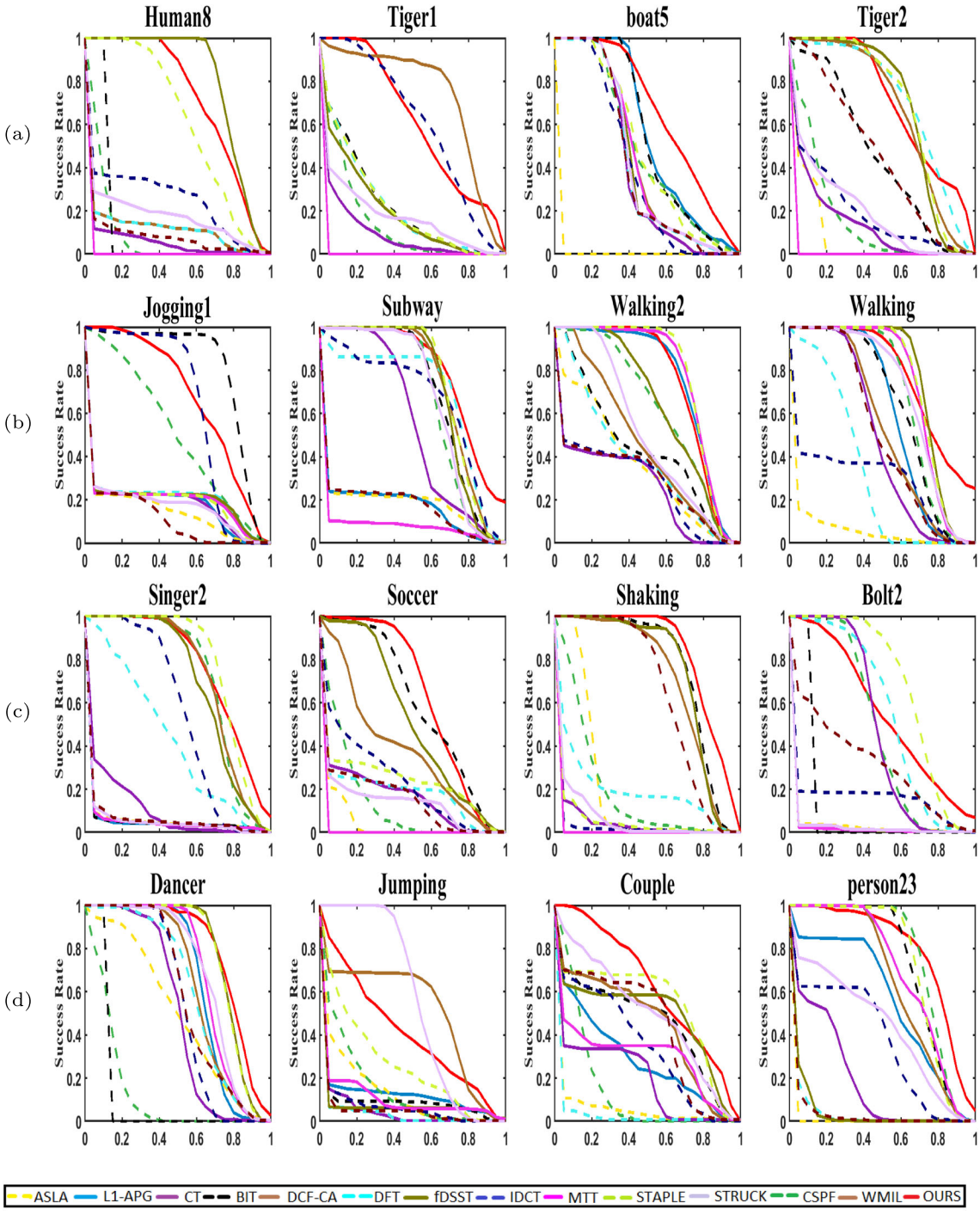


Fig. 8 Success *vs* Overlap threshold (in %) under various visual tracking challenges: **a** Illumination Variations (*Human8*, *Tiger1*, *boat5*, *Tiger2*) **b** Full or Partial Occlusion (*Jogging1*, *Subway*, *Walking2*, *Walking*) **c** Background Clutters (*Singer2*, *Soccer*, *Shaking*, *Bolt2*) **d** Deformation and Fast Motion (*Dancer*, *Jumping*, *Couple*, *person23*)

the target through optical flow which exploits the motion cue of the target. Tracking results for *Surfer* sequence are depicted in Fig. 6f. At frame #119, the appearance model of FRAG, L1-APG, DFT, MTT and IDCT get corrupted and lose the target. OURS, STRUCK and DCF-CA has handled the variations due to fast motion and in plane rotations. This may be due to periodic updation of the tracker which removes the erroneous unreliable fragments from the tracker. In addition, the rough localization of the target by optical flow exploits the motion cue in the appearance model. Transductive reliability is also integrated at each time step to prevent drift under

target's deformation when accompanied by several pose variations, fast motion and motion blur.

4.2 Quantitative analysis

In order to evaluate the performance of the proposed tracker under quantitative analysis four performance metrics are exploited. Two performance evaluations are tabulated which includes center location error (CLE) and F-Measure [52]. In addition, Success plots and Precision plots [42] are plotted to prove the robustness of the proposed approach. CLE is the Euclidean distance between the bounding box predicted by the tracker and the ground truth. F-measure (F_m) is calculated as the harmonic mean of the precision (p_r) and Recall (r_c) as $F_m = \frac{2 \times p_r \times r_c}{1 + p_r + r_c}$. Precision ($p_r$) is defined as the ratio of overlap of target bounding box and ground truth bounding box with respect to target bounding box. Recall (r_c) is calculated as the ratio of overlap between target bounding box and ground truth bounding box with respect to ground truth bounding box. Precision plot depicts the mean precision in accordance with various centre error thresholds (in pixels). Success plot presents the ratio of successfully tracked frames over various overlap thresholds (in %). Overlap precision is defined as the percentage of frames where overlap of bounding box is above a threshold.

Average CLE and average F-Measure of the proposed tracker alongwith other state-of-the-art trackers on challenging video sequences are tabulated in Tables 1 and 2 respectively. Our tracker has attained average CLE of 8.88 (in pixels) and average F-Meaure of 0.760. In addition, in Figs. 7 and 8 precision plot and success plot are illustrated respectively. Plots reveal that our tracker has attained high precision values and high success rate. In sequences with illumination variation challenge, the proposed tracker has achieved superior performance due to exploited complementary HOG and LBP features. High precision values during full or partial occlusion confirm the effectiveness of the proposed fusion of multi-cue features. During background clutters, deformation and fast motion, the suppression of deteriorating cue over the effective cues by the transductive reliability alongwith the proposed classifier mechanism validates the strength of the proposed method. Also, rough localization of the target using optical flow estimates the motion information more accurately. Our tracker performs well in tough situations by catering various dynamic environmental challenges.

Our tracker has shown eminent performance in comparison to the existing state-of-the-art trackers. Low mean CLE and high mean F-Measures validates the strength of proposed tracker. In addition, Precision plot and success plot reveals that the proposed tracker outperforms the other trackers. Generative trackers, like MTT, ASLA, and L1-APG are not able to handle background clutters and full or partial occlusion. Discriminative trackers, like WMIL and CT fail to handle target's deformation due to pose variations. STAPLE and CSPF are not able to handle target's deformation due to fast motion and motion blur efficiently. DCF-CA has shown comparable performance but not suitable for real-time computation whereas fDSST has shown resistance to occlusion and fast motion. STRUCK is able to handle background clutters challenge to a great extent but is relatively slow due to complex learning methods in the update model. WMIL has random Haar-like features in appearance model and not able to cater target deformation due to rotation. MTT and L1-APG has exploited sparse representation using single feature in the appearance model and drift away from the target during tough environmental challenges. In sum, the proposed tracker has performed better than the other trackers by exploiting complementary multi-cue LBP and HOG in the appearance model during precise localization of the target. During rough localization, proposed approach has exploited the motion cue using optical flow to handle the target deformation due to fast motion. The fusion of multi-cue generates a robust unified feature to overcome occlusion and background clutters. Nevertheless, the proposed K-Means based classifier mechanism classifies the positive reliable fragments and hence, prevent the false updation of the reference dictionary during environmental challenges.

5 Conclusion

In this paper, we have proposed a multi-stage tracking framework for object tracking. Optical flow exploits the motion cue for rough estimation of the target. During precise localization, complementary features viz. scale-invariant LBP and HOG in appearance model handles scale variations and occlusion. The proposed fusion of feature vectors improve the computational efficiency of the tracker during background clutters. K-Means based classifier mechanism classify the positive fragments and the negative fragments and hence, prevents the false updation of the reference dictionary. Nevertheless, transductive reliability of each fragment is integrated for quick adaptation of the target during environmental challenges. Quantitative and qualitative results on challenging video sequences infers the competitive performance of the proposed tracker against 13 other state-of-the-art trackers. As an average of outcome, the proposed tracker has attained CLE of 8.88 and F-Measure of 0.760.

In future, we look forward to fuse more visual cues in the appearance model. Context information and spatial information between the consecutive frames in the tracking framework can be further explored.

References

1. Zhang T, Ghanem B, Liu S, Ahuja N (2012) Robust visual tracking via multi-task sparse learning. In: 2012 IEEE Conference on computer vision and pattern recognition. IEEE, pp 2042–2049
2. Jia X, Lu H, Yang M-H (2012) Visual tracking via adaptive structural local sparse appearance model. In: IEEE Conf. on computer vision and pattern recognition, pp 1822–1829
3. Adam A, Rivlin E, Shimshoni I (2006) Robust fragments-based tracking using the integral histogram. In: 2006 IEEE CS Conf. on comp. visi. and pattern recog, vol 1, pp 798–805
4. Zhang K, Song H (2013) Real-time visual tracking via online weighted multiple instance learning. Pattern Recog 46:397–411
5. Yang Y, Hu W, Zhang W, Zhang T, Xie Y (2015) Discriminative reverse sparse tracking via weighted multitask learning. IEEE Trans Circ Syst Video Technol 27(5):1031–1042
6. Zhang K, Zhang L, Yang M-H (2012) Real-time compressive tracking. In: European conf. on computer vision. Springer, pp 864–877
7. Walia GS, Raza S, Gupta A, Asthana R, Singh K (2017) A novel approach of multi-stage tracking for precise localization of target in video sequences. Expert Syst Appl 78:208–224
8. Walia GS, Kumar A, Saxena A, Sharma K, Singh K (2019) Robust object tracking with crow search optimized multi-cue particle filter. Pattern Anal Applic. 1–17
9. He Z, Yi S, Cheung Y-M, You X, Tang YY (2016) Robust object tracking via key patch sparse representation. IEEE Trans Cybern 47(2):354–364
10. Kang B, Zhu W-P, Liang D (2017) Robust multi-feature visual tracking via multi-task kernel-based sparse learning. IET Image Process 11(12):1172–1178
11. Walia GS, Ahuja H, Kumar A, Bansal N, Sharma K Unified graph-based multicue feature fusion for robust visual tracking. IEEE Transactions on Cybernetics, https://doi.org/10.1109/TCYB.2019.2920289
12. Lan X, Zhang S, Yuen PC, Chellappa R (2017) Learning common and feature-specific patterns: a novel multiple-sparse-representation-based tracker. IEEE Trans Image Process 27(4):2022–2037
13. Lan X, Ma AJ, Yuen PC, Chellappa R (2015) Joint sparse representation and robust feature-level fusion for multi-cue visual tracking. IEEE Trans Image Process 24(12):5826–5841
14. Xiao J, Stolkin R, Gao Y, Leonardis A (2017) Robust fusion of color and depth data for rgb-d target tracking using adaptive range-invariant depth models and spatio-temporal consistency constraints. IEEE Trans Cybern 48(8):2485–2499
15. Wang H, Nguang SK, Wen J (2018) Robust video tracking algorithm: a multi-feature fusion approach. IET Comput Vis 12(5):640–650
16. Li C, Cheng H, Hu S, Liu X, Tang J, Lin L (2016) Learning collaborative sparse representation for grayscale-thermal tracking. IEEE Trans Image Process 25(12):5743–5756
17. Hu M, Liu Z, Zhang J, Zhang G (2017) Robust object tracking via multi-cue fusion. Signal Process 139:86–95
18. Meng OK, Pauline O, Kiong SC, Soong LE (2019) Effective moving object tracking using modified flower pollination algorithm for visible image sequences under complicated background. Applied Soft Computing, 105625
19. Duffner S, Garcia C (2015) Using discriminative motion context for online visual object tracking. IEEE Trans Circ Syst Video Technol 26(12):2215–2225
20. Wang W, Wang C, Liu S, Zhang T, Cao X (2017) Robust target tracking by online random forests and superpixels. IEEE Trans Circ Syst Video Technol 28(7):1609–1622
21. Li Z, Zhang J, Zhang K, Li Z (2018) Visual tracking with weighted adaptive local sparse appearance model via spatio-temporal context learning. IEEE Trans Image Process 27(9):4478–4489
22. Sun S, Liu S, Kang S, Xia C, Dan Z, Lei B, Wu Y (2018) Improved dual-mode compressive tracking integrating balanced colour and texture features. IET Comput Vis 12(8):1200–1206
23. Kanagamalliga S, Vasuki S (2018) Contour-based object tracking in video scenes through optical flow and gabor features. Optik 157:787–797
24. Li X, Hu W, Shen C, Zhang Z, Dick A, Hengel AVD (2013) A survey of appearance models in visual object tracking. ACM Trans Intell Syst Technol (TIST) 4(4):58
25. Walia GS, Kapoor R (2016) Recent advances on multicue object tracking: a survey. Artif Intell Rev 46:1–39
26. Yang Y, Hu W, Xie Y, Zhang W, Zhang T (2016) Temporal restricted visual tracking via reverse-low-rank sparse learning. IEEE Trans Cybern 47(2):485–498
27. Bai T, Li Y-F, Zhou X (2014) Learning local appearances with sparse representation for robust and fast visual tracking. IEEE Trans Cybern 45(4):663–675
28. Fu C, Duan R, Kayacan E (2019) Visual tracking with online structural similarity-based weighted multiple instance learning. Inform Sci 481:292–310
29. Feng W, Han R, Guo Q, Zhu J, Wang S (2019) Dynamic saliency-aware regularization for correlation filter-based object tracking. IEEE Trans Image Process 28(7):3232–3245
30. Zhou T, Liu F, Bhaskar H, Yang J (2017) Robust visual tracking via online discriminative and low-rank dictionary learning. IEEE Trans Cybern 48(9):2643–2655
31. Fan B, Cong Y, Du Y (2014) Discriminative multi-task objects tracking with active feature selection and drift correction. Pattern Recogn 47(12):3828–3840
32. Yu Y, Wu Q, Kirubarajan T, Uehara Y (2018) Robust discriminative tracking via structured prior regularization. Image Vis Comput 69:68–80
33. Wu Y, Pei M, Yang M, Yuan J, Jia Y (2015) Robust discriminative tracking via landmark-based label propagation. IEEE Trans Image Process 24(5):1510–1523
34. Lan X, Zhang S, Yuen PC (2016) Robust joint discriminative feature learning for visual tracking. In: IJCAI, pp 3403–3410
35. Ma B, Huang L, Shen J, Shao L (2015) Discriminative tracking using tensor pooling. IEEE Trans Cybern 46(11):2411–2422
36. Danelljan M, Bhat G, Gladh S, Khan FS, Felsberg M (2019) Deep motion and appearance cues for visual tracking. Pattern Recogn Lett 124:74–81
37. Cao Y, Ji H, Zhang W, Xue F (2018) Learning spatio-temporal context via hierarchical features for visual tracking. Signal Process Image Commun 66:50–65
38. Zeng X, Xu L, Ma L, Zhao R, Cen Y (2018) Visual tracking using global sparse coding and local convolutional features. Digital Signal Process 72:115–125
39. Wang L, Zhang L, Yi Z (2017) Trajectory predictor by using recurrent neural networks in visual tracking. IEEE Trans Cybern 47(10):3172–3183
40. Horn BK, Schunck BG (1981) Determining optical flow. Artif Intell 17(1–3):185–203

41. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
42. Wu Y, Lim J, Yang M-H (2013) Online object tracking: a benchmark. In: IEEE Conf. on comp. vis. and pattern recog., pp 2411–2418
43. Kristan M, Matas J, Leonardis A, Vojir T, Pflugfelder R, Fernandez G, Nebehay G, Porikli F, Čehovin L (2016) A novel performance evaluation methodology for single-target trackers. IEEE Trans Pattern Anal Mach Intell 38(11):2137–2155
44. Mueller M, Smith N, Ghanem B (2016) A benchmark and simulator for uav tracking. In: European conference on computer vision. Springer, pp 445–461 s
45. Asvadi A, Mahdavinataj H, Karami MR, Baleghi Y Online visual object tracking using incremental discriminative color learning. The CSI Journal on Comp. Sc. and Engg
46. Sevilla-Lara L, Learned-Miller E (2012) Distribution fields for tracking. In: IEEE Conference on computer vision and pattern recognition (CVPR), pp 1910–1917
47. Bao C, Wu Y, Ling H, Ji H (2012) Real time robust l1 tracker using accelerated proximal gradient approach. In: 2012 IEEE Conference on computer vision and pattern recognition. IEEE, pp 1830–1837
48. Danelljan M, Häger G, Khan FS, Felsberg M (2017) Discriminative scale space tracking. IEEE Trans Pattern Anal Mach Intell 39(8): 1561–1575
49. Cai B, Xu X, Xing X, Jia K, Miao J, Tao D (2016) Bit: biologically inspired tracker. IEEE Trans Image Process 25(3):1327–1339
50. Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr PHS (2016) Staple: complementary learners for real-time tracking. In: The IEEE Conference on computer vision and pattern recognition (CVPR)
51. Mueller M, Smith N, Ghanem B (2017) Context-aware correlation filter tracking. In: Proc. of the IEEE Conference on computer vision and pattern recognition (CVPR), vol 2, p 6
52. Lazarevic-McManus N, Renno J, Makris D, Jones GA (2008) An object-based comparative methodology for motion det. based on the f-measure. Comp Vis Image Un 111:74–85

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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

A Parallel Military Dog based Algorithm for Clustering Big data in Cognitive Industrial Internet of Things

Ashish Kumar Tripathi, Kapil Sharma, *Member, IEEE*, Manju Bala, Akshi Kumar, *Member, IEEE*, Varun G Menon, *Senior Member, IEEE*, Ali Kashif Bashir, *Senior Member, IEEE*

***Abstract*—With the advancement of wireless communication, internet of things, and big data, high performance data analytic tools and algorithms are required. Data clustering, a promising analytic technique is widely used to solve the IoT and big data based problems, since it does not require labeled datasets. Recently, meta-heuristic algorithms have been efficiently used to solve various clustering problems. However, to handle big data sets produced from IoT devices, these algorithm fail to respond within desired time due to high computation cost. This paper presents a new meta-heuristic based clustering method to solve the big data problems by leveraging the strength of MapReduce. The proposed methods leverages the searching potential of military dog squad to find the optimal centroids and MapReduce architecture to handle the big data sets. The optimization efficacy the proposed method is validated against 17 benchmark functions and the results are compared with 5 other recent algorithms namely, bat, particle swarm optimization, artificial bee colony, multiverse optimization, and whale optimization algorithm. Further, a parallel version of the proposed method is introduced using MapReduce (MR-MDBO) for clustering the big datasets produced from industrial IoT. Moreover, the performance of MR-MDBO is studied on 2 benchmark UCI datasets and 3 real IoT based datasets produced from industry. The F-measure and computation time of the MR-MDBO is compared with the 6 other state-of-the-art methods. The experimental results witness that the proposed MR-MDBO based clustering outperforms the other considered algorithms in terms of clustering accuracy and computation times.**

***Index Terms*—Optimization, benchmark, clustering, big data, industrial IoT.**

I. Introduction

The rapid advances in the Internet of Things (IoT) to provide smart services to the people is the trending research area in the last few years. On the same footprints, industrial IoT system is also working to make smart factories by enabling structural health monitoring, remote diagnosis, condition monitoring, and automation of various services [1] [2]. Meanwhile, industrial IoT comes with a number of challenges for making smart, reliable, and optimized system [3] [4]. Large number of sensors are deployed in the industries to collect the streaming and high volume data [5]. Generally, the data produced from these sensors is huge, streaming and unstructured. To analyze such high volume and streaming data is a challenging issue. Moreover, the streaming data arriving from the sensors is also time depended, which may loose the patterns over the time [6]. The efficient analysis of such data is a prominent part for the smart industries, as it may improve the services and reduce the risks involved in the different processes. The timely and efficient analysis of the information such as humidity, temperature, pressure and gas composition collected from large number of sensors can mitigate the chances of abnormalities. However, the traditional data analysis techniques can not handle such data sets due to limited memory and processing unites [7]. Therefore, for the efficient analysis of big industrial data sets, advanced tools and algorithms are required. Data clustering is an efficient unsupervised method of data analysis. Generally, the data generated from the sensors is not labeled, therefore clustering is a promising tool for such systems [8]. In the recent years, meta-heuristic based clustering methods have been efficiently used for solving the clustering problem due to their ability to avoid local optima [9]. Generally, meta-heuristics are used for optimization of real world problems, where traditional methods are not able to respond in the polynomial time. These algorithms starts with random set of solution in search space to find the optima in defined time interval. However, No Free Lunch theorem proves that there exists no algorithm, which can be efficiently applied for all kind of problems [10]. Therefore, a number of meta-heuristic based methods have been developed in the literature for unfolding various real world problems. Moreover, the existing sequential algorithms are not suitable to cope with high computation cost of big data sets. To remedy this challenge, in this paper, a novel meta-heuristic algorithm is introduced which mimics the searching process of suspicious objects by the trained military dog squad. Military dogs have the high smell senses, by which they can search any object like bomb, body, blood or currency. They may sense the object in the range of 200 meters by their smell power in the absence of wind. However, with the wind factor, the same can go up to 1000 meters depending the direction and speed of the wind. Moreover, some other factors such as vegetation may deviate the direction of sound or smell. In this paper, the ability of the

Ashish Kumar Tripathi is with the Department of CSE, Malviya National Institute of Technology, Jaipur, India. (e-mail:mail2ashish07@gmail.com)

Kapil Sharma is with Department of IT, Delhi Technological University, India. (e-mail: kapil@ieee.org).

Manju Bala is with Department of CSE, IP College Of Women, India. (e-mail: manjugpm@gmail.com).

Akshi Kumar is with Department of CSE, Delhi Technological University, India. (e-mail: akshikumar@dce.ac.in)

Varun G Menon is with Department of CSE, SCMS School of Engineering and Technology, Ernakulam, India. (e-mail: vaunmenon@ieee.org)

Ali Kashif Bashir is with Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom. (e-mail: dr.alikashif@ieee.org)

1551-3203 (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 June 02,2020 at 07:20:47 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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

trained military dogs is leveraged to find the optimal centroids in clustering big data produced by industrial IoT. The major contribution of this paper is summarized as follows.

1) A novel meta-heuristic based clustering method is introduced, which can handle the complexities of big industrial IoT based datasets. The mathematical model of the proposed meta-heuristic is presented. Further, the optimization efficacy of the proposed algorithm is validated against 17 benchmark functions and performance is evaluated in term of 2 parameters namely, mean fitness value and standard deviation.
2) The benchmark function results are compared with 5 existing state-of-the-art meta-heuristics namely, particle swarm optimization (PSO) [11], bat algorithm (BA) [12], artificial bee colony optimization (ABC) [13], multiverse optimization (MVO) [14], and whale optimization algorithm (WOA) [15].
3) For efficient clustering of big datasets generated from industrial IoT, a parallel version of the proposed algorithm is introduced and named MR-MDBO.
4) The performance of MR-MDBO is validated on 2 benchmark UCI datsets and 3 real datasets produced from industrial IoT. The results are compared against six MapReduce based state-of-the-art clustering algorithms namely, MR-KPSO [11], improved PSO [16], DFBP-KBA [17], MR-ABC [18], and MR-EGWO [19].

Rest of the paper is presented as follows. Section II discusses the related work. Section III introduces the mathematical model of the proposed algorithm. Section IV presents parallel MDBO model using MapReduce. Section V provides the experimental results. Finally, the conclusion and future work is elucidated in section VI.

II. Related Work

A. Big data in cognitive IoT: Opportunities and challenges

In the first decade of IoT from 2009 to 2018, several mature products have been introduced to develop the IoT based systems, containing sensors, cloud platform, and big data analytic systems [20] [21]. Now a days, we are in the second decade of IoT, where the novel techniques for the data analysis produced from the IoT based systems are essential research topics [22] [23]. Meanwhile, the industrial IoT comes with a number of challenges, for smart control and management of high volume and unstructured data, produced from IoT based devises and sensors [23] [24]. For the same, researchers are working to develop efficient methods for analyzing data produced from IoT. Recently, some techniques and frameworks for the big data analysis have been introduced for the industrial applications [21]. Cao et al. [25] presented an outlier detection framework for the big streaming data. Rosa et al. [26] introduced a non parametric approach for the classification of industrial data. Fong et al. [6] introduced a lightweight particle swarm based feature selection method for the streaming big data analysis. Zhang et al. [1] introduced incremental CFS algorithm for clustering the big data in industrial IoT. Subsequently, Yuan et al. [16] introduced spark based improved particle swarm optimizer for the anomaly data mining in mass sensor networks. However, in the last few years, the data mining challenges of IoT based problems have been research trend [3]. Some meta-heuristic based algorithms are introduced in the literature for solving the computationally complex problems using MapReduce. However, for the IoT based framework, no study have been perform in the literature. To bridge this gap, a novel MDBO based algorithm have been introduced in this work for clustering big data in industrial IoT.

B. Challenges of Meta-heuristic based clustering

Producing acceptable solutions in a reasonable time is one of the key challenge for the meta-heuristic algorithms. Resources like memory have always been a concern, an algorithm is efficient if it uses the resources in an optimized way and always maintains a trade-off between time and space complexity.

At present, due to the tremendous growth in industrial data, the complexity of real-world problems related to optimization tasks in the smart industries has been also increased. The conventional meta-heuristic computation is facing new challenges due to computation intensive objective functions. To mitigate these concerns, nowadays, researchers are working with parallel and distributed meta-heuristic computation [27]. A number of platforms are available for solving the computation-intensive problems using distributed meta-heuristic computation, such as CUDA, GPU, and MapReduce based programming. MapReduce is an open source parallel computation framework, that has been successfully utilized for the parallel computations of meta-heuristics algorithms for managing large-scale and streaming datasets [28], [29]. In the last one decade, Hadoop and MapReduce based model for parallel processing have been used by a number of researchers for solving complex real-world problems. Thus, in this paper Hadoop and MapReduce programming framework is leveraged for the parallel processing of the complex dataset produced from industrial internet of things.

III. Military Dog Squad based Optimization

This section details the behavior and mathematical model of the trained military dogs for finding the optimal solution in the predefined search space. Military dogs undergo hard training through special training coaches, where they learn to search any suspicious object using their strong smell sense. A rigorous training is performed to train them for working in team. Military dogs uses barking to pass message to each other. They cooperate each other by passing the message, by different ways of barking and loudness. The way of barking and loudness represents some specific message and loudness helps them to identify the distance from the suspicious object. When a trained military dog squad is left for finding the suspicious object in a tertiary, it start searching the area randomly. Military dogs make a move in the area, using their smell sensation and barking sound from other dogs of the squad. Each military dog defines its fitness value in terms of loudness. The loudest barking represents the best fitness value, as it is assumed to be closest from the target object. Further,

1551-3203 (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 June 02,2020 at 07:20:47 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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

each military dog uses scent smell to reach close to the target object which represents the exploitation phase of the MDBO. However, they also take moves based on the loudness of any randomly chosen dog, which helps to explore the search space and avoid the local optima.

A. MDBO Definitions and Algorithm

To design an efficient algorithm, the searching behavior of trained military dog squad is mathematically modeled. This section presents the definitions and mathematical model of trained military dog squad. First, the formal definitions of MDBO are presented. Thereafter, the complete MDBO algorithm is explained. In the following definitions, the set of empty numbers and real numbers is represented with ϕ and R respectively, whereas, the set of integers is represented by Z.

Definition 1: MDS^N :; is defined as military dog squad of N trained military dog. The value of N in the squad remains constant with time. In future, the value of N can be variable for the further improvements.

Definition 2: $MDSI : MD \rightarrow R$:; represents the military dog smell index. It is defined as the closeness of a MD with the target object. In other meta-heuristic based algorithms, this factor is commonly known as fitness of individual.

Definition 3: FSV^d; is a feasibility solution vector, which denotes the position of a trained military dog in the military dog squad. Further, $FSV \in R^d$ is a set of all real numbers which represents the position of a military dog in d^{th} dimension.

$Definition4$: Sniffing movement $\delta(p, P_m,) : MD \rightarrow MD$; represents an operator, which modifies the FSV^d of military dogs randomly, based on the $MDSI$ of fittest military dog. P_m is the movement probability, which controls the movement of a military dog. The mathematical model of sniffing movement is defined by the following equation.

$$FSV_i^j(i+1) = \begin{cases} FSV_{loudest}^j, \ p \leq P_m \\ FSV_i^j + R(0,1) \times mstep(i), \ p > P_m \end{cases} \tag{1}$$

where,

$mstep(i) = WC \times R(0,1) \times (FSV_i^j - FSV_{loudest}^j)$

$R(0,1)$ is any randomly chosen number between 0 and 1. WC is a wind constant and p is any randomly generated number in the range (0,1).

Definition 5: $\omega(P_m, X, \alpha) : MD^n \rightarrow$ MD; is an operator which defines the barking movement of the military dog. The barking movement of a trained military dog is controlled by FSV^d of fittest dog and it is also induced by any randomly chosen military dog. $X \in (1, 2, 3..., d)$ is the randomly chosen index and P_m is the movement probability which remains constant. Further, α is the vegetation or smog factor which may effect the intensity of the sound in the air. The barking movement modifies the feasible solution vector (FSV^d) of a military dog, which is defined as follows:

$$FSV_i^j(i+1) = \begin{cases} FSV_i^j(i), \ p \leq \alpha \\ FSV_i^j(i) + B_m \times R(0,1), \ p > \alpha \end{cases} \tag{2}$$

where $B_m = (FSV_{loudest} - FSV_q)$ and $R(0,1)$ represents a random number in the range (0,1).

Definition 6: The transition function $\phi = (m, d, \delta, \omega, P_m) : MD^m \rightarrow MD^m$; represents a 5-tuple MDS transition function, which updates the MDS in each iteration. It starts with the computation of FSV^d and $MDSI$. Further, modification operator is applied on each MD of MDS and the $MDSI$ value is redefined.

Definition 7: Finally, MDBO algorithm is defined as three tuple, $MDBO = (H, \phi, T)$, which obtains the solution for any optimization function. The initial population is generated using function $H :\rightarrow \{MD^n, MDSI^n\}$, and corresponding $MDSI$ is calculated. ϕ is a MDS transition function defined earlier. H generates the FSV of each military dog in the specified search space. $T : MD^n \rightarrow \{true, false\}$ represents the termination criterion of the algorithm.

The complete procedure of the MDBO is described with the following steps.

1) In the first step, all the parameters of $MDBO$ algorithm are initialized. A method is derived to map the problem to FSV^d and MDS as defined in definition 1 and 2. This process is dependent on the nature of problem. Further, the maximum population size (military dogs), movement probability P_m, vegetation constant α, and wind factor W are fixed as per the nature of the optimization function.
2) In the second step, the position (FSV) of each MD in the search space is initialized according to the search boundaries defined in the optimization problem. This process is performed by the H operator explained in definition 7.
3) Sniffing movement (exploitation step): Further, each MD redefines its FSV which is influenced by the fittest (loudest barking) military dog in the squad. During the search, each MD makes a random walk and look around the location. MD approaches the target object using the smell sensation. MD may take move towards the loudest barking dog with probability P_m. In the other case, it may take random move influenced by loudest baking dog.

   The complete pseudocode of the sniffing movement is defined as follows:

```
for (i = 1 to N) do
    W = rand(0, 1)
    IF(W < P_m)
    FSV_i^j(i + 1) = FSV_loudest^j
    step(i) = WC × R(0, 1) × (FSV_i^j − FSV_loudest^j);
    FSV_i^j(i + 1) = FSV_i^j + R(0, 1) × step(i)
end for
```

4) After the sniffing movement, barking movement is performed to explore the search space. Generally, it is observed that military dogs bark loud, where they sense any suspicious object. This barking sound makes a global movement, as other military dogs follow the loudest barking dog. Each military dog makes a movement, according to the fittest (loudest) one, and any other randomly chosen barking dog from the squad. The

1551-3203 (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 June 02,2020 at 07:20:47 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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

TABLE I: Ranking of the considered algorithms using Friedman Test

Parameter Name	PSO	Bat	ABC	MVO	WOA	MDBO
Rank	6	5	4	3	2	1
Mean Rank	5.52	4.50	4.20	3.46	2.82	1.32

FSV of each military dog is redefined by the barking movement, which is explained as follows.
The complete pseudocode for barking movement is presented as follows:

for $(i = 1\ to\ N)$ **do**
  $K = rand(0, 1)$
  S=IF($K < \alpha$)
  $FSV_i^{(i+1)} = FSV_i(t) + B_m \times S$
  $B_m = rand * (FSV_{loudest} - FSV_q)$;
end for

5) Start the next iteration from step 3. This loop is continued until the maximum number of iterations are not reached, or the required solution is achieved. This step implements the T operator as explained in definition 6.

B. MDBO based clustering

In the MDBO based clustering, the FSV of each military dog denotes a set of centroids, $C = \{C_1, C_2, \cdots, C_M\}$ for M number of clusters. The $MDSI$ value represents the fitness (squired Euclidean distance) of MD as shown in Eq. (3).

$$SED(N, M) = \sum_{i=1}^{N} \sum_{j=1}^{M} W_{ij} \mid Z_i - C_j \mid \qquad (3)$$

Where, M represents the umber of cluster, N denotes the number of data items, $\mid N_i - C_j \mid$ presents the Ecludian distance between I^{th} data point J^{th} center. Subsequently, W_{ij} is an association weight corresponding to i^{th} vector in j^{th} cluster. The value of W_{ij} is 1, if i^{th} data point belongs to j^{th} cluster, else it is set as 0. In the clustering process, each solution is represented by a FSV, which represents the position of MD in the search space. The algorithm starts by randomly generating N random $FSVs$, where N is the size of MDS. The FSV of each MD is updated in each iteration to improve the $MDSI$ value. Finally, after the termination criteria is reached, the FSV corresponding to the fittest MD is returned which represents the position of final cluster centers.

Algorithm 1 :Map Function

Input: $Key - RecordID, Value - Record.$
Output: Key-MD-ID, Value-Minimum distance
Map (Key : dataId, Value : data)
MD-ID = read (file);
for each MD in MDS **do**
  MD-ID = read(MD-ID);
  Centroid =read(centroids) // FSV represents the location of centroids
  Min-D= getMinD(record, Centroid); //getMinD returns centroid nearest to the record
  centroid-ID = i //index of the centroid with minimum distance
  new-key = MD-ID+centroid-ID;
end for
write (new-key, Min-D);

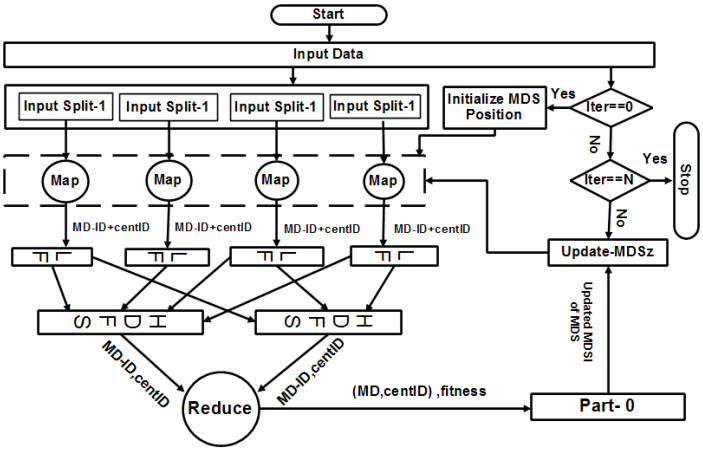


Fig. 1: Model of parallel MDBO for data clustering

Algorithm 2 :Reduce Function

Input: $Key - (MD - Id, centroid - ID), Value - (Min - Distance)$
Output: MD-ID, Value-MD-fitness // sum of intra-cluster distance
Reduce (Key:MD-Id, centroid-ID, Value-list: Min-Distance) // value contains list of distances of data point from their nearest centroids
Initialization
MD-fitness=0
for each distance in Min-Distance list **do**
  MD-fitness=sum-Dist+distance
end for
Update the positions of MD
write(key, MD-fitness)

IV. Parallel MDBO for IoT based Industrial Big Data Clustering

In this section, parallel model of MDBO algorithm using Hadoop MapReduce is explained. In MapReduce based MDBO (MR-MDBO), two main operations are performed namely, updating of the cluster centroid and the MDSI computation, which represents the fitness of the clustering solution. In the proposed algorithm for data clustering, the centroid of the clusters are updated according to MR-MDBO with the purpose of minimizing sum of the squared Euclidean distance. The complete MapReduce architecture of the MDBO for the clustering of large datasets is shown in Fig. 1. As shown in Fig., first the dataset is divided into smaller data blocks called input split. In the first iteration of the algorithm, population of the MR-MDBO is initialized and supplied to each mapper running on the different nodes. In the clustering process using MR-MDBO, the main computation intensive task is to compute the sum of the squared Euclidean distance. The MapReduce model is thus employed to compute the fitness value. Each iteration of MR-MDBO runs in two phases called MR-MDBO-Map and MR-MDBO-Reduce. In the proposed MR-MDBO, the task of fitness computation is done in the MR-MDBO-Map phase, in which parallelism is achieved, since each machine have only some fraction of the whole dataset. As shown in in Algo 1, the MR-MDBO-Map function starts with extracting the centroids of each cluster from the population which is stored in the HDFS (hadoop distributed file system). The MR-MDBO-Map function then retrieves FSV vector of a military dog that represents location of the centroids.

1551-3203 (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 June 02,2020 at 07:20:47 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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

Further, the distance of centroid is calculated with each data object and minimum distance along with its $centroid-ID$ is returned. The MR-MDBO-Map function writes ($MD-ID$, $centroid-ID$) as the new key and the data point is written as the new value. After the completion of the MR-MDBO-Map phase, output of all the mappers are merged and grouped by keys.

In the MR-MDBO-reduce phase, the reduce function is called on each key, value pair generated by the MR-MDBO-Map phase. The reducer function aggregates all the values with the identical key's to compute the fitness value that we aim to minimize. The main function of the MR-MDBO-map is to decompose the task and that is merged by the MR-MDBO-reduce phase. Finally, as shown in the Algo. 2 the reduce function computes the sum of the squared Euclidean distance between each data object and the respective cluster centroids. The newly computed fitness value is used to update the FSV of the military dog squad in the next iteration. The whole MapReduce cycle is repeated and this process continue until the stopping criterion or maximum iterations are not reached. Algo. 2 presents the Pseudo-code of the MR-MDBO-reduce function.

The time complexity of MR-MDBO based clustering is proportional to the number of data points, number clusters, and number of dimensions in the dataset. In the MR-MDBO based clustering, the optimal cluster centroids are generated with O($N\times K\times D\times I$) operations, where N, K, D, and I represents the number of data points, number of clusters and number of dimensions, and number of iterations respectively. For M number of agents, the time complexity of the MR-MDBO based clustering can be represented as O($M\times N\times K\times D\times I$).

V. Performance Analysis and Results

The performance analysis of MDBO is presented in two folds. First, the optimization ability of MDBO is vindicated against 17 standard benchmarks. The results are are detailed in section V-A. Second, the clustering ability of MDBO is validated against big data sets produced from industries, and the results are presented in section V-A3.

A. Benchmark Function Results

This section presents the optimization efficacy and uniqueness of MDBO. For the fair compression, seventeen benchmark functions are used, and the results are presented in terms of mean and standard deviation of the fitness value. Furthermore, Wilcoxon rank sum and FriedMan's test of MDBO is performed with the other algorithms, which validates the dissimilarity of MDBO and other considered algorithms. Each algorithms is run on system with 3.0 Ghz frequency, Intel (R) core(TM) i7 processor and 16 GB RAM with Matlab 2015a.

1) Comparison with existing algorithms: The proposed MDBO is first tested on benchmark functions to validate the optimization efficacy. Table II presents the detail of each benchmark function, including name, and category (Uni-model or Multi-model). Among seventeen benchmark functions, nine are the uni-model, whereas rest eight function belong to multi-model class. Generally, uni-model functions are used to validate the convergence rate whereas, multi-model functions are used to test the ability of local optima avoidance. The optimal value for each benchmark function is defined as 0. The parameter values of each algorithm as taken from the respective literature. Furthermore, each algorithm is run thirty times to obtain the mean and standard deviation of the fitness value. Table III presents the results on each benchmark function in terms of mean and the standard deviation. It is observed from the Table III that MDBO outperformed five other state-of-the-art existing algorithms on 17 considered benchmark functions in terms of mean of the fitness value. However, on one function i.e F6, WOA has shown competitive results. Moreover, in terms of standard deviation, MDBO has outperformed all the algorithms on 16 benchmark functions, while for one function i.e., F1, again WOA has produced minimum value. Further, it can be analyzed that MDBO has shown superior results on all 9 uni-model function, which validates its stronger ability for the local search. However, on the multi-model functions MDBO outperformed in seven among eight, which shows stronger exploration property. Thus, it is concluded that MDBO can be successfully utilized for solving real world optimizations problems.

2) Wilcoxon and Friedman's test: The Wilcoxon rank sum and FriedMans's tests are performed to validate the dissimilarity and efficacy of MDBO with other compared algorithms. The test is performed over fiver percent significance level (*ninety five percent confidence*) using NULL hypothesis. Further, p value of the MDBO and other algorithms has been computed using the best fitness of each iteration. Null hypothesis is reject, if p<0.05 and presented with '+' or '-' symbol. However, if the value of p>0.05, NULL hypothesis is accepted and represented with symbol '='. The '+' symbol shows superior result of the MDBO whereas '-' indicate inferior results of MDBO as compared to other algorithm. The results of Wilcoxon rank sum test of MDBO with the other considered algorithms over 17 benchmark functions are presented in Table II. The performance of MDBO is analyzed with 5 other state-of-the-art algorithms namely, PSO, Bat, ABC, MVO, and WOA in terms of p value. The p value on each benchmark function is calculated by running thirty iterations of each algorithm. The significant value (SG) is calculated on the basis of p value and mean of the fitness value. As shown in the table, the value of SG is positive if $p < 0.05$ and mean fitness value of MDBO is better than the compared algorithm. However, if the mean fitness value of MDBO is less than the compared algorithm, the SG is represented with '-' symbol. The $p > 0.05$ indicate that MDBO has failed the test of dissimilarity. It can be concluded from the Table II, that proposed algorithm has surpassed WOA on all seventeen benchmark functions except F6. Moreover, MDBO has outperformed Bat algorithm on all the seventeen benchmark functions. Also, MDBO produced better results than MVO on sixteen benchmark functions among seventeen. However, for F6, PSO has given competitive results. Also, MDBO has shown positive SG value on all the functions against PSO and ABC except one function i.e, F6. Further, the Friedman's test was also performed to find the mean rank of each algorithm. Table I presents the ranking of each algorithm

1551-3203 (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 June 02,2020 at 07:20:47 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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

TABLE II: p and SG values of the Wilcoxon rank-sum test at α = 0.05 on standard benchmark functions

Fun	Name	Type	PSO		Bat		ABC		MVO		WOA	
			p-value	SG	p-value	SG	p-value	SG	p-value	SG	p-value	SG
F_1	Ackley	Multi-Model	4.24E-06	+	4.44E-08	+	4.28E-06	+	4.28E-06	+	5.28E-06	+
F_2	Alpine	Multi-Model	9.07E-06	+	6.44E-08	+	6.32E-08	+	8.07E-06	+	8.07E-06	+
F_3	Dixon and Price	Uni-modal	4.48E-06	+	4.32E-06	+	6.44E-08	+	4.39E-06	+	4.59E-06	+
F_4	Griewank	Multi-model	6.21E-06	+	4.31E-06	+	4.48E-08	+	3.31E-06	+	5.21E-06	+
F_5	Levy	Multi-Model	2.29E-06	+	6.44E-06	+	6.48E-07	+	4.39E-06	+	4.39E-06	+
F_6	Pathological	Multi-model	6.16E-09	=	5.22E-06	+	0.167962	=	0.167962	=	0.212258	=
F_7	Perm	Multi-Model	6.87E-07	+	6.32E-09	+	6.87E-07	+	6.87E-07	+	8.22E-08	+
F_8	Powell	Uni-model	4.39E-06	+	4.30E-07	+	3.44E-07	+	3.44E-07	+	4.48E-04	+
F_9	PowellSum	Uni-Model	6.87E-07	+	7.78E-04	+	6.87E-07	+	8.87E-07	+	6.87E-07	+
F_{10}	Rastrigin	Uni-Model	4.39E-06	+	2.40E-08	+	4.44E-06	+	4.51E-06	+	4.20E-06	+
F_{11}	Rosenbrock's	Multi-Model	4.39E-06	+	3.39E-08	+	3.48E-06	+	4.48E-06	+	4.46E-06	+
F_{12}	Rotated Hyper-Ellipsoid	Uni-Model	4.42E-08	+	3.42E-07	+	3.39E-06	+	4.290E-06	+	6.32E-06	+
F_{13}	Schumer Steiglitz	Uni-model	4.48E-07	+	3.39E-08	+	3.39E-06	+	3.39E-06	+	2.32E-08	+
F_{14}	Schwefel	Multi-Model	4.48E-08	+	3.38E-08	+	3.38E-06	+	3.38E-06	+	4.44E-08	+
F_{15}	Sphere	Uni-Model	4.48E-08	+	3.39E-08	+	3.39E-06	+	3.39E-06	+	4.22E-04	+
F_{16}	Step	Uni-model	2.34E-08	+	1.26E-08	+	1.26E-06	+	1.26E-06	+	2.22E-07	+
F_{17}	Trigonometric	Uni-Model	2.80E-06	+	1.90E-08	+	1.90E-06	+	1.90E-06	+	2.28E-07	+

TABLE III: Mean of fitness value & standard deviation for 30 runs on seventeen benchmark functions

Fun	PSO		Bat		ABC		MVO		WOA		MDBO	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
F_1	4.51E+00	0.19E+00	4.44E+00	0.30E+00	2.18E+00	4.32E-01	1.63E+00	4.50E-01	6.31E+00	**1.88E-01**	**6.24E-04**	6.22E-01
F_2	4.89E+02	3.54E+01	2.22E+00	5.01E+00	5.29E+01	1.15E+01	1.27E+02	3.42E+01	4.91E+02	3.67E+01	**1.11E+00**	**0.22E+00**
F_3	8.56E+09	1.84E+09	4.11E+10	3.03E+08	1.05E+06	1.06E+06	5.71E+03	5.32E+03	8.26E+09	2.54E+09	**1.20E+02**	**4.22E+02**
F_4	1.46E+00	0.06E+00	2.22E+00	0.04E+00	1.07E+00	2.57E-02	1.50E-02	1.25E-02	1.54E+00	4.98E-02	**2.32E-04**	**2.40E-03**
F_5	3.69E+03	5.89E+02	6.31E+04	2.13E+02	5.32E+01	1.86E+01	4.83E+02	3.14E+02	3.70E+03	3.60E+02	**1.30E+01**	**1.10E+01**
F_6	9.37E+00	0.66E+00	6.44E+00	0.52E+00	5.65E+00	4.94E-01	9.93E+00	3.62E-01	**4.48E+00**	**3.19E-01**	4.68E+00	8.03E-01
F_7	2.00E+117	3.60E+117	4.10E+99	1.20E+99	5.59E+58	2.15E+59	1.05E+17	4.05E+17	4.50E+117	6.50E+117	**2.00E+08**	**0.00E+00**
F_8	4.62E+04	93.54E+02	2.18E+03	7.40E+03	1.86E+02	1.22E+02	7.78E+00	5.49E+00	1.24E+05	2.91E+04	**6.32E-034**	**4.20E-04**
F_9	4.46E+41	4.99E+41	2.12E+22	4.80E+54	3.33E+16	1.20E+17	2.96E+11	4.10E+11	8.41E+43	1.70E+44	**1.00E+08**	**0.00E+00**
F_{10}	1.54E+02	1.17E+01	2.12E+00	3.94E+01	1.40E+01	6.49E+00	1.23E+02	3.56E+01	4.14E+02	1.96E+01	**1.04E+01**	**4.22E+00**
F_{11}	1.39E+08	3.50E+07	4.12E+07	3.10E+06	9.47E+03	1.03E+04	4.58E+02	5.30E+02	1.50E+08	2.87E+07	**1.10E+01**	**3.00E+01**
F_{12}	2.46E+01	1.99E+04	9.56E+04	7.77E+04	4.88E+03	2.50E+03	1.83E+01	1.31E+01	2.79E+05	2.70E+04	**8.89E-12**	**5.32E-12**
F_{13}	1.87E+08	3.56E+08	7.40E+06	6.77E+06	4.25E+04	3.49E+04	1.01E-01	5.52E-02	1.55E+08	4.13E+07	**3.55E-16**	**5.74E-16**
F_{14}	8.65E+00	3.32E+02	4.12E+03	1.07E+03	3.22E+03	6.44E+02	4.53E+03	7.70E+02	8.41E+03	3.79E+02	**5.48E+02**	**2.20E+02**
F_{15}	4.52E+00	4.14E+03	1.02E+04	2.70E+03	1.05E+03	4.98E+02	6.76E-01	1.91E-01	4.62E+04	4.21E+03	**2.07E-12**	**1.58E-12**
F_{16}	9.32E+02	6.70E+00	4.82E+02	7.82E+01	1.82E+02	5.65E+01	5.60E+00	4.73E+00	9.12E+02	5.87E+01	**2.67E-01**	**7.04E-01**
F_{17}	0.89E+00	2.36E+00	7.67E+03	6.86E+00	5.05E+00	1.96E+01	1.85E+00	9.18E-01	7.68E+03	2.36E+03	**2.74E-01**	**5.10E-01**

obtained by Friedmans test. As shown in the table, again MDBO algorithm is ranked first as it has achieved minimum mean rank. Thus, it is concluded from the experimental results that MDBO algorithm is significantly different and superior as compared to five state-of-the art algorithms namely, PSO, Bat, ABC, MVO, and WOA for all the considered methods.

3) Clustering Analysis: In section V-A1, MDBO has shown superior results on the seventeen benchmark functions. In this section, the clustering efficacy of the MapReduce based MDBO, (MR-MDBO) is analyzed on 2 UCI and 3 big industrial IoT based data sets. Table V presents the details of datasets, including number of data-points, number of clusters, and number of dimensions. Poker-hand and SUSY are the large size benchmark data-sets widely used for testing the clustering performance, whereas DLemp, sIoT, and IoT_bonet are the three real industrial IoT based big datsets. For the experiments, a Hadoop cluster of 10 computers having Intel Core-i7 processor with 3.20GHz, 16GB of memory and 2TB hard disk is designed. All the methods were run with Hadoop version 3.6.2, jdk 1.8.0, and Ubuntu version 14.04. The results of the MR-MDBO are presented in terms of computation time and F-measure. Table IV shows the mean of F-measure and computation time of five state-of-the-art MapReduce based methods, obtained by running each method on a cluster of 10 nodes. It is clearly observed from the table, that MR-MDBO has surpassed all the considered methods on four datasets in terms of F-measure, whereas for Pokerhand, MR-ABC has given competitive results. Moreover, the computation time of MR-MDBO is minimum among all the MapReduce based methods on all the datasets, where as I-PSO has performed well on two datasets namely, pokerhand and IoT_bonet. Thus, it can be inferred from the experimental results, that MR-MDBO can be used for efficient analysis of IoT based big datasets.

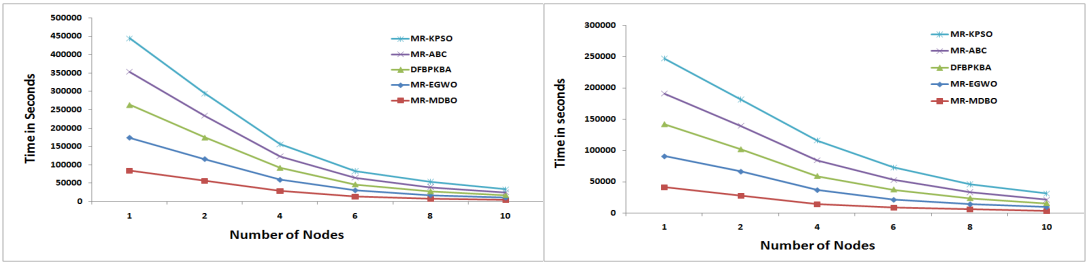
4) Speedup Analysis: Furthermore, the speedup performance of all the considered methods is also studied on a cluster with ten nodes. For analyzing the speedup efficacy, two big industrial IoT datasets namely, sIot and IoT_Bonetdata sets have been used as described in Table V. The computation time of each MapReduce based method has bee recorded by adding 2 nodes in the cluster in each run. It is pertinent from the Fig. 2 that the running time of the MDBO reduces gradually

1551-3203 (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 June 02,2020 at 07:20:47 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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

TABLE IV: Computation time and F-measure of proposed and considered method over 30 runs

#DS	Criteria	MR-K-PSO	I-PSO	MR-ABC	DFBPKBA	MR-EGWO	MR-MDBO
Susy	F-Measure	0.684	0.690	0.718	0.738	0.732	**0.744**
	Computation Time	12.20E+04	9.33E+04	12.20E+04	12.12E+04	11.02E+04	**8.12E+04**
Pokerhand	F-Measure	0.664	0.674	**0.688**	0.684	0.718	0.733
	Computation Time	13.20E+04	**10.30E+04**	12.14E+04	15.14E+04	18.16E+04	12.88E+04
DLePM	F-Measure	0.534	0.544	0.526	0.515	0.522	**0.658**
	Computation Time	14.12E+04	11.30E+04	13.22E+04	12.21E+04	13.23E+04	11.21E+04
sIoT	F-Measure	0.642	0.632	0.594	0.610	0.654	**0.754**
	Computation Time	14.08E+04	11.08E+02	10.24E+04	10.38E+04	10.88E+04	**10.10E+03**
IoT_Bonet	F-Measure	0.788	0.712	0.786	0.822	0.810	**0.846**
	Computation Time	09.66E+04	**08.20E+04**	10.44E+04	09.88E+04	9.98E+04	9.10E+04



(a) (b)

Fig. 2: The speedup analysis of (a) Slot (b) sIoT_Bonet

TABLE V: Datasets description

SN	Name	Number of Cluster	Dimension	Number of datapoints
1	Susy [30]	2	9	10,000,050
2	Pokerhand [30]	10	10	1025010
3	DLePM [31]	3	52	50,300
4	sIoT [32]	3	255	60,000,0
5	IoT_Bonet [30]	10	115	7062606

with the increasing the number of nodes. It is also concluded from the Fig, that the performance of MR-MDBO is close to ideal for the datasets large in size and having more number of features and clusters. Moreover, it is clearly vindicated from the Fig, that the computation time of MR-MDBO is minimal among all the MapReduce based considered method on both the datasets. Hence, it can be concluded that, as the size and number of clusters, features of the dataset grew, the advantage of the proposed parallel MDBO are increased.

VI. Conclusion

This paper presents a novel method for clustering big IoT based data sets, produced from industry. The proposed method finds the optimal centroids by using the ability of trained military dogs to sense the suspicious object by their strong smell sense and communication. The contribution of the paper has three folds, (i) a novel meta-heuristic algorithm, inspired from trained military dogs has been introduced (ii) the optimization ability of the proposed algorithm has been tested on seventeen standard benchmark functions and the results are compared with 5 other state-of-the-art meta-heuristics. The proposed MDBO outperformed the all the considered algorithms in terms of standard deviations and fitness value on all the benchmark functions. The uniqueness and superiority of the MDBO is also validated using the Wilcoxon rank-sum and Friedman's test. Finally, a parallel model of the proposed algorithm (MR-MDBO) is introduced for clustering the IoT based big datasets from industry. The clustering efficacy of the MR-MDBO is validated against F-measure and computation time. The clustering performance of the MR-MDBO is compared with 6 other state-of-the-art MapReduce based parallel methods. Five standard benchmark datasets, 2 UCI and 3 real IoT based data-sets from industry are used to perform the test. The experimental results validated that proposed MR-MDBO has surpassed all the considered methods in terms of F-measure and computation time. Also, the speedup analysis of the parallel algorithm has been analyzed on a cluster with 10 nodes to test the scalability. The results witnesses that the MR-MDBO has outperformed all the compared algorithms in clustering big data sets. Thus, it can be concluded that MR-MDBO can serve as an alternative tool for handling real world big data and IoT based problems. The future work will include the applications of the proposed method in other real world applications pertaining to IoT and big data. Also, some other tools such as spark can be tested to accelerate the computation efficiency of the algorithm.

References

[1] Q. Zhang, C. Zhu, L. T. Yang, Z. Chen, L. Zhao, and P. Li, "An incremental cfs algorithm for clustering large data in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1193–1201, 2017.

[2] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "Iot-based big data storage systems in cloud computing: perspectives and challenges," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 75–87, 2016.

[3] C.-W. Tsai, S.-J. Liu, and Y.-C. Wang, "A parallel metaheuristic data clustering framework for cloud," *Journal of Parallel and Distributed Computing*, vol. 116, pp. 39–49, 2018.

[4] R. C. Luo and T.-J. Hsiao, "Indoor localization system based on hybrid wi-fi/ble and hierarchical topological fingerprinting approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 791–10 806, 2019.

1551-3203 (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 June 02,2020 at 07:20:47 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/TII.2020.2995680, IEEE Transactions on Industrial Informatics

[5] C. Zhu, L. Shu, T. Hara, L. Wang, S. Nishio, and L. T. Yang, "A survey on communication and data management issues in mobile sensor networks," *Wireless Communications and Mobile Computing*, vol. 14, no. 1, pp. 19–36, 2014.

[6] S. Fong, R. Wong, and A. V. Vasilakos, "Accelerated pso swarm search feature selection for data stream mining big data," *IEEE transactions on services computing*, vol. 9, no. 1, pp. 33–45, 2015.

[7] L. Gu, D. Zeng, S. Guo, Y. Xiang, and J. Hu, "A general communication cost optimization framework for big data stream processing in geo-distributed data centers," *IEEE Transactions on Computers*, vol. 65, no. 1, pp. 19–29, 2015.

[8] A.-R. Al-Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, "A smart home energy management system using iot and big data analytics approach," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 426–434, 2017.

[9] T. Ashish, S. Kapil, and B. Manju, "Parallel bat algorithm-based clustering using mapreduce," in *Networking Communication and Data Knowledge Engineering*. Springer, 2018, pp. 73–82.

[10] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation*, vol. 1, pp. 67–82, 1997.

[11] R. Eberhart and J. Kennedy, "Particle swarm optimization," in *Proceedings of the IEEE international conference on neural networks*, vol. 4. Citeseer, 1995, pp. 1942–1948.

[12] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature inspired cooperative strategies for optimization (NICSO 2010)*. Springer, 2010.

[13] D. Karaboga, "Artificial bee colony algorithm," *scholarpedia*, vol. 5, no. 3, p. 6915, 2010.

[14] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, "Multi-verse optimizer: a nature-inspired algorithm for global optimization," *Neural Computing and Applications*, vol. 27, no. 2, pp. 495–513, 2016.

[15] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, vol. 95, pp. 51–67, 2016.

[16] J. Yuan, "An anomaly data mining method for mass sensor networks using improved pso algorithm based on spark parallel framework," *Journal of Grid Computing*, pp. 1–11, 2020.

[17] A. K. Tripathi, K. Sharma, and M. Bala, "Dynamic frequency based parallel k-bat algorithm for massive data clustering (dfbpkba)," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 4, pp. 866–874, 2018.

[18] A. Banharnsakun, "A mapreduce-based artificial bee colony for large-scale data clustering," *Pattern Recognition Letters*, vol. 93, pp. 78–84, 2017.

[19] A. K. Tripathi, K. Sharma, and M. Bala, "A novel clustering method using enhanced grey wolf optimizer and mapreduce," *Big data research*, vol. 14, pp. 93–100, 2018.

[20] D. V. Medhane and A. K. Sangaiah, "Pcca: position confidentiality conserving algorithm for content-protection in e-governance services and applications," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 3, pp. 194–203, 2018.

[21] A. K. Sangaiah, D. V. Medhane, T. Han, M. S. Hossain, and G. Muhammad, "Enforcing position-based confidentiality with machine learning paradigm through mobile edge computing in real-time industrial informatics," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4189–4196, 2019.

[22] S. Rajesh, V. Paul, V. G. Menon, and M. R. Khosravi, "A secure and efficient lightweight symmetric encryption scheme for transfer of text files between embedded iot devices," *Symmetry*, vol. 11, no. 2, p. 293, 2019.

[23] A. K. Sangaiah, M. Sadeghilalimi, A. A. R. Hosseinabadi, and W. Zhang, "Energy consumption in point-coverage wireless sensor networks via bat algorithm," *IEEE Access*, vol. 7, pp. 180 258–180 269, 2019.

[24] V. Anavangot, V. G. Menon, and A. Nayyar, "Distributed big data analytics in the internet of signals," in *2018 International Conference on System Modeling & Advancement in Research Trends (SMART)*. IEEE, 2018, pp. 73–77.

[25] L. Cao, D. Yang, Q. Wang, Y. Yu, J. Wang, and E. A. Rundensteiner, "Scalable distance-based outlier detection over high-volume data streams," in *2014 IEEE 30th International Conference on Data Engineering*. IEEE, 2014, pp. 76–87.

[26] R. De Rosa, F. Orabona, and N. Cesa-Bianchi, "The abacoc algorithm: a novel approach for nonparametric classification of data streams," in *2015 IEEE International Conference on Data Mining*. IEEE, 2015, pp. 733–738.

[27] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Sensing as a service model for smart cities supported by internet of things," *Transactions on emerging telecommunications technologies*, vol. 25, no. 1, pp. 81–93, 2014.

[28] A. Kumar Tripathi, K. Sharma, and M. Bala, "Fake review detection in big data using parallel bbo," *International Journal of Information Systems & Management Science*, vol. 2, no. 2, 2019.

[29] A. K. Tripathi, K. Sharma, and M. Bala, "Parallel hybrid bbo search method for twitter sentiment analysis of large scale datasets using mapreduce," *International Journal of Information Security and Privacy (IJISP)*, vol. 13, no. 3, pp. 106–122, 2019.

[30] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[31] Q. Zhang and Z. Chen, "A weighted kernel possibilistic c-means algorithm based on cloud computing for clustering big data," *International Journal of Communication Systems*, vol. 27, no. 9, pp. 1378–1391, 2014.

[32] Q. Zhang, L. T. Yang, and Z. Chen, "Privacy preserving deep computation model on cloud for big data feature learning," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1351–1362, 2015.

Ashish Kumar Tripathi has received his M.Tech and PhD from he department of Computer Science & Engineering, Delhi Technological University. He is currently working as an Assistant Professor in the department of Computer Science & Engineering at Malviya National Institute of Technology (MNIT), Jaipur, India. He is active reviewer of several journals of repute. His research interests are Big data analytics, Social Media Analytics, Soft Computing and Natural Language Processing.

Prof. Kapil sharma received his Ph.D in Computer Engineering in the area of software reliability. He is currently working as Head in the Department of Information Technology at Delhi Technological University. He is active reviewer of many reputed journals of IEEE, IET, ACM, Springer and Elsevier. He is also a senior member of IEEE. His research interests are in the area of Software Reliability, Soft Computing, and Big data analytics.

Manju Bala completed her PhD in Computer Science from Jawaharlal Nehru University, New Delhi, India in 2011. She is currently an Assistant Professor with the department of Computer Science, Indraprastha College for Women, University of Delhi. Her research interests include computer graphics, pattern classification and machine learning. She has authored research papers in many national and international journals.

Akshi Kumar received her Ph.D in Computer Engineering from University of Delhi, India. She is currently an Assistant Professor in the Department of CSE at Delhi Technological University. She is active reviewer of many top journals of IEEE, IET, ACM, Springer and Elsevier. Her research interests are in the area of Affective Computing, Sentiment Analysis, Social Media Analytics, Soft Computing and Natural Language Understanding.

Varun G. Menon completed his Ph.D in Computer Science and Engineering from Sathyabama University, India in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. His research interests include IoT, data analytics, brain signal analysis, fog computing, and underwater acoustic sensor networks.

Ali Kashif Bashir is a Senior Lecturer at the Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom. He received his Ph.D. in computer science and engineering from Korea University South Korea. His research interests include internet of things, wireless networks, distributed systems, machine learning, etc. He is serving as the Editor-in-chief of the IEEE FUTURE DIRECTIONS NEWSLETTER. He is also editor of several reputed journals.

1551-3203 (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 June 02,2020 at 07:20:47 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

Adaptive Droop based Control for Active Power Sharing in Autonomous Microgrid for Improved Transient Performance

Surya Prakash, Student Member IEEE, Vaibhav Nougain, and Sukumar Mishra, *Senior Member, IEEE*

***Abstract*—This paper presents a decentralised adaptive droop based control for active power sharing between parallel inverters in autonomous microgrids. The work is focused on improving the transient performance of the system proposing a robust controller. The analysis does not require load current measurement where the Lyapunov analysis based control takes load current as a disturbance to the system. The mathematical analysis takes a more practical form of *LC* filter configuration where the dynamics due to resistances across the *LC* is also taken into consideration. Controller gain conditions are taken into consideration employing a detailed Lyapunov analysis. A comparative study of the conventional droop control with the proposed adaptive droop control is included in the work. MATLAB has been employed as the simulink platform for the comparative study. In addition, for experimental verification, FPGA based hardware setup has been developed and the related results are presented to verify the real-time operation of proposed adaptive control.**

***Index Terms*—Point of Common Coupling (PCC), Low Voltage (LV), Medium Voltage (MV), Backstepping Control, Voltage Source Converter (VSC), Phase Locked Loop (PLL), Constant Power Load (CPL), Adaptive Droop**

NOMENCLATURE

p	Instantaneous active power
q	Instantaneous reactive power
P	Filtered active power
Q	Filtered reactive power
V_{odq}	PCC point voltage in *dq* frame of reference
V^*_{odq}	Virtual PCC point voltage in *dq* frame of reference
V_{idq}	Inverter output voltage in *dq* frame of reference
I_{idq}	Inverter current in *dq* frame of reference
I^*_{idq}	Virtual Inverter current in *dq* frame of reference
I_{odq}	Load current in *dq* frame of reference
L_f	Filter Inductance
R_f	Resistance associated with the filter inductance
C_f	Filter Capacitance
R_d	Equivalent leakage Resistance associated with the filter capacitance
ω^*	Nominal system frequency
ω_o	Reference system frequency
m	Static droop for Real Power sharing
α	Adaptive droop for Real Power sharing
e_1, e_2	Current Tracking Errors
e_3, e_4	Voltage Tracking Errors

This work was supported in part by the Science & Energy Research Board (SERB), Department of Science & Technology (DST), Govt. of India through the Distinguished Investigator Award (DIA) scheme (File number - DIA/2018/000032).

The authors are with the Department of Electrical Engineering, Indian Institute of Technology Delhi, India. (e-mail: eeesuryaprakash@gmail.com, nougainvaibhav@gmail.com and sukumar@ee.iitd.ernet.in)

I. INTRODUCTION

THE penetration of power electronic interfaces in the power system has changed the course of how electricity is perceived. Power electronic interfaces have made it possible for the renewable sources (solar, wind, fuel cell etc.) to be either interface with the conventional utility grid or to be able to get locally consumed upon generation forming an autonomous microgrid [1]. The sources of such microgrid, i.e., distributed generators (DGs) are interfaced with Voltage source converters (VSCs). The VSCs, running in parallel, is responsible for controlling the frequency and voltage profile of the system due to the absence of grid [2]. The multi-DG configuration, in parallel shares the total power of the microgrid. The discussed control objective could be met by employing centralised [3], distributed [4] or decentralised methods [5].

The primary challenge for the autonomous microgrid is maintaining voltage and frequency of operation for proper operation of the loads. Once the primary challenge is dealt, the secondary challenge surfaces which are meeting the load demand in a microgrid. Once the secondary problem is addressed, the tertiary and final challenge is, maintaining the power-sharing between different sources proportionately to ensure the sources not hitting their current limit for their operational longevity. The droop control based power-sharing is widely accepted due to its decentralised nature [6]-[7]. However, droop based approach has the inherent problem of a trade-off between the voltage regulation and power-sharing accuracy [8]. To elaborate, a larger value of droop gives better power-sharing. However, the oscillations increase for such cases [9]. Transient operations in power-sharing in such microgrid setups cause overstressing, which affects the life of DG units. Mitigation of such transients is necessary to ensure the longevity of VSCs [10]-[11]. The transient and steady-state performance of the system is dependent on the control architecture employed. The conventional Proportional-Integral (PI) control has **I** (Integral) which is responsible for the robust steady-state performance. However, it is also responsible for the sluggish transient performance of the controller. The hierarchical droop based control [12]-[14] has been discussed employing radial basis

2168-6777 (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 May 22,2020 at 05:46:56 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

function neural network (RBFNN). However, such schemes are sluggish due to multiple control layers. The idea of single control layer modern non-linear controllers have surfaced in the power domain for improved transients performance of such microgrid setups. Sliding mode control (SMC) [15]-[16] has been proposed to improve the transient response of the system. However, SMC inherently suffers from the problem of chattering along the sliding surface. The authors in [17] employ Adaptive Neural Network (ANN) control to conclude improved transient performance. However, the gain re-tuning for different configurations is a problem of control in [17]. The authors in [18] introduce a robust Takagi-Sugeno (T-S) fuzzy control for a parallel islanded AC microgrid. However, the complexity of such control makes its implementation difficult [19].

The idea of the proposed work is improvement of transient performance for active power sharing employing an Adaptive control. Virtual inputs [20] are introduced to frame a non-linear systematic recursive control design. The transient performance of such virtual input based control has been validated in [21]. The proposed adaptive control is a simple, robust approach without the problem of chattering as in SMC. A detailed Lyapunov analysis is used to design the gain conditions for the proposed control. The performance of the proposed adaptive control is tested under dynamic AC (RLC) transients, CPL transients and fault transients.

The rest of the paper is organised as follows. Section II presents the system architecture of the autonomous microgrid. Section III gives the modelling of parallel VSCs, including system dynamics, control problem formulation, control synthesis, generation of firing signal and tuning of control gains. Section IV gives the simulation results. Section V presents the hardware results for the proposed adaptive control.

II. System Architecture

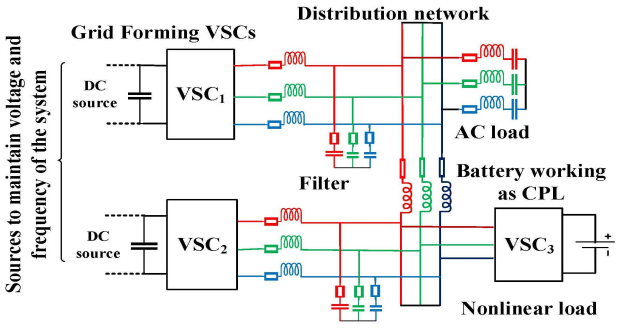


Fig. 1: Autonomous Microgrid setup

The system under consideration is an autonomous microgrid with its abstract presentation depicted in Fig. 1. Since the network is a distribution type in nature, the operating voltage of the system is 400V (phase to phase). The coupling between active and reactive power is not taken into consideration since the filter inductance value is considerable in such cases [22]. A Constant Power Load (CPL) has been taken into account due to its nature of negative incremental resistance which tends to destabilise the distribution network [23]. The different loads are distributed in nature, and hence, if the load current is a requirement for controlling the autonomous microgrid, a large number of current sensors are required. The proposed analysis does not require load current measurement, saving the cost of multiple load current sensors. The DG units and the loads are interfaced employing filter and distribution network, as shown in Fig. 1. The filter configuration employed is *LC* type where the resistances R_f (across L_f) is also included in the dynamics of the proposed analysis. The control architecture is built around the degree of freedom of the filter parameters (L_f, R_f and C_f). The power electronic interfaces, i.e., VSCs, are modulated using a control strategy obtained from Lyapunov based Adaptive technique which has been discussed in the following section.

III. Modelling of Parallel VSCs in Autonomous Microgrid

The control objective of the VSCs operating in parallel is to maintain the voltage and frequency of the autonomous microgrid, sharing the total load power in the microgrid. The instantaneous active power, p shared by each VSC is the microgrid can be calculated as

$$p = V_{od} I_{id} + V_{oq} I_{iq} \tag{1}$$

where the parameters employed can be seen in Fig. 3. The fundamental component of power, P is obtained using a low pass filter and is expressed as

$$P = \frac{\omega_c}{s + \omega_c} p \tag{2}$$

where ω_c is the cut-off frequency of the low pass filter.

The approach applied widely for such control objective is Proportional-Integral (PI) based static droop control [24]-[25]. The proposed adaptive method employs an adaptive droop control governed by,

$$\omega_0 = \omega^* - (m + \alpha)P \tag{3}$$

where ω_0 is the reference frequency of the autonomous microgrid, ω^* is the nominal frequency which is taken to be 314 *rad/s*, m is the static droop and α is the adaptive droop. Intuitively, the adaptive droop term is responsible for the improved transient response of the power-sharing in the autonomous setup. The idea behind introducing α is, increasing static droop, m improves the power-sharing; however, a low-frequency is introduced in the system if the value of static droop is very high [24]. Hence, a low value of m with adaptive α improves the power-sharing and helps to eliminate the problem of low-frequency oscillations in the system.

A. Control Design Using Lyapunov analysis based Adaptive Technique

The control design is derived using Lyapunov energy function of the autonomous system. The analysis employs adaptive technique to introduce parameter estimate ($\hat{\alpha}$) for adaptive term (α) and virtual control inputs (i^*_{1dq}, V^*_{odq}) in

2168-6777 (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 May 22,2020 at 05:46:56 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

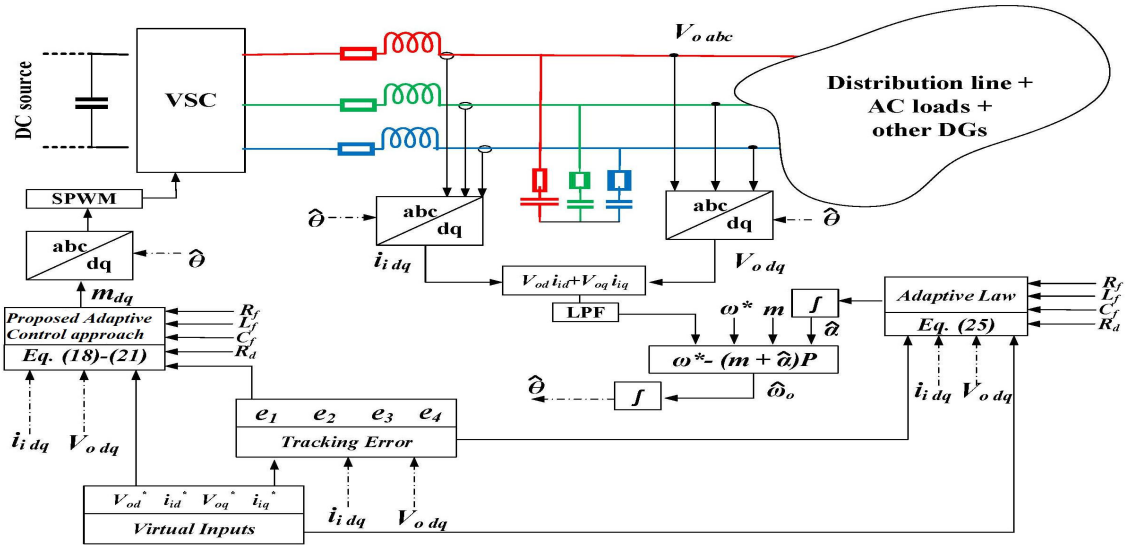


Fig. 2: Proposed Adaptive Control

the system dynamics [20]-[21], [26] (the terms are elaborated subsequently). The control technique takes load current (i^*_{odq}) as disturbance focusing on the notion of boundedness of the tracking error (e_1, e_2, e_3, e_4) dynamics. The results ensure exponential convergence of the tracking errors to a bound. Control gains introduced (γ, k_1, k_2, k_3, k_4) in the analysis are employed to vary the bound for the errors.

B. Mathematical analysis

The dynamics of the 3ϕ autonomous inverter setup is given using Kirchoff's law where the system is transformed into the direct-quadrature (*d-q*) form from the 3ϕ form. R_f are considered in the dynamics to present a more practical filter configuration. The system after the *d-q* transformation can be represented with the following differential equations where the notations are as indicated in Fig. 3.

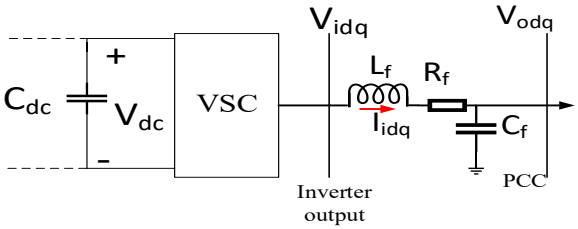


Fig. 3: The direct-quadrature equivalent of the 3ϕ (abc) single VSC configuration

1) System Dynamics:

$$\dot{i}_{1d} = -R_f i_{1d}/L_f - V_{od}/L_f + V_{id}/L_f + \omega_0 i_{1q} \quad (4)$$

$$\dot{i}_{1q} = -R_f i_{1q}/L_f - V_{oq}/L_f + V_{iq}/L_f - \omega_0 i_{1d} \quad (5)$$

$$\dot{V}_{od} = i_{1d}/C_f - i_{od}/C_f + \omega_0 V_{oq} \quad (6)$$

$$\dot{V}_{oq} = i_{1q}/C_f - i_{oq}/C_f - \omega_0 V_{od} \quad (7)$$

The discussed technique does not require load current measurement and hence i_{od} and i_{oq} are considered to be disturbances with an upper bound.

$$|i_{od}| < \bar{d_1} \qquad |i_{oq}| < \bar{d_2} \quad (8)$$

2) Control Problem Formulation: The closed loop system is considered for the control problem formulation. The tracking error dynamics for the system is defined as,

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} i_{id} - i_{id}* \\ i_{iq} - i_{iq}* \\ V_{od} - V_{od}* \\ V_{oq} - V_{oq}* \end{bmatrix}$$

where $i_{1d}*$ and $i_{iq}*$ are the virtual input current references to the *d-axis* and *q-axis*; $V_{od}*$ and $V_{oq}*$ are the static virtual *d-axis* and *q-axis* voltage references. Inclusion of the tracking error dynamics modify eq. (4)-(7) to represent the closed loop error dynamics for the system (eq. (9)-(12)).

$$\dot{e}_1 = -R_f e_1/L_f - R_f i^*_{1d}/L_f - V_{od}/L_f + V_{id}/L_f + \omega_0 i_{1q} - \dot{i}^*_{1d} \quad (9)$$

$$\dot{e}_2 = -R_f e_2/L_f - R_f i^*_{1q}/L_f - V_{oq}/L_f + V_{iq}/L_f + \omega_0 i_{1d} - \dot{i}^*_{1q} \quad (10)$$

$$\dot{e}_3 = e_1/C_f + i^*_{1d}/C_f + \omega_0 e_4 + \omega_0 V^*_{oq} - i_{od}/C_f \quad (11)$$

$$\dot{e}_4 = e_2/C_f + i^*_{1q}/C_f - \omega_0 e_3 - \omega_0 V^*_{od} - i_{oq}/C_f \quad (12)$$

3) Adaptive Control Synthesis: The Lyapunov Energy Function for the state dynamics of the autonomous inverter is defined as;

$$V_1(e_1, e_2, e_3, e_4) = \frac{1}{2} L_f e_1^2 + \frac{1}{2} L_f e_2^2 + \frac{1}{2} C_f e_3^2 + \frac{1}{2} C_f e_4^2 \quad (13)$$

2168-6777 (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 May 22,2020 at 05:46:56 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

where the energy function is positive definite and radially unbounded. The derivative of the energy function is further realized as;

$$\dot{V}_1(e_1,e_2,e_3,e_4) = L_f e_1\dot{e}_1 + L_f e_2\dot{e}_2 + C_f e_3\dot{e}_3 + C_f e_4\dot{e}_4 \quad (14)$$

Parameter estimation error for the adaptive term is defined as,

$$\tilde{\alpha} = \alpha - \hat{\alpha} \quad (15)$$

where $\hat{\alpha}$ is the known estimate of adaptive droop, α.

$$\eta = m + \hat{\alpha} \quad (16)$$

where η is a parameter used for simplification. The estimated reference system frequency is defined as,

$$\hat{\omega_o} = \omega^* - \eta P \quad (17)$$

The direct-quadrature derivative of virtual current inputs (i^*_{1d} and i^*_{1q}) and the direct-quadrature modulated inverter output voltages (V_{id} and V_{iq}) of the autonomous inverter are defined as,

$$V_{id} = R_f i^*_{1d} + V^*_{od} + L_f \dot{i}^*_{1d} - \hat{\omega_o} L_f i^*_{1q} - k_1 e_1 \quad (18)$$

$$V_{iq} = R_f i^*_{1q} + V^*_{oq} - L_f \dot{i}^*_{1q} - \hat{\omega_o} L_f i^*_{1d} - k_2 e_2 \quad (19)$$

$$i^*_{1d} = -C_f \hat{\omega_o} V^*_{oq} - k_3 e_3 \quad (20)$$

$$i^*_{1q} = -C_f \hat{\omega_o} V^*_{od} - k_4 e_4 \quad (21)$$

Hence, incorporating the derivative of virtual current inputs and the modulating inverter output voltages (eq. (18)-(21)), the derivative of the energy function is given as;

$$\begin{aligned}\dot{V}_1(e_1,e_2,e_3,e_4) &= -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 \\ &-k_4 e_4^2 - e_3 i_{od} - e_4 i_{oq} + L_f e_1 i^*_{1q}\tilde{\alpha}P - L_f e_2 i^*_{1d}\tilde{\alpha}P \\ &+C_f e_3 V^*_{oq}\tilde{\alpha}P - C_f e_4 V^*_{oq}\tilde{\alpha}P\end{aligned} \quad (22)$$

Now, the idea is to introduce another energy function to take the terms with $\tilde{\alpha}$ into consideration.

$$V_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = V_1 + 0.5\tilde{\alpha}^2 \quad (23)$$

Taking the derivative of new energy function,

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = \dot{V}_1 - \tilde{\alpha}\dot{\hat{\alpha}} \quad (24)$$

The derivative of the estimate of adaptive term is defined as,

$$\dot{\hat{\alpha}} = -P(L_f e_1 i^*_{1q} + L_f e_2 i^*_{1d} - C_f e_3 V^*_{oq} + C_f e_4 V^*_{od}) \quad (25)$$

The idea of designing $\dot{\hat{\alpha}}$ is to shape the energy function, V_2 eliminating the terms which are impeding the decay of energy function derivative. Accounting eq. (25), the derivative of the energy function is modified to the form,

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 - k_4 e_4^2 - e_3 i_{od} - e_4 i_{oq} \quad (26)$$

The idea is to upper bound the derivative of the energy function considering the direct-quadrature profile of load current and their derivative as disturbances in the system.

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) \le -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 - k_4 e_4^2 + |e_3|\bar{d_1} + |e_4|\bar{d_2} \quad (27)$$

The evaluation of control gains (k_1–k_4) is based on the philosophy of exponential convergence of tracking errors (e_1–e_4) towards a bounded region around equilibrium point. The exponential convergence of tracking errors is guaranteed if the derivative of the energy function can be expressed in the form, $\dot{V}_2 \le -\gamma V_2 + c$ ensuring uniform ultimate boundedness of the energy function and the tracking errors.

$$\begin{aligned}\dot{V}_2 \le &-(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - \gamma C_f e_3^2 - \gamma C_f e_4^2 - \\ &|e_3|\left[(k_3-\gamma C_f)|e_3| - \bar{d_1}\right] - |e_4|\left[(k_4-\gamma C_f)|e_4| - \bar{d_2}\right]\end{aligned} \quad (28)$$

The control gains (k_1–k_4) are defined as,

$$k_1 = k_2 = \gamma L_f - R_f \quad (29)$$

$$k_3 \ge \gamma C_f + \bar{d_1}/|e_3| \quad (30)$$

$$k_4 \ge \gamma C_f + \bar{d_2}/|e_4| \quad (31)$$

4) Generation of the Firing Signal for the inverter: The autonomous inverter is fired using the proposed control philosophy. The SPWM technique employed for firing includes generation of m_{dq}, inverse Park's Transformation (m_{dq} to m_{abc}), comparing with carrier sinusoid to generate the firing signal. The generation of m_{dq} is defined employing eq. (18)-(19) as,

$$\begin{bmatrix} m_d & m_q \end{bmatrix} = \frac{2}{V_{dc}} \begin{bmatrix} V_{id} & V_{iq} \end{bmatrix}$$

5) Eigen Value Synthesis: The control objective of exponential convergence of tracking error towards a bound requires tuned value of introduced control gains (k_1–k_4). Eq(29)-(31) give the range of control gains. The dynamics is then employed to do the eigenvalue synthesis studying variation of control gains on the system performance. The synthesis gives the idea of small-signal stability of the system with the variation of control gain parameters. The state-space representation is expressed as,

$$\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) + \boldsymbol{D}d(t) \quad (33)$$

where $\boldsymbol{A}$ is the state matrix, $\boldsymbol{B}$ is the input matrix and $\boldsymbol{D}$ is the disturbance matrix [see eq. (32)]. $x = [e_1\ e_2\ e_3\ e_4]^T$ gives the state variables, $u = [V^*_{od}\ V^*_{oq}]^T$ gives the inputs, $d = [d_1\ d_2\ d_3\ d_4]^T$ gives the disturbances in VSC state-space representation. The eigenvalues are plotted for varying control gains. Fig. 4 shows the eigenvalue loci for varying k_1,k_2,k_3,k_4. Table I & II shows the eigenvalues for different k_1,k_2,k_3,k_4. It can be seen from table I that $k_{1min} > 0$. Similar studies are presented subsequently to verify the values of k_{2min}, k_{3min}, k_{4min} for system stability.

TABLE I: Eigen value loci with the variation of gain parameters, k_1 and k_2

λ	k_1=1000	k_1=500	k_1=0
λ_1	-471.23	-471.08	52.57
λ_2, λ_3	-436.61± 399.13i	-436.61±399.17i	-426.49±392.64i
λ_4	-1443.82	-848.15	-546.69
λ	k_2=1000	k_2=500	k_2=0
λ_1	-471.23	-471.08	52.57
λ_2, λ_3	-436.61± 399.13i	-436.61±399.17i	-426.49±392.64i
λ_4	-1443.82	-848.15	-546.69

2168-6777 (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 May 22,2020 at 05:46:56 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

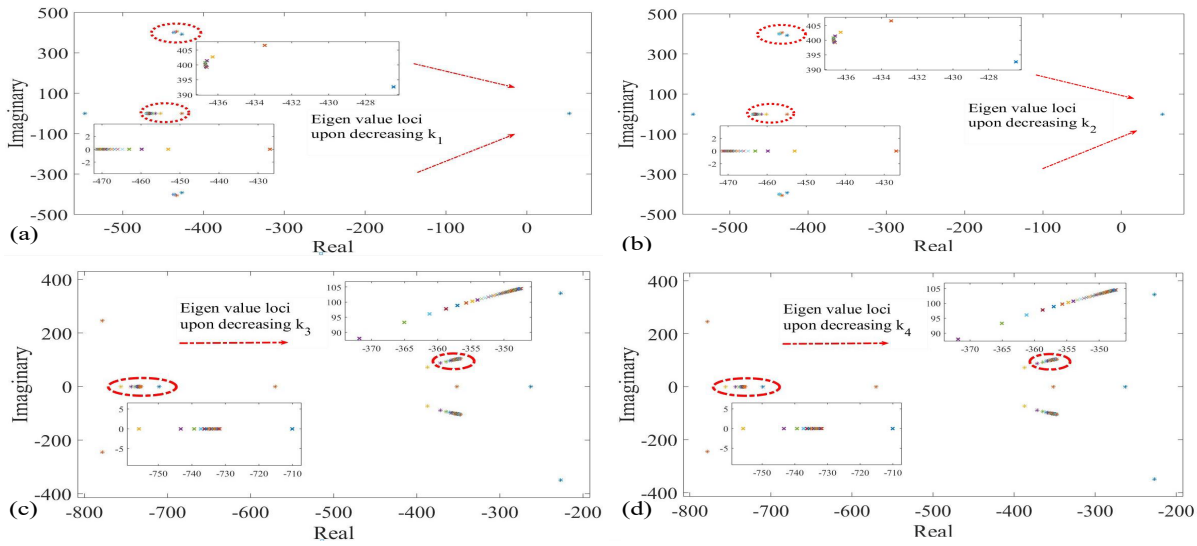


Fig. 4: Eigen Value loci for variation of (a) k_1, (b) k_2, (c) k_3, (d) k_4

$$\frac{d}{dt}\begin{bmatrix}\Delta e_1\\ \Delta e_2\\ \Delta e_3\\ \Delta e_4\end{bmatrix} = \begin{bmatrix} -\frac{R_f+k_1}{L_f} - \eta V_{od} e_2 & \hat{\omega_o} - \eta V_{oq} e_2 & -\frac{1}{L_f} - \eta i_{1d} e_2 & -\eta i_{1q} e_2 \\ \hat{\omega_o} - \eta V_{od} e_1 & -\frac{R_f+k_2}{L_f} - \eta V_{oq} e_1 & -\eta i_{1d} e_1 & -\frac{1}{L_f} - \eta i_{1q} e_1 \\ \frac{1}{C_f} - \eta V_{od} e_3 & -\eta V_{oq} e_4 & -\frac{k_3}{C_f} - \eta i_{1d} e_4 & \hat{\omega_o} - \eta i_{1q} e_4 \\ \eta i_{1d} e_3 & \frac{1}{C_f} + \eta V_{oq} e_3 & -\hat{\omega_o} + \eta i_{1d} e_3 & -\frac{k_4}{C_f} + \eta i_{1q} e_3 \end{bmatrix} \begin{bmatrix}\Delta e_1\\ \Delta e_2\\ \Delta e_3\\ \Delta e_4\end{bmatrix} + \eta \begin{bmatrix} -i_{1d} e_2 & -i_{1q} e_2 \\ -i_{1d} e_1 & -i_{1q} e_1 \\ -i_{1d} e_4 & -i_{1q} e_4 \\ i_{1d} e_3 & i_{1q} e_3 \end{bmatrix} \begin{bmatrix}\Delta V_{od}^* \\ \Delta V_{oq}^*\end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -\frac{1}{C_f} & 0 \end{bmatrix} \begin{bmatrix}\Delta d_1 \\ \Delta d_2\end{bmatrix} \quad (32)$$

TABLE II: Eigen value loci with the variation of gain parameters, k_3 and k_4

λ	k_3=1000	k_3=500	k_3=0
λ_1	-731.71	-531.83	-262.70
λ_2, λ_3	-347.55± 390.81i	-347.87± 104.24i	-226.64± 349.02i
λ_4	-2105.25	-1052.61	-710.05
λ	k_4=1000	k_4=500	k_4=0
λ_1	-731.71	-531.83	-262.70
λ_2, λ_3	-347.55± 390.81i	-347.87± 104.24i	-226.64± 349.02i
λ_4	-2105.25	-1052.61	-710.05

IV. Simulation Results and Discussion

MATLAB/Simulink platform is employed for the simulation-based validation. The system considered is shown in Fig. 1. The simulation events considered for the comparison between the proposed control and the conventional PI control are elaborated in Table III. The events are kept similar to ensure proper comparison of control actions.

A. Simulation Events

1) AC load transients: Initially, both VSCs operating in parallel are contributing *320W* each (*m=22.86e-4*, α*=0*) when the AC load is consuming *640W* as can be seen in Fig. 6(a). A load transient of *640W* occurs at *t=1.3s* making the total load *1280W* as explained in Table III. The value of the adaptive droop increases to *10.27e-4* from *0* for both VSCs

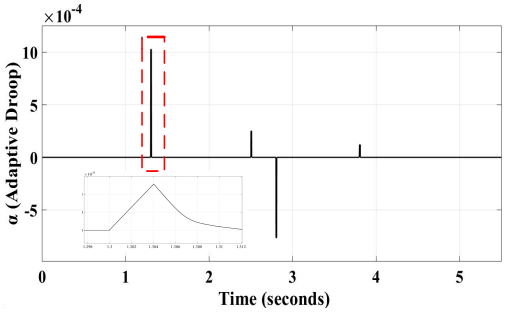


Fig. 5: Variation of adaptive droop, α with time with Event transients indicated in Table III

TABLE III: Event transients captured for comparison in Fig. 5

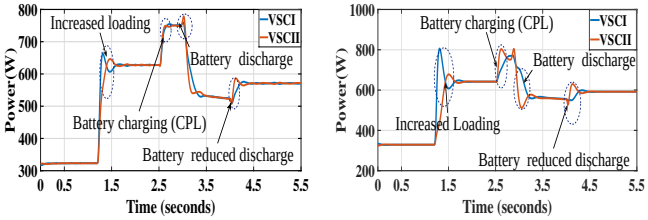
Parameter	Time of operation	Value
$P_1 + P_2$ (W)	0.0 s to 1.3 s	640 W (320 W each)
$P_1 + P_2$ (W)	1.3 s to 2.5 s	1280 W (640 W net change)
$P_1 + P_2$ (W)	2.5 s to 2.8 s	1480 W (200 W net change)
$P_1 + P_2$ (W)	2.8 s to 3.8 s	1020 W (460 W net change)
$P_1 + P_2$ (W)	3.8 s to 5.5 s	1120 W (100 W net change)

2168-6777 (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 May 22,2020 at 05:46:56 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

and again reduces to *0*. The idea is to obtain better power sharing and also to eliminate the low-frequency oscillations with the adaptive term. The variation of α can be seen in Fig. 5. The proposed adaptive control also mitigates the transients better than PI control as can be seen, comparing Fig. 6(a) and Fig. 6(b). The peak overshoot for the proposed control is lesser than that for PI control. Fig. 6(c) shows the frequency profile for the proposed control and Fig. 6(d) for PI control. Similar to the power response, the proposed control mitigate the transients in frequency in a better manner.

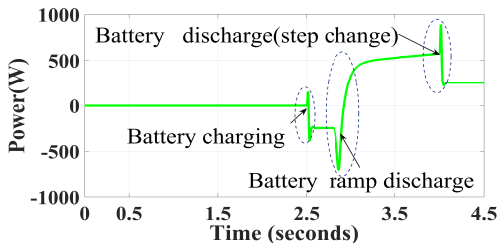
2) CPL transients: A *42V* lead-acid battery connected on the DC side of VSC_3 acts as CPL for the autonomous microgrid. At *t=2.5s*, CPL switching of *200W* (battery charging) occurs in the system. Another CPL switching of *460W* (battery discharging) occurs at *t=2.8s*. Fig. 7 validates the performance of proposed controller mitigating the CPL transients. The performance is validated to be better than the conventional PI control.



(a) Active power(proposed control) (b) Active power(PI)

(c) Frequency(proposed control) (d) Frequency(PI)

Fig. 6: MATLAB/simulink responses for transient events explained in Table III

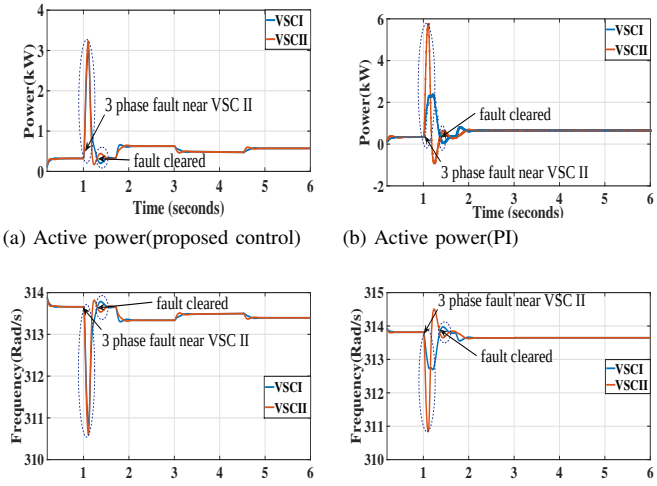


(a) CPL response

Fig. 7: MATLAB/Simulink responses of CPL operation explained in Table III

3) Fault transients: The case of transient mitigation during an AC faults contingency is discussed here. A 3ϕ fault occurs near VSC_2 for *100ms* at *t=1s* and is cleared at *t=1.1s* [see Fig. 8]. The contribution to the fault is dependent on the control employed. It can be seen that power-sharing has a better profile for the proposed control than for PI control. Fig. 8(a) shows both VSCs contributing *P=3.2 kW* whereas Fig. 8(b) shows VSC_2 contributing *P=5.8 kW* alone while VSC_1 contributing around *P=2 kW*. The frequency output of both VSCs are shown in Fig. 8(c) & (d). The proposed controller has better performance in terms of maintaining the same frequency to the distribution system. As it can be seen that during the transient, both VSC maintain the almost identical frequency when the proposed controller is applied. However, with the conventional PI controller VSC, I & VSC II have different frequency response during the transient as the drop in VSC II frequency output is higher than the VSC I frequency output. The settling time for the proposed control is also lesser than that for PI control. It means that for power restoration after a fault isolation, the proposed approach restores the system at *t=1.45s* whereas the conventional PI restores the system at *t=1.92s*.

However, the steady-state error is zero for the case of PI control, whereas the steady-state error is in a bound for the proposed approach. The proposed analysis gives uniform ultimate boundedness of the tracking errors. Hence, the error converges exponentially to a bound around the equilibrium point (zero).



(a) Active power(proposed control) (b) Active power(PI)

(c) Frequency(proposed control) (d) Frequency(PI)

Fig. 8: MATLAB responses for 3 phase ground fault.

V. Experimental Results

A. Experimental setup

The hardware validation of the proposed adaptive backstepping is carried out using an FPGA based NI- PXIe system. The scaled-down assembly for hardware validation is shown in Fig. 9 where the system shown in Fig. 1 has been implemented in a lab scale setup at *110V*. The two VSCs responsible for maintaining the voltage and frequency of the system and also for sharing the load proportionately are shown in the figure as VSC1 and VSC2. For constant power load (CPL), series combination of *12V* lead-acid batteries are connected on the DC side of a VSC through a bi-directional converter. The voltages and currents measurements are taken employing voltage sensors (LV25-P) and current sensors (LA55-P), shown in Fig. 9. Also, the RLC type loads connected to the system is shown as load banks in the figure. FPGA board (NI sbRIO 9606 and

2168-6777 (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 May 22,2020 at 05:46:56 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

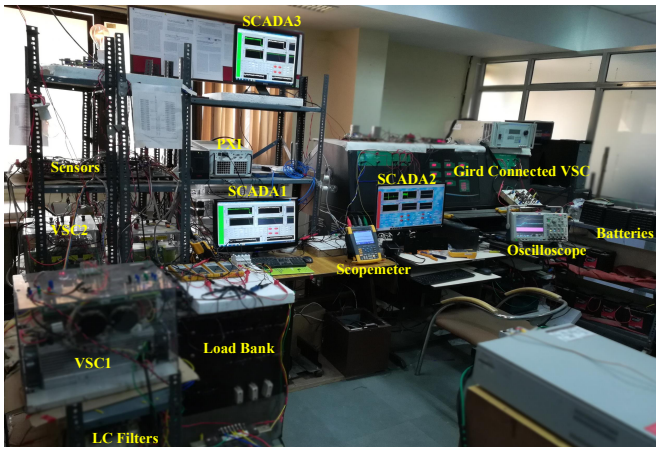
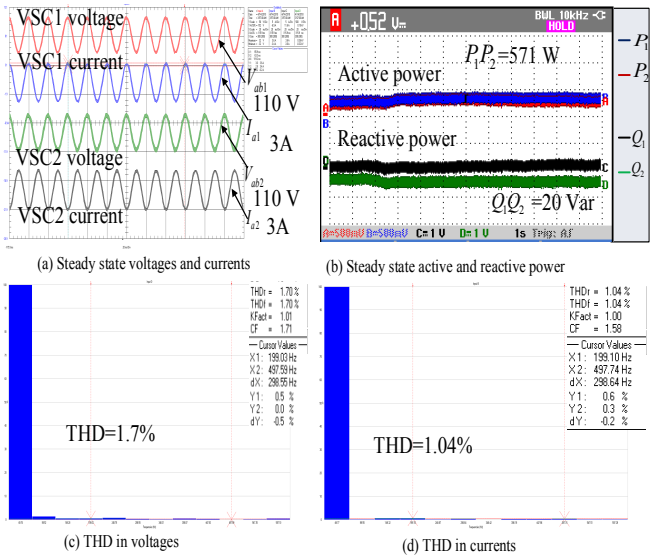


Fig. 9: Hardware setup for experimental validation

9683) is used for data acquisition and processing of signals. NI LabView is employed to design the control algorithms shown as SCADA1, SCADA2 and SCADA3 in Fig. 9.



(a) Steady state voltages and currents

(b) Steady state active and reactive power

(c) THD in voltages

(d) THD in currents

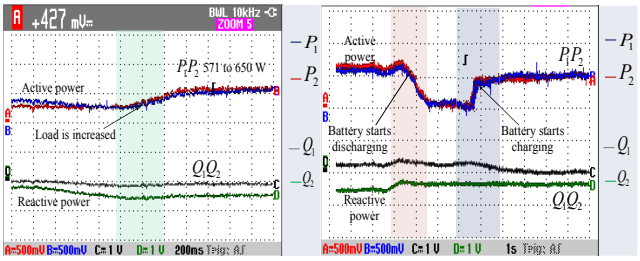
Fig. 10: Experimental Results (steady state)

B. Steady state response (VSC_1 & VSC_2)

Fig. 10(a) shows the voltage and current profile of VSCs running in parallel. Fig. 10(a) includes the VSC_1 and VSC_2 voltages and currents while the system is operating in steady-state. The output voltage of VSCs is at 110 V while both are supplying 3 A current to the loads connected to the system. The active and reactive powers of both VSC combined is shown in Fig. 10(b). The net active power transferred by VSC1 and VSC2 is of 571 W and reactive power transferred is of 20 Var. The VSCs share *571W* as seen in Fig. 10(b). Total Harmonic Distortion (THD) in AC phase voltage for the proposed backstepping control is 1.7% (see Fig. 10(c)) while THD in the current plot is 1.04%. (see Fig. 10(d)).

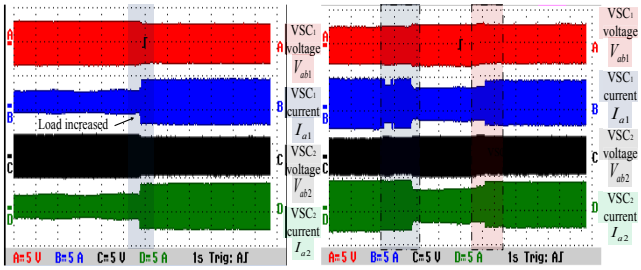
C. Dynamic state response (VSC_1 & VSC_2)

The performance of power-sharing during a load transient (*571W* to *650W*) is studied next. Fig. 11(a) shows the pro-



(a) Active and reactive power sharing in dynamic loading

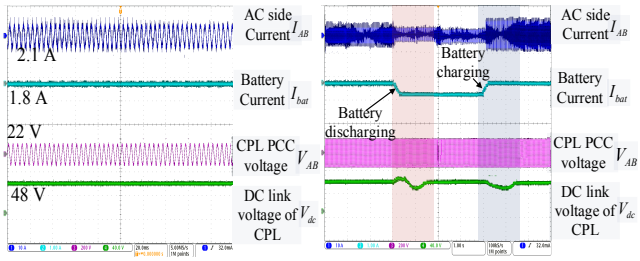
(b) Active and reactive power sharing in CPL loading



(c) Voltages and currents in increased loading

(d) Voltages and currents in CPL loading

Fig. 11: Experimental Results (load changes)



(a) Steady state operation of battery discharging

(b) Battery changed from charging to discharging

Fig. 12: Experimental Results (load changes)

portional power-sharing profile of both VSCs for the load transient. Fig. 11(c) shows the voltage and current profile of VSCs for such an operation. It can be seen that the current by each VSC increases from *3A* to *4.1A*. The loads are increased to make the loading dynamic, and it can be seen that the active power is increased from 571 W to 650 W and similarly the reactive power demand is also increased from 20 Var to 32 Var. However, this increased in active power demand is proportionally shared by both VSCs, which support the claim of the proposed control philosophy.

D. Dynamic state response with CPL (VSC_1, VSC_2 & VSC_3)

CPL operations in the form of battery charging and discharging are studied next. Three lead-acid batteries (*12V, 42Ah* each) in series are interfaced with a bi-directional converter operating in current control mode. The output of the bi-directional converter is connected to the DC link of VSC_3. The DC link of VSC is maintained at *48V*. Initially, battery discharges which reduce the power requirement (*571W* to *450W*) from the VSC set. The battery is made to charge next,

2168-6777 (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 May 22,2020 at 05:46:56 UTC from IEEE Xplore. Restrictions apply.

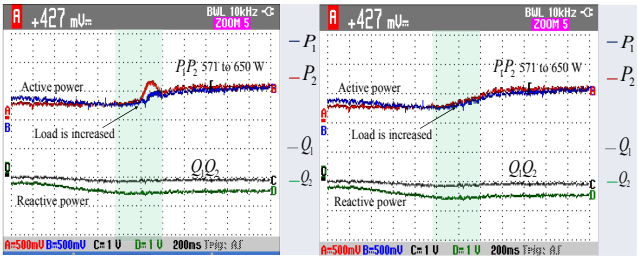
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

increasing power requirement from *450W* to *560W*. Fig. 11(b) shows the contribution of each VSC in the CPL operation. The current contribution drops from *3A* to *1.9A* when the battery starts to discharge. The current contribution starts to increase again from *1.9A* to *2.9A* when the battery starts charging. The sequence of events and the following current contributions are elaborated in Table IV. CPL dynamics has been presented in Fig. 12(a), 12(b). In steady-state, the battery is discharging at *1.8A*. A step-down transformer (*5:1*) is connected between distribution lines and VSC_3. The AC side voltage of VSC_3 is at *22V*.

TABLE IV: Events during real time CPL operation

Parameter	Time of operation	Value
$P_1 + P_2$ (W)	0.0 s to 1.5 s	571 W (285.5W each)(Fig. 9(b))
$P_1 + P_2$ (W)	1.5 s to 4.5 s	450 W (79W net change)(Fig. 9(b))
$P_1 + P_2$ (W)	4.5 s to 6 s	560 W (110W net change)(Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	0 s to 2.5 s	3A each (Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	2.5 s to 4.5 s	1.9A each (Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	4.5 s to 6 s	2.9A each (Fig. 9(b))

E. Comparision between PI based control and proposed control reponse (VSC_1 & VSC_2)



(a) Active power sharing (PI control) (b) Active power sharing (proposed control)

Fig. 13: Experimental Results (comparision between PI and proposed control)

Similar to the simulation results, a comparison between the PI-based control and the proposed control is carried out in real-time. The performance of power-sharing during a load transient (*571W* to *650W*) is studied next. Fig. 13(a) shows the proportional power-sharing profile of both VSCs for the load transient when PI control is applied for both VSCs. The loads are increased to make the loading dynamic, and it can be seen that the active power is increased from 571 W to 650 W. The same steps are carried out for the proposed adaptive droop based control to maintain the uniformity. The active power-sharing dynamics for the proposed control is shown in Fig. 13(b) The proposed adaptive backstepping control mitigates the transients better than PI control as can be seen in Fig. 13, as the peak overshoot for the proposed control, in case of VSC_2 is lesser than that for PI control.

VI. Conclusion

The work proposes an adaptive droop based control for active power-sharing between VSCs in an autonomous microgrid. The adaptive term introduced in the droop control acts as a feed-forward term and is responsible for the transient performance of the system. A Lyapunov based technique has been proposed, which ensures uniform ultimate boundedness (UUB) of the tracking errors. The focus of the work is on transient mitigation for different transients in the system. A thorough comparison of the proposed control with the conventional PI control is presented employing MATLAB/Simulink as the platform. AC load variations, Constant Power Load (CPL) switching, AC faults restoration transients are considered for the comparison. FPGA based hardware setup is employed to study the real-time performance of the proposed adaptive control.

References

[1] IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems," in *IEEE Std 1547.4-2011* , vol., no., pp.1-54, 20 July 2011.

[2] Q. Zhong, "Robust Droop Controller for Accurate Proportional Load Sharing Among Inverters Operated in Parallel," in *IEEE Transactions on Industrial Electronics,* vol. 60, no. 4, pp. 1281-1290, April 2013.

[3] J. He and Y. W. Li, "An Enhanced Microgrid Load Demand Sharing Strategy," in *IEEE Transactions on Power Electronics,* vol. 27, no. 9, pp. 3984-3995, Sept. 2012.

[4] R. Han, L. Meng, G. Ferrari-Trecate, E. A. A. Coelho, J. C. Vasquez and J. M. Guerrero, "Containment and Consensus-Based Distributed Coordination Control to Achieve Bounded Voltage and Precise Reactive Power Sharing in Islanded AC Microgrids," in *IEEE Transactions on Industry Applications,* vol. 53, no. 6, pp. 5187-5199, Nov.-Dec. 2017.

[5] T. Wu, Z. Liu, J. Liu, S. Wang and Z. You, "A Unified Virtual Power Decoupling Method for Droop-Controlled Parallel Inverters in Microgrids," in *IEEE Transactions on Power Electronics,* vol. 31, no. 8, pp. 5587-5603, Aug. 2016.

[6] T. Nguyen, H. Yoo and H. Kim, "A Droop Frequency Control for Maintaining Different Frequency Qualities in a Stand-Alone Multimicrogrid System," in *IEEE Transactions on Sustainable Energy,* vol. 9, no. 2, pp. 599-609, April 2018.

[7] G. Lou, W. Gu, Y. Xu, M. Cheng and W. Liu, "Distributed MPC-Based Secondary Voltage Control Scheme for Autonomous Droop-Controlled Microgrids," in *IEEE Transactions on Sustainable Energy,* vol. 8, no. 2, pp. 792-804, April 2017.

[8] X. Huang, K. Wang, J. Qiu, L. Hang, G. Li and X. Wang, "Decentralized Control of Multi-Parallel Grid-Forming DGs in Islanded Microgrids for Enhanced Transient Performance," in *IEEE Access,* vol. 7, pp. 17958-17968, 2019.

[9] D. E. Olivares et al., "Trends in Microgrid Control," in *IEEE Transactions on Smart Grid,* vol. 5, no. 4, pp. 1905-1919, July 2014.

[10] H. R. Baghaee, M. Mirsalim, and G. B. Gharehpetian, Real-time verification of new controller to improve small/large-signal stability and fault ridethrough capability of multi-DER microgrids, *IET Gener. Transmiss. Distrib.,* vol. 10, no. 12, pp. 30683084, Sep. 2016.

[11] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetian, and H. A. Talebi, A generalized descriptor-system robust H control of autonomous microgrids to improve small and large signal stability considering communication delays and load nonlinearities, *Int. J. Elect. Power Energy Syst.,* vol. 92, no. 1, pp. 6382, Nov. 2017.

[12] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetan and H. A. Talebi, "Nonlinear Load Sharing and Voltage Compensation of Microgrids Based on Harmonic Power-Flow Calculations Using Radial Basis Function Neural Networks," in *IEEE Systems Journal,* vol. 12, no. 3, pp. 2749-2759, Sept. 2018.

[13] H. R. Baghaee, M. Mirsalim and G. B. Gharehpetian, "Power Calculation Using RBF Neural Networks to Improve Power Sharing of Hierarchical Control Scheme in Multi-DER Microgrids," in *IEEE Journal of Emerging and Selected Topics in Power Electronics,* vol. 4, no. 4, pp. 1217-1225, Dec. 2016.

[14] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetian and H. A. Talebi, "Unbalanced harmonic power sharing and voltage compensation of microgrids using radial basis function neural network-based harmonic power-flow calculations for distributed and decentralised control structures," in *IET Generation, Transmission & Distribution,* vol. 12, no. 7, pp. 1518-1530, 10 4 2018.

[15] M. B. Delghavi and A. Yazdani, "Sliding-Mode Control of AC Voltages and Currents of Dispatchable Distributed Energy Resources in Master-Slave-Organized Inverter-Based Microgrids," in *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 980-991, Jan. 2019.

2168-6777 (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 May 22,2020 at 05:46:56 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/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

[16] M. Cucuzzella, G. P. Incremona and A. Ferrara, "Decentralized Sliding Mode Control of Islanded AC Microgrids With Arbitrary Topology," in *IEEE Transactions on Industrial Electronics,* vol. 64, no. 8, pp. 6706-6713, Aug. 2017.

[17] D. O. Amoateng, M. Al Hosani, M. S. Elmoursi, K. Turitsyn and J. L. Kirtley, "Adaptive Voltage and Frequency Control of Islanded Multi-Microgrids," in *IEEE Transactions on Power Systems,* vol. 33, no. 4, pp. 4454-4465, July 2018.

[18] T. Hosseinalizadeh, H. Kebriaei and F. R. Salmasi, "Decentralised robust T-S fuzzy controller for a parallel islanded AC microgrid," in *IET Generation, Transmission & Distribution,* vol. 13, no. 9, pp. 1589-1598, 7 5 2019.

[19] Graeme Heald "Issues with Reliability of Fuzzy Logic" Published in *International Journal of Trend in Scientific Research and Development (ijtsrd)*, ISSN: 2456-6470, Volume-2 — Issue-6, October 2018, pp.829-834,

[20] H. K. Khalil, Nonlinear Systems, 2nd ed., *Englewood Cliffs,* NJ, USA: Prentice-Hall, 1996.

[21] Yaolong Tan, Jie Chang, Hualin Tan and Jun Hu, "Integral backstepping control and experimental implementation for motion system," in Proceedings of the 2000 *IEEE International Conference on Control Applications,* Anchorage, AK, USA, 2000 pp. 367,368,369,370,371,372.

[22] D. Rajicic and A. Bose, "A modification to the fast decoupled power flow for networks with high R/X ratios," in *IEEE Transactions on Power Systems,* vol. 3, no. 2, pp. 743-746, May 1988.

[23] A. Emadi, A. Khaligh, C. H. Rivetta and G. A. Williamson, "Constant power loads and negative impedance instability in automotive systems: definition, modeling, stability, and control of power electronic converters and motor drives," in *IEEE Transactions on Vehicular Technology,* vol. 55, no. 4, pp. 1112-1125, July 2006.

[24] A. Firdaus and S. Mishra, "Mitigation of Power and Frequency Instability to Improve Load Sharing Among Distributed Inverters in Microgrid Systems," in *IEEE Systems Journal,* vol. 14, no. 1, pp. 1024-1033, March 2020.

[25] M. Yazdanian and A. Mehrizi-Sani, "Internal Model-Based Current Control of the RL Filter-Based Voltage-Sourced Converter," in *IEEE Transactions on Energy Conversion,* vol. 29, no. 4, pp. 873-881, Dec. 2014.

[26] P. D. Achlerkar, V. Nougain and B. K. Panigrahi, "Backstepping Technique based Customer-end Voltage Control Strategy of DC Distribution Network," in *IEEE Journal of Emerging and Selected Topics in Power Electronics.*

Surya Prakash received the B.Tech. degree in Electrical Engineering from School of Engineering, CUSAT, Kochi, India in 2009 and the M.Tech. degree in Power System from the NIT Trichy, Tamilnadu, India, in 2015. He has worked as Design Engineer at Hydro Machinery Development Station, BHEL, Bhopal India for 3 years. He is currently working toward the Ph.D. degree at Electrical Engineering Department, Indian Institute of Technology, Delhi, India.

His current research interests include control of grid connected inverter system under varying grid conditions.

Vaibhav Nougain Vaibhav Nougain received B.Tech degree in electrical engineering in 2017 from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India. He is currently working toward the doctoral degree at the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India. He was a Visiting Student with The Center for Electric Power and Energy (CEE), Technical University of Denmark in 2019. His research interests include protection and control of DC and AC microgrids, islanding detection, and protection of HVDC systems.

Sukumar Mishra (M'97-SM'04) received the B.Sc (Engg) degree in electrical engineering from University College of Engineering Burla, Odisha, India, in 1990, and the M.Sc (Engg) and Ph.D. degrees in electrical engineering from National Institute of Technology Rourkela, Odisha, India, in 1992 and 2000, respectively.

He a Professor with the Indian Institute of Technology Delhi, New Delhi, India, and has been a part of the institute for the past 15 years. He has authored or coauthored over 200 research articles (including papers in international journals, conferences, and book chapters). His research interests include power systems, power quality studies, renewable energy, and smart grid.

Prof. Mishra is a recipient of many accolades throughout his academic tenure of over 25 years. He has been a recipient of Indian National Science Academy Medal for Young Scientist (2002), Indian National Academy of Engineering (INAE) Young Engineer Award (2009), INAE Silver Jubilee Young Engineer Award (2012), and has recently won the Samanta Chandra Shekhar Award (2016). He has been granted fellowship from many prestigious technical societies like IET (U.K.), NASI (India), INAE (India), IETE (India), and IE (India) and is also recognized as the INAE Industry Academic Distinguish Professor. He is currently serving as an Editor for the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, IEEE TRANSACTIONS ON SMART GRID, and an Associate Editor for the IET Generation, Transmission, & Distribution journal. Currently, he is the Vice Chair of Intelligent System Subcommittee of Power and Energy society (PES) of IEEE. Apart from all research and academic collaborations, he is very actively involved in industrial collaborations. He is currently acting as INAE Chair Professor and was previously delegated as the NTPC and Power Grid Chair Professor. He is also serving as an Independent Director of Cross Border Power Transmission Company Ltd., Gurugram, India, and River Engineering Pvt. Ltd., Greater Noida, India.

2168-6777 (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 May 22,2020 at 05:46:56 UTC from IEEE Xplore. Restrictions apply.

Asian Journal of Business Ethics
https://doi.org/10.1007/s13520-020-00102-x

Analyzing the inter-relation between workplace spirituality and constructive deviance

Naval Garg [1] · Anubhuti Saxena [2]

Received: 29 October 2019 / Accepted: 1 April 2020/Published online: 18 May 2020
© Springer Nature B.V. 2020

Abstract

Researchers advocate that workplace spirituality has the potential to increase "constructive deviant behavior" among employees across different types of organizations and professions by engaging individuals in meaningful ways. This research examines the link between workplace spirituality and constructive deviant behavior. Literature on the issue of workplace spirituality suggests that meaningful work has the potential to increase positive organizational outcomes. This study was carried out on a purposively selected sample of 152 managers from the oil and gas industry in India. The paper investigates the relationship between workplace spirituality and constructive deviance with the help of a wide range of statistical tools including mean, factor analysis, correlation, and regression. Six Indian constructs of workplace spirituality were examined to assess the relationship between workplace spirituality and constructive deviance.

Keywords Workplace spirituality · Constructive deviance · Sense of community · *Karma* capital

Introduction

It is a well-known fact that Jesus Christ, the Son of God, was crucified for not conforming to the most popular beliefs and values of his time. The famous philosopher, Socrates, was also sentenced to death for impiety. Lord Buddha, who left his kingdom

✉ Naval Garg
naval.garg@dtu.ac.in

Anubhuti Saxena
anubhuti.jhingran88@gmail.com

[1] University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

[2] JIMS, Greater Noida, India

and his family at the age of only twenty-nine years in search of elusive knowledge about the world, was initially criticized for this action by his own father and the people of his own state. Mahavir Jain, the founder of Jainism, gave up his clothes for which he was ostracized from a number of villages and attacked with stones. Mahatma Gandhi proposed and practiced "non-violence" as an effective tool for gaining independence from foreign rule, and while doing so, deviated from the most common practice of perpetration of bloodshed to foment a revolution for earning independence. In the contemporary world, the founder of Microsoft, Bill Gates, has donated most of his wealth amid the materialism and blind race for the accumulation of wealth that characterizes the twenty-first century. All the above-mentioned incidences are examples of what can be described as "deviant behaviors" of historical characters, which are following the modern day definition of "constructive deviance."

According to Spreitzer and Sonenshein (2003), constructive deviance represents "intentional behaviors that depart from the norms of a referent group in honorable ways." Further, according to Warren (2003), constructive deviance signifies "behavior that deviates from the reference group norms but conforms to hyper-norms." Donaldson and Dunfee (1994) defined hyper-norms as the beliefs and values that are globally accepted. Galperin (2003), on the other hand, defined constructive deviance as "voluntary behavior that violates significant organizational norms and in doing so contributes to the well-being of an organization, its members, or both." Most recently, Vadera et al. (2013) described constructive deviance as "[a set of] behaviors that deviate from the norms of the reference group such that they benefit the reference group and conform to hyper-norms." Based on the various prevalent definitions of "constructive deviance," Vadera et al. (2013) delineate and discuss the following three main characteristics associated with this concept:

a. Deviation from the standard of the reference group;
b. Gain of the reference group; and
c. Behavior that must be conventional to hyper-norm.

As the above examples show, all the protagonists, including Jesus Christ, Lord Buddha, Mahatma Gandhi, and Mahavir Jain, deviated from their respective groups but conformed to the hyper-norms of humanity, love, affection, and welfare of all. Even after the passage of hundreds of years, their teachings and directives are benefiting thousands of people around the world. There is not even a single iota of doubt about the utility and significance of their teachings. One of the underlying communalities of such deviant behavior is the pursuit of spiritualism. These super-human characters were spiritually enriched, and through their deviant behavior, they, in a way, redefined the relationship of human beings with supernatural powers.

A number of researchers, including Ahmad and Omar (2014), Krishnakumar and Neck (2002), Cropanzano et al. (2007), Bennett and Robinson (2003), Ehtesham (2011), Gabriela (2013), Kelloway (2010), Lian et al. (2012), Paula and Sofia (2003), Penney and Spector (2005), and Prasanna and Madhavaiah (2017), have reported that the concept of "workplace spirituality" helps to counter deviant behavior in the workplace. Workplace spirituality has been reported to work positively for a wide range of deviant behaviors in the workplace, such as the playing of pranks among employees, the act of falsifying, theft, office politics, corruption, sabotaging, making

fun of colleagues, violating organizational policies, and manipulation in the workplace. Although some previous studies have investigated the inter-relation between workplace spirituality and deviant behavior at work, the relationship between workplace spirituality and constructive deviance is yet to be properly explored. This paper attempts to plug this gap in research by investigating the linkages between six Indian constructs of workplace spirituality, that is, *Swadharma*, *Lokasangraha*, sense of community, authenticity, *karma* capital, and *Kritagayata*, and constructive deviance. Through this research, the authors aim to explore how people indulge in constructively deviant behavior as a manifestation of the way in which they recognize and understand their organizations in terms of various spiritual dimensions based on Indian spiritual traditions. The main goal of this paper is to explore the inter-relation between the concepts of workplace spirituality and constructive deviance in the context of the Indian oil and gas industry.

The article is broadly divided into four sections. The first section delineates the concept of constructive deviance. The second section describes the concept of workplace spirituality and its dimensions in the Indian context. The third section of the article uses correlation and regressions to explore the relationship between workplace spirituality and constructive deviance. This paper concludes with the fourth and final section, which highlights the theoretical and practical implications of the study, along with its limitations and scope for further research on the subject.

Literature review

Constructive deviance

The concept of workplace deviance has gained enormous attention by researchers in recent years and has become an issue of great concern for modern organizations (Bennett and Robinson 2003). A major part of the literature on the topic conceptualizes deviance with negative connotations. Workplace deviance is mostly considered against the norms of the organization and is also perceived to be damaging for the organization (Bord 1976). Previous research studies have argued that any deviation from organizational norms may be harmful (Mitchell and Ambrose 2007; Robinson and Bennett 1995). However, deviating from standard norms could also prove to be a blessing in disguise for the organization (Galperin 2003; Spreitzer and Sonenshein 2004). In an era where "no innovation is new," higher degrees of creativity and innovativeness are desired from the employees (Zhang and Bartol 2010). Thus, there are possibilities of employees deviating from the established paths to become more creative, better performers, and a more relevant workforce in an unpredictable and abruptly changing world. Such beneficial deviation has been termed as "constructive deviance," which has been defined in many ways but which reflects similar underlying patterns, intentions, and dimensions (Vadera et al. 2013). As discussed above, constructive deviance is argued to have three basic dimensions: deviation from the standard of the reference group, benefit of the reference group, and behavior following hyper-norms. It is symbolic of intentional and voluntary conduct that disregards and disobeys significant organizational norms but subsequently helps in the betterment of an organization and its members.

Such deviant behaviors have been divided into two major categories: *interpersonal constructive deviance* and *organizational constructive deviance* (Bodankin and Tziner 2009). Interpersonal constructive deviance is directed towards an individual, like a colleague, subordinate, or superior, and involves deviant behavior such as disobedience towards a boss to improve organizational performance. Organizational constructive deviance is directed towards an organization, and it may challenge or even break organizational rules, processes, and procedures (Bodankin et al. 2009). Constructive deviance is rapidly gaining acceptance among both researchers and its practitioners because of its potential to act as a harbinger of organizational change (Robbins and Galperin 2010). Researchers have argued that constructive deviance is a wide concept and covers different types of behaviors such as creative performance (Amabile et al. 1996; Oldham and Cummings (1996); issue selling (Dutton and Ashford 1993); extra-role behaviors (Van Dyne et al. 1995); pro-social rule-breaking (Morrison 2006); taking charge; and counter-role behavior. Constructive deviance also embraces concepts such as organizational citizenship behavior (OCB), whistle-blowing, and role innovation.

Creative performance is defined as the generation of innovative and valuable ideas to resolve organizational problems (Oldham and Cummings 1996). Issue selling signifies "voluntary behaviors which organizational members use to influence the organizational agenda by getting those above them to pay attention to an issue" (Dutton and Ashford 1993, pp. 398). Extra-role behavior is "behavior which benefits the organization and is intended to benefit the organization, which is discretionary and which goes beyond existing role expectations" (Van Dyne et al. 1995, pp. 218). Pro-social rule-breaking is "intentional violation of a formal organizational policy, regulation, or prohibition with the primary intention of promoting the welfare of the organization or one of its stakeholders" (Morrison 2006, pp. 6). OCBs represent work behaviors that are useful for optimizing organizational functioning though OCBs are not critical or indispensable for the performance of assigned tasks like offering a helping hand to colleagues and supervisors, and not availing of excessive breaks (Borman and Motowidlo 1993; Organ 1988).

Whistle-blowing behavior is commonly referred to as the disclosures made by one or more employee(s) of some illegal or illegitimate act done by some other employee in the organization in their official capacity, and which may have a detrimental effect on organizational growth and development (Vadera et al. 2013). All these organizational behaviors explicitly encompass (a) deviation from norms and (b) well-being of the reference group.

In a meta-analysis of studies of constructive deviance published in the *Journal of Management*, Vadera et al. (2013) theorized about "an emergent model of constructive deviance." The model proposes various antecedents, mediators, and constructs of constructive deviance based on previous empirical and theoretical researches. The different predictors include employee characteristics (innovative cognitive style, self-worth, efficacy, extroversion), supervisor-focused characteristics (transformational leadership, leader-membership exchange), job-focused characteristics (positive job attitude), organization-focused characteristics (organizational culture and climate, organizational support), and group-focused characteristics (attachment to the group, support of co-workers). The authors also encourage future researchers to examine the model and its part empirically. Based on the same fundamentals, the present study tends to investigate the relationship between workplace spirituality and constructive deviance. The relevant literature on workplace spirituality has been discussed below.

Workplace spirituality

It has been found that with time, the elemental structure in the development of workplace spirituality has undergone an exemplary transformation. Workplace spirituality can be broadly defined as amalgamation of live experiences and expressions of one's spirituality in the context of work and the workplace. It not only relates to personal spirituality but also focuses upon the spiritual environment provided by the organization for its employees. Some authors have broadened the notion by arguing that workplace spirituality is all about interconnectedness, not only with other people but also with the environment and the entire universe at large (Comer 2008). Ashmos and Duchon (2000) defined the concept of spirituality at work (SAW) as a "phenomenon through which individuals can gratify their inner selves by doing meaningful work while simultaneously experiencing the feeling of working in a community." Milliman et al. (2003) extended the definition of "work" propounded by Ashmos and Duchon and defined a spiritual workplace as a place where individuals find meaning in work and contentment by working in a community. According to them, the third component in workplace spirituality should be an alignment with organizational values rather than the inner life, as they perceived inner life to be highly personal. Petchsawang and Duchon (2009) claimed that individuals perform meaningful work at the spiritual workplace, where they experience a feeling of compassion, transcendence, and stay mentally aware. Workplace spirituality appears to be an ephemeral phenomenon replete with multiple perspectives, and, consequently, contrasting with the neat containment and normalization of feelings and perceptions as expounded by many proponents of the new paradigm (Case and Gosling 2010). Giacalone and Jurkiewicz (2004) defined workplace spirituality as "a framework of organizational values evidenced in the culture that promotes employees' experience of transcendence through the work process, facilitating their sense of being connected to others in a way that provides feelings of completeness and joy." "Spiritual Intelligence is the intelligence through which we understand our earnest meanings, purposes, and highest motivations" (Zohar and Marshall 2000). Amid the multiple definitions of workplace spirituality, Karakas (2010) found 70 more and different definitions in several research papers. This wide diversity in defining workplace spirituality makes it a difficult area to explore and comprehend (Kinjerski and Skrypnek 2004). Several researchers (Dehler and Welsh 2003; George and Mallery 2003; Tepper 2003) accepted the fact that workplace spirituality (WSP) is a multidimensional phenomenon that inherently assumes a meaningful and purposeful job experience, aligned with the authenticity of work and a sense of belongingness and interconnection felt by the worker. The rationale behind this definition is that organizations have now started following the spiritual dimension, which deals less with rules and order, and revolves more around meaningful work and a sense of community (Armour 1998; Galen and West 1995). The concept has also been explored by using different constructs and variables because of the prevalence of the diverse definitions, understanding, and the culture-specific nature of workplace spirituality (Garg 2017).

Indian constructs of workplace spirituality

Swadharma

The term *Swadharma* consists of the words *swa* and *dharma*. *Swa* can be said to be the self while *dharma* can be referred to as righteous action or lawful conduct. In both the

traditional and modern settings, *Swadharma* usually alludes to one's *dharma*, duties, and responsibilities. *Swadharma* is unique to every individual because everyone has different capacities and a sense of morality. Lord Krishna said that the act of *Swadharma*, even if entrenched with imperfection, is better than someone else's *dharma* entrenched with perfection. It is better to depart from this life while performing *Swadharma* as the performance of someone else's *dharma* only resonates with fear and uncertainty (Dutt 2018).

Karma Capital

One of the important philosophical Indian spiritual texts, *The Bhagavad Gita*, calls attention to the significance of action or *karma yoga* and promotes it as a way of life for attaining salvation (or yoga). *Karma* is a Sanskrit term and draws its meaning from the root word *kri*, which means doing or an activity. According to Tilak (1915/2000), the *Gita* mainly talks about *karma yoga* as having three connotations: (i) as a skill for acting (*Gita* Chapter 2, verse 50), (ii) as a calmness of mind towards the results of one's actions (*Gita* Chapter 2, verse 48), and (iii) as a device for eradicating a sense of attachment (*Gita* Chapter 2, verse 50). Even Buddhism talks about the concept of *karma* and views it as an action. It is defined as an "overall psychological impulse behind an action, that which sets going a chain of causes culminating in a *karmic* fruit" (Harvey 2000, p. 17).

As per the Indian viewpoint, no one can stay even for a moment without doing work, and it is only by performing an action that a person accomplishes the highest satisfaction, but the determination in action is needed without much thought of the fruit (Chakraborty 1987; Radhakrishnan 1926). Mulla and Krishnan (2007) established that employees high on the performance of *karma yoga* are more prone to follow "others-oriented" (rather than "self-oriented") terminal values such as ensuring that the "world is at peace," and such individuals are regarded as more responsible and obedient than those who do not believe in *karma yoga*.

Lokasangraha

The term "*Lokasangraha*" is composed of two words: *lok*, which means the world, and *sangraha*, which means holding together. The message of the *Bhagavad Gita* can help in smoothening of the progress of the dimension of interconnectedness. *Lokasangraha* means the "interconnectedness of society" (Radhakrishnan 1970). The society as a whole relies on the interdependence of its citizens, including individuals, to function properly. As per several Indian scholars, the concept of *Lokasangraha* is a vital example of signifying action and performance in the modern corporate world (Chakraborty 1995). It entails that all the actions performed by individuals and organizations should focus on ensuring only the well-being of both societies as well as its stakeholders (Garg 2017). Further, in today's world characterized by business dealings and values, it is extremely important to deal with the concerns of justice and positive relations, for which both employees and organizations must attempt to propagate the wisdom of the spiritual tradition of *Lokasangraha* to the corporate world (Chattopadhyay and Agrawal 1977). This would be possible by promoting thoughts of selfless service and concern for the well-being of all employees (*sarva-bhuta-*

hiteratah), (Chapter 12, verse 4, *Bhagavad Gita*) and by encouraging employees to uphold the same mindset towards all their colleagues (*samah sarveshu bhuteshu*) (Chapter 18, verse 54, *Bhagavad Gita* "As It Is," p. 738).

Authenticity

The concept of authenticity is used both in psychology as well as existentialist philosophy. It refers to the harmonization of the actions and behaviors of the executives in an organization with their central and internalized cultural values and belief systems (Pandey et al. 2016). It is what other people see in you, and can, therefore, be controlled by you, to a great extent. Had it been purely an innate quality, there would be very little one could do to manage it. If a person is true and honest to one's self as well as to others, he is being authentic (Garg 2017). Authenticity would be operational in the workplace if the employees are sincere, open, and supportive to fellow workers. It facilitates an environment devoid of politics, biasness, misuse, harassment, and prejudice.

Sense of community

McMillan and Chavis (1986) described the sense of community as a feeling of belongingness among members, a shared belief that if all members are committed to be together, the needs of each one of them would be fulfilled. According to Jurkiewicz and Giacalone (2004), employees develop a sense of community if they achieve interdependence and interconnectedness among themselves. This sense of community can be developed within organizations with the help of a shared vision and mission, and the perpetration of core organizational values (Naylor et al. 1996). When an employee perceives him/her as an indispensable cog in the wheel of the organization, it results in the development of a true sense of community (Duchon and Plowman 2005).

Kritagayata

The term *kritagayata* refers to a feeling of gratitude to the divine. In India, the attainment of spiritual values represents the pinnacle or the most advanced and extended version of gratefulness. Indian values honor every aspect of life, be it animate or inanimate (Garg 2018; Garg and Gera 2019). Apart from honoring God in the human form, the Indian tradition also adores and reveres God in animal form (such as a cow or a snake) and also in non-living forms (such as water, fire, and earth). A feeling of gratefulness for every form of life is the basis of spiritualism in India (*Shrimad Bhagwad Gita*, 18.2).

Workplace spirituality and work attitudes (constructive deviance)

Sufficient evidence is available to authenticate the positive relationship between workplace spirituality and work attitudes such as commitment and engagement (Pradhan and Jena 2016); job involvement (Kolodinsky et al. 2008); job satisfaction (Madhav and Kumar 2016; Garg 2017; Altaf and Awan 2011); organization citizenship behavior (Garg et al. 2019); and innovative work behavior (Afsar and Rehman 2015). Also, quite a few other studies seek to

investigate the relationship that workplace spirituality has with organizational performance (Giacalone and Jurkiewicz 2004; Duchon and Plowman 2005; Garg 2017), and with organizational development (Pawar 2008). Other studies have also investigated and found an encouraging association between WSP and other dimensions of the workplace like leadership perspective (Sengupta 2010); leadership effectiveness (Abdullah et al. 2009); knowledge-sharing intention and transformational leadership (Khari and Sinha 2017); emotional intelligence (Pradhan and Jena 2016); team learning (Pandey et al. 2016; Garg and Saxena 2019); employee well-being (Garg 2018); and low incidences of absenteeism (Chawla and Guda 2010; Pawar 2009). It is quite evident from past research that workplaces practicing spirituality have not only satisfied and more involved employees but also a more committed and productive workforce. These engaged, motivated, and productive employees become the harbingers of better organizational performance.

Although the relationship between workplace spirituality and a variety of work attitudes has been sufficiently investigated, the relationship between workplace spirituality and constructive deviance has not been as thoroughly explored. Despite the availability of theoretical evidence, our perception of the role played by workplace spirituality in determining and defining work attitudes such as constructive deviance remains narrow. We need a clearer depiction of why these linkages exist to develop a deeper insight into the psychological mechanisms involved in them. Spirituality is indisputably a human need for many people (Hart and Brady 2005), and a "reality" that should not be overlooked either by the society or by organizations (Judge 1999; Sanders et al. 2003). This research thus signifies a significant effort to add to the literature on workplace spirituality and the several ways in which it influences the management, employees, and the corporate organization, as a whole.

Research methodology

Objectives, sample, and procedures

The primary objective of this research paper is to investigate the linkage between workplace spirituality and constructive deviance. Also, it explores the status of constructive deviance and workplace spirituality in the oil and gas industry of India. The data for the study has been collected by surveying 152 managers working in various oil and gas companies in India like Reliance Petrochemicals, Bharat Petroleum, Cairn India, Oil, and Natural Gas Corporation, and Indian Oil Company Limited, using the convenience sampling techniques. The oil and gas industry is purposefully chosen for this study. Being a capital-intensive industry, it demands constructively deviant behavior like organizational citizenship behavior, whistle-blowing, and pro-social behavior. Convenience sampling is most appropriate in this study since data has to be collected from both off-shore and on-shore employees. Structured questionnaires were administered to the respondents using hard copies and emails. The demographic distribution of the sample is depicted in Table 1.

Table 1 Demographic profile of the sample

Variable	Category	Total	Percentage (%)
Gender	Male	100	65
	Female	52	35
Marital status	Married	87	57
	Unmarried	65	43
Age	Less than 30 years	40	25
	30 to 40 years	63	41
	40 to 50 years	31	20
	More than 50 years	18	14
Experience	0–10 years	84	55
	10–20 years	40	28
	20–30 years	20	13
	More than 30 years	08	5

Source: primary data

Measures

Workplace spirituality

Workplace spirituality is measured using the 18-item questionnaire of Pandey et al. (2009) and Garg (2017). Four constructs of workplace spirituality, that is, *swadharma*, *lokasangraha*, sense of community, and authenticity, were measured with the help of 11 statements taken from Pandey et al. (2009). The other two dimensions of workplace spirituality, that is, *karma* capital and *kritagayata*, were measured using seven statements presented by Garg (2017). A 5-point rating scale, ranging from 1 ("strongly disagree") to 5 ("strongly agree") was used to rate all the statements. The combined questionnaire was subjected to reliability and validity tests (Garg 2018; Garg et al. 2019). The spirituality scale includes six dimensions of workplace spirituality: *Swadharma* (for example, "Working here is a means for realizing my real self," "Work itself is enjoyable for me"), authenticity (for example, "People here own up to mistakes with others in the group"), *Lokasangraha* (for example, "People here try to avoid wastage of any kind," "People here are concerned about the natural environment while working here"), sense of community (for example, "When stuck with a problem, people here feel free to ask for advice from colleagues/advice from a superior/help from their colleagues/help from a superior"), *karma* capital (for example, "I put my 100 percent in my work," "Working itself provides me satisfaction"), and *kritagayata* ("I am grateful for being a part of this organization," "My job makes me understand that I have numerous things to cherish in my life").

Constructive deviance

The constructive deviance is assessed with the help of ten statements used by Galperin (2012). The questionnaire includes statements like "Bend or break the rules to perform

your job" and "I violated company procedures to solve a problem," which were rated on a 7-point rating scale. This questionnaire has been referred to by several other authors like Yildiz et al. (2015). These statements are rated on a 7-point rating scale.

Reliability and validity

Reliability is the consistency of results under similar conditions (Malhotra and Das 2011). The present study investigated internal consistency reliability using Cronbach's alpha. The proposed reliability estimates measure split-half correlation through splitting items into two sets, and Cronbach's alpha represents mean of all possible split-half correlations for a set of items. George and Mallery (2003) proposed a threshold value of 0.70 of Cronbach's alpha for statistically significant reliability. Table 2 shows acceptable values of internal reliability estimates (*Swadharma* = .922, *Lokasangraha* = .845, sense of community = .786, authenticity = .888, *kritagayata* = .823, *karma* capital = .712). Later on, Cronbach and Shavelson (2004) suggested confirming reliability with composite reliability (CR) along with Cronbach's alpha. CR could be calculated using the below-mentioned formula. Again, CR estimates are desired to be greater than 0.70 (Hair et al. 2010). Table 2 reports adequate values of CR (*Swadharma* = .867, *Lokasangraha* = .931, sense of community = .810, authenticity = .912, *kritagayata* = .850, *karma* capital = .754).

$$\mathrm{CR} = \left(\sum \lambda_i\right)^2 / \left(\sum \lambda_i\right)^2 + \left(\sum 1 - \lambda_i^2\right)$$

where (lambda) is the standardized factor loading for an item.

As far as validity is concerned, it is defined as the degree to which an instrument actually measures what it is supposed to. Here, convergent validity is examined with the help of average variance extracted (AVE), which is calculated using the below-mentioned formula. An AVE value of .5 or higher indicates an adequate level of convergent validity (Fornell and Larcker 1981). Values of AVE presented in Table 2 confirm validity.

$$\mathrm{AVE} = \left(\sum \lambda^2{}_i\right)/N$$

Table 2 Descriptive statistics, reliability, validity, and multicollinearity estimates

Variable	Mean	SD	α	CR	AVE	VIF
Karma capital	3.47	0.832	0.712	0.754	0.700	4.55
Authenticity	2.87	0.748	0.888	0.912	0.842	5.67
Sense of community	3.86	0.783	0.786	0.810	0.838	4.75
Lokasangraha	2.99	0.841	0.922	0.931	0.868	3.93
Swadharma	3.16	0.854	0.845	0.867	0.745	4.80
Kritagayata	3.90	0.836	0.823	0.850	0.821	5.96
CD	4.52	1.004	0.788	0.794	0.838	---

SD standard deviation, α Cronbach's alpha, *CR* composite reliability, *AVE* average variance extracted, *VIF* variance inflation

Source: primary data

where (lambda) is the standardized factor loading for an item and *N* is the number of items in a factor.

Multicollinearity and common method variance

Now, the problem of multicollinearity is assessed using variance inflation factor (VIF) (Kline 1998). Multicollinearity represents a state of high inter-correlation and inter-association between independent variables. VIF values (see Table 2) reveal that this study does not have a statistically significant problem of multicollinearity. Also, common method biases arise because of use of common method for data collection. It has been categorized in the following four categories: common rater effects, items context effect, measurement effect, and items characteristics effect. Common method variance (CMV) may lead to serious issues of inflated or deflated estimates of the relationship between two constructs, poor reliability and validity estimates, hypothesis testing error, and higher type I and type II errors (Podsakoff et al. 2012). Both procedural and statistical remedies are used to minimize the biasness. Firstly, pre-validated questionnaires were used to collect information regarding both independent and dependent variables. Secondly, questionnaires did not seek names of the respondents. Also, respondents were assured that procured data will remain confidential and will only be used for academic purpose only. According to Podsakoff et al. (2012), anonymity of respondents helps in minimizing CMV. Thirdly, different rating scales—5-point rating scale for the independent variable and 7-point rating scale for the dependent variable—are used to collect the data. Further, Harman's single-factor method was used to check common method biases (Chang et al. 2010). When all items are subjected to factor analysis, several factors were extracted, which means common method biases do not exist in this study.

Methods and procedure

The ten statements of constructive deviance were subjected to factor analysis using varimax rotation. Prior to factor analysis, the Kaiser-Meyer-Olkin (KMO) and Bartlett tests were administered to investigate the adequacy of sample for factor analysis. The relationships between the extracted factors and the six dimensions of workplace spirituality have been explored with the help of the correlation and multiple regressions.

Result and analysis

Table 2 shows the descriptive statistics including the mean and standard deviation of the various items of the study. As far as workplace spirituality is concerned, employees of the Indian oil and gas industry rated *kritagayata* (mean = 3.90) as the dimension of workplace spirituality associated with maximum experience. *Kritagayata* implies a feeling of gratitude and gratefulness for whatever a person possesses. Indian epics and spiritual masters have referred to *kritagayata* as a divine virtue, which needs to be cherished in every aspect of life. The sense of community with a mean value of 3.86 is not far behind *kritagayata*. This dimension of workplace spirituality implies respect and value for the feeling of interdependence and interconnectedness among employees. The

least rated construct of workplace spirituality is authenticity (mean = 2.87). As far as the overall constructive deviance in concerned, it has been measured with the help of ten statements using a 7-point rating scale. The mean value of 4.52 on this scale is well above the mid-value of 3.5 on the 7-point rating scale. Further, values of Cronbach's alpha, composite reliability, AVE, and VIF confirm reliability, validity, and lack of multicollinearity.

Now, the ten statements of constructive deviance are subjected to factor analysis for facilitating an in-depth analysis of the relationship between workplace spirituality and constructive deviance. Prior to factor analysis, the adequacy of the sample data for application of factor analysis was adjudged using the KMO and Bartlett tests.

Table 3 depicts the results of the KMO and Bartlett tests. The KMO test checks whether the given sample is appropriate for factor analysis or not. The KMO statistic depicts the proportion of common variance among the variables. Low common variance is a pre-requisite for judging the suitability of data for factor analysis. The value of KMO lies between 0 and 1, and a value above 0.8 indicates adequacy of data for factor analysis. Bartlett's test of sphericity is based on the null hypothesis that the correlation matrix is an identity matrix. A *P* value of less than 0.05 confirms the fitness of the sample for factor analysis. Now, the ten statements of constructive deviance are subjected to factor analysis using varimax rotation with the help of SPSS 16.0. Factor loading of a value greater than or equal to 0.6 is considered to segment various statements under extracted factors. One statement, that is, "Departed from dysfunctional organizational policies or procedures to solve a problem," has been discarded due to low factor loading.

Table 4 shows the result of the factor analysis. Following is a brief description of the three extracted factors:

F1: Procedure and Rule Deviant Behaviour: This factor includes the statements that explore the employees' perceptions regarding deviation from established organizational rules and procedures. The factor has an eigenvalue of 28.84, and it explains 33.85% of the variance. The deviations from rules and procedures discussed above are considered to be constructive as the deviation leads to a resolution of problem and also helps in the attainment of better results.

F2: Supervisor and Work Group Deviant Behaviour: The second factor appreciates the statements supporting deviation from the supervisor's instructions and work group norms. It has an eigenvalue of 16.91, and is responsible for 28.84% of the variance.

F3: Whistle-blowing Behaviour: This factor includes two statements that support reporting of wrong-doing in an organization. The reporting of wrong acts is

Table 3 KMO and Bartlett tests

Kaiser-Meyer-Olkin measure of sampling adequacy		0.87
Bartlett's test of sphericity	Approx. chi-square	389.940
	Df	40
	Sig.	0

Source: primary data

Table 4 Result of the factor analysis

Factors	Loading	Eigenvalue	Percentage of variance
F1: Procedure and Rule Deviant Behaviour		28.84	33.85
Bend or break rules in order to perform your job	0.757		
Violated company procedures to solve a problem	0.734		
Departed from organizational procedures to solve a job-related problem	0.814		
Bent a rule to maximize result from a job	0.817		
F2: Supervisor and Work Group Deviant Behaviour		16.91	28.84
Did not follow the orders of supervisor to improve work procedures	0.683		
Disobeyed supervisors' instructions to perform more efficiently	0.788		
Disagreed with work group to improve current work procedures	0.865		
F3: Whistle-blowing Behaviour		12.27	22.47
Reported wrong doing of co-worker to bring positive organizational change	0.853		
Reported a wrong doing to another person in your company to bring positive organizational change	0.837		

Source: primary data

constructive and it leads to adherence to organizational policies, ethical guidelines, and the law of the land. The factor has an eigenvalue of 12.27, and it explains 22.47% of the variance.

The three factors delineated above have been extracted through administration of factor analysis on ten statements of constructive deviance. After exploratory factor analysis, confirmatory factor analysis (CFA) is also applied to check model fit of the data. Essential statistical estimates including chi-square = 17.01, RMR = .075, TLI = 0.95, CFI = 0.98, and GFI = 0.97 report good model fit. Now, the relationship between the six dimensions of workplace spirituality and the three extracted factors of constructive deviance is explored with the help of the correlation matrix (Table 5).

Table 5 reports the statistically significant and positive correlation between factor 1 (Procedure and Rule Deviant Behaviour) and all dimensions of workplace spirituality. Factor 2 (Supervisor and Work Group Deviant Behaviour) and factor 3 (Whistle-blowing Behaviour) report positive correlation with five dimensions of workplace spirituality. Only *karma* capital is reported to have a non-significant correlation with factor 2 and factor 3.

Table 6 presents the results of multiple regressions which were applied considering the six constructs of workplace spirituality as independent variables and the three extracted factors of constructive deviance as three different dependent variables. Based on Tables 6 and 7, the model considers the six dimensions of workplace spirituality as independent variables and factor 1, i.e., Procedure and Rule Deviant Behaviour, as the dependent variable. The result shows that all regression coefficients are statistically significant and workplace spirituality explains 60% variance in constructive deviant behavior. Similarly, model 2 and model 3 report that the six constructs of workplace

Table 5 Correlation matrix

	SD	LS	AU	SC	KC	KT	F1	F2	F3
SD	1								
LS	0.574**	1							
AU	0.693**	0.584**	1						
SC	0.596**	0.473**	0.593**	1					
KC	0.820**	0.593**	0.522**	0.583**	1				
KT	0.743**	0.603**	0.449**	0.498**	0.802**	1			
F1	0.493*	0.384*	0.429*	0.603*	0.488*	0.284*	1		
F2	0.476*	0.312*	0.392*	0.394*	0.203	0.394*	0	1	
F3	0.645**	0.457*	0.683**	0.473*	0.304	0.638**	0	0	1

Source: primary data

**Correlation is significant at the 0.01 level; *correlation is significant at the 0.05 level

SD Swadharma, *LS Lokasangraha*, *AU* authenticity, *SC* sense of community, *KC karma* capital, *KT Kritagayata*, *F1* Procedure and Rule Deviant Behaviour, *F2* Supervisor and Work Group Deviant Behaviour, *F3* Whistle-blowing Behaviour

spirituality explain 68% and 65% variance in factor 2 (Supervisor and Work Group Deviant Behaviour) and factor 3 (Whistle-blowing Behaviour) respectively.

Discussion and conclusion

This research study examined the relationship between workplace spirituality and constructive deviance. More precisely, the study explored the correlation and regression between constructive deviance and six Indian dimensions of workplace spirituality. Prior to the generation of the correlation matrix, factor analysis was performed, and three factors were extracted from ten statements of constructive deviance. All

Table 6 Results of multiple regressions

Dependent variables →	Factor 1		Factor 2		Factor 3	
Independent variables ↓	Beta	Sig.	Beta	Sig.	Beta	Sig.
Constant	2.451		1.622		3.691	
Karma capital	0.455	0.00*	0.210	0.02*	0.234	0.02*
Authenticity	0.421	0.00*	0.384	0.00*	0.632	0.05*
Sense of community	0.573	0.01*	0.291	0.04*	0.445	0.04*
Lokasangraha	0.350	0.03*	0.233	0.00*	0.419	0.00*
Swadharma	0.444	0.00*	0.372	0.00*	0.623	0.00*
Kritagayata	0.274	0.05*	0.349	0.03*	0.615	0.01*

Source: primary data. **Significant at the 0.01 level, *significant at the 0.05 level

Factor 1 Procedure and Rule Deviant Behaviour, *Factor 2* Supervisor and Work Group Deviant Behaviour, *Factor 3* Whistle-blowing Behaviour

Table 7 Model summary

Model	Dependent variable	R	R^2	Adjusted R^2	F-value	Sig.
Model 1	Factor 1	0.79	0.62	0.60	12.35	0.00*
Model 2	Factor 2	0.85	0.72	0.68	6.42	0.04*
Model 2	Factor 3	0.84	0.70	0.65	4.35	0.03*

Source: primary data. **Significant at the 0.01 level, *significant at the 0.05 level

Factor 1 Procedure and Rule Deviant Behaviour, *Factor 2* Supervisor and Work Group Deviant Behaviour, *Factor 3* Whistle-blowing Behaviour

constructs of workplace spirituality were found to have a positive correlation with all the three extracted factors of constructive deviance. In simple words, positive variations in these dimensions of workplace spirituality may result in corresponding favorable variations in procedure and rule-related constructive deviance. The findings may be understood with the help of an appropriate amalgamation of modern management theories and the ancient Indian ethos. Spiritual texts define *Swadharma* as the pursuit of self for righteous action. And the righteous pursuit for modern organizations signifies wealth maximization for shareholders, which is directly related to the performance of employees. Thus, any constructively deviant behavior, which enables better performance, is supported by *Swadharma*. In the *Mahabharata*, Lord Krishna deviated from the established rules of the war to end the unethical, immoral, and exploitative rule of the king. Lord Jesus prayed to God for mercy on those who crucified him. This behavior of the "Son of God" is righteous but again it deviates from the modern practiced rule of "tit for tat." Modern literature on constructive deviance also seems to support *Swadharma.* Pro-social rule breaking behavior is defined as the "intentional violation of a formal organizational policy, regulation, or prohibition with the primary intention of promoting the welfare of the organization or one of its stakeholders" (Morrison 2006, p. 6). This definition of pro-social rule clearly expresses deviation from established organizational policies for the betterment of stakeholders or an organization. "Such behavior reflects a desire to do things better or to "do good" in the context of one's organizational role" (Morrison 2006, p. 8). Crant (2000) described proactive behavior as "taking initiative in improving current circumstances or creating new ones; it involves challenging the status quo rather than passively adapting to present conditions" (p. 436). Again, the intention behind proactive behavior is in the direction of influencing the organizational environment for ensuring the wellness of the organization (Grant 2008; Vadera et al. 2013).

The findings of the study also highlight a significant and positive correlation between *karma* capital and constructive deviance. Lord Krishna has devoted an entire chapter in the *Shrimad Bhagavad Gita* to explain the significance of *karma* in personal and professional life. *Karma capital* basically entails three learnings: (i) excellence in work (*YogahKaramsu Kosalam*, *Bhagavad Gita*, Chapter 2, verse 50), (ii) non-attachment to results (*karmanyevadhikaraste ma phalesukadacana ma karma-phala-heturbhur ma te' sangostvakarmani*, *BhagavadGita*, Chapter 2, verse 47), and (iii) 100% work efficiency and devotion. Growing acceptability of extra-role behavior and voice (forms of constructive deviance) may be explained with the help of the employees' latent pursuit for excellence and utmost devotion in the workplace. Extra-role

behavior is behavior which benefits and/or is intended to benefit the organization, which is discretionary and which goes beyond existing role expectations (Van Dyne et al. 1995). Constructively deviant organization citizenship behavior, which includes "assisting colleagues and supervisors, not indulging in excessive breaks, and paying attention to functions that are not required" (Organ 1988), may also be seen in the light of work excellence. Vadera et al. (2013) highlighted three reasons for indulgence in constructive deviant behaviors: intrinsic motivation, felt obligation, and psychological empowerment. *Karma* theory represents the way ahead for defined drivers (interest and enjoyment) of intrinsic motivation, as it equates work to worship. Indian spiritual texts have claimed that the notion of *karma* may even lead to salvation (Chakrabarti 1999).

Kritagayata (gratitude) is one of the special and unique features of Indian spiritual teachings (Sarkar and Garg 2020). Here, the findings indicate that the *kritagayata* component of workplace spirituality is also positively associated with constructive deviance, which is in line with previous research results. *Kritagayata* has a positive relation with the motivation to act pro-socially (Naito et al. 2005; Tsang 2007; Froh et al. 2008). Various constituents of organization citizenship behavior were found to be related to gratitude (Emmons and McCullough 2003; Wood et al. 2009). Felt obligation may increase the propensity of constructive deviance (Morrison and Phelps 1999; Vadera et al. 2013). A grateful employee feels morally obliged to return. The social exchange theory (Blau 1964) seems to support this stated fact. According to the social exchange theory, employees develop a strong reciprocal obligation in return for pro-employee initiatives adopted by the employer. When the employer makes some contribution to the welfare of the employees, the latter feel grateful towards the former and develop a strong desire to reciprocate these measures.

Also, authenticity commands genuineness and openness, which are nourished with meditation and self-reflection (Gardner et al. 2005). Pandey et al. (2016) defined authenticity as "the alignment of people's actions and behaviors with their core, internalized values, and beliefs." Thus, it may be inferred from the results that the employees who have internalized the values of honesty and integrity will resort to whistle-blowing behavior. "Pressure decentralization, flexibility, and stronger performance orientation, employees are expected to be more creative, more productive, and more involved in the performance and survival of the organization" (Vadera et al. 2013, p. 1266). Employees may resort to deviant means to remain useful and suitable in this age of throat-cut competition. Researchers have argued that deviance is not always detrimental; rather, positive deviance will contribute significantly to organizational growth (Galperin 2003; Spreitzer and Sonenshein 2004). This paper provides empirical support for channelizing spiritual energy to promote constructive deviant behavior in the workplace. This study is one of the pioneer research works that empirically investigate the relationship between workplace spirituality and constructive deviance. The study proudly introduces Indian conceptualization of the model of workplace spirituality and constructive deviance. Various dimensions of workplace spirituality have been taken from diverse concepts of the research studies on workplace spirituality, keeping in view the great Indian epics and scriptures. To promote innovation, pro-social behavior, and a stress-free, happy, and balanced work-life, modern organizations have started implementing various spiritual interventions in the workplace. Leading companies like the World Bank, Hewlett-Packard, AT&T, Ford Motors, Apple

Computers, Google, Microsoft, Yahoo, and IBM all have realized the importance of spirituality in the workplace.

This research study provides at least three precious contributions in the field of organization behavior and human resource management literature. Firstly, it offers theoretical and empirical support of the association between spirituality in the workplace and constructive deviance. Several previous researchers, including Sheep (2006), Gatling et al. (2016), and Garg (2015), promoted further research in the field of spirituality. And thus, the present study enriches existing literature of workplace spirituality. Secondly, it also provides statistical evidence of workplace spirituality leading to constructive deviance. This is one of the pioneer studies that investigate the relationship between these two variables. Thirdly, the study reports major factors of constructive deviance that are extracted with the help of factor analysis.

The findings of the study also have significant practical implications for human resource managers and practitioners in the workplace. Since constructive deviance is regarded as the "new normal" in modern organizations, human resource managers could utilize the services of workplace spirituality in ensuring constructive deviant behavior, especially rule- and procedure-deviant behavior, and whistle-blower behavior. An organization should institutionalize a holistic and all-encompassing spiritual climate by incorporating spiritual values in the overall organizational vision, mission, and policies to enable the realization of the true potential of a positive association between workplace spirituality and constructive deviance. Thus, modern organizations are suggested to facilitate workplace spirituality through institutionalizing several interventions like meditation, yoga, vipassana, introspection, and mindfulness. Dent et al. (2005) advocated a holistic organization-wide spirituality development program. Several companies like Apple, Google, Amazon, Facebook, and ONGC have already started to implement these practices at their workplace.

Although the study has new findings to offer, it also has a few limitations. Firstly, the present research work is based on a sample size of only 152. Larger sample sizes would have helped in the better generalization of results. Secondly, data related to all variables of the study (independent and dependent variables) were collected at the same time from the same source. This could lead to a common method error. And thirdly, Table 5 indicates a high positive correlation between the various constructs of workplace spirituality, which is acceptable given the nature of the variables. However, Harman's single-factor test indicates that the common method error is not significant. Future researchers may opt for a longitudinal research design to obtain more valid findings and results. Other future studies may investigate the moderating effect of various demographic variables like gender, age, experience, and academic qualification in the relationship between workplace spirituality and constructive deviance. The questionnaire used for constructive deviance is a general one, but further experiments in the field could use specialized questionnaires for different constructive deviant behaviors like pro-social behavior, whistle-blower behavior, and issue selling. Vadera et al. (2013) reported three antecedent variables, that is, felt obligation, psychological obligation, and intrinsic motivation, to constructive deviance. New-age scholars could evaluate the mediating role of these three variables amid the positive association between workplace spirituality and constructive deviance.

Compliance with ethical standards

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

References

Abdullah, A. G. K., Alzaidiyeen, N. J., & Aldarabah, I. T. (2009). Workplace spirituality and leadership effectiveness among educational managers in Malaysia. *European Journal of Social Sciences, 10*(2), 304–316.

Afsar, B., & Rehman, M. (2015). 'The relationship between workplace spirituality and innovative work behavior: the mediating role of perceived person–organization fit', Journal of Management. *Spirituality & Religion, 12*(4), 329–353.

Ahmad, A., & Omar, Z. (2014). Reducing deviant behaviour through workplace spirituality and job satisfaction. *Asian Social Science, 10*(19), 107–112.

Altaf, A., & Awan, M. A. (2011). Moderating effect of workplace spirituality on the relationship of job overloads and job satisfaction. *Journal of Business Ethics, 104*(1), 93–99.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *The Academy of Management Journal, 39*(5), 1154–1184.

Armour, S. (1998). Employers hope spirituality lifts workers' morale. *USA Today.*

Ashmos, D. P., & Duchon, D. (2000). Spirituality at work: a conceptualization and measure. *Journal of Management Inquiry, 9*(2), 134–145.

Bennett, R. J., & Robinson, S. L. (2003). The past, present, and future of workplace deviance research. In J. Greenberg (Ed.), *Organizational behavior: the state of the science* (2nd ed.). Mahwah, NJ: Erlbaum.

Blau, P. M. (1964). *Exchange and power in social life*. New York: John Wiley.

Bodankin, M., & Tziner, A. (2009). Constructive deviance, destructive deviance and personality: how do they interrelate? *Amfiteatru Economic Journal, 11*, 549–564.

Bord, R. J. (1976). The impact of imputed deviant identities in structuring evaluations and reactions. *Sociometry, 39*, 108–115.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco: Jossey-Bass.

Case, P., & Gosling, J. (2010). The spiritual organization: critical reflections on the instrumentality of workplace spirituality. *Journal of Management, Spirituality and Religion, 7*(4), 257–282.

Chakrabarti, K. K. (1999). *Classical Indian philosophy of mind: The Nyaya dualist tradition*. New Delhi: Motilal Banarsidass Publishers.

Chakraborty, S. K. (1987). *Managerial effectiveness and quality of work life*. New Delhi: Tata McGraw-Hill.

Chakraborty, S. K. (1995). *Ethics in management: Vedantic perspective*. New Delhi: Oxford University Press.

Chang, S. J., Van Witteloostuijn, A., & Eden, L. (2010). From the editors: Common method variance in international business research. *Journal of International Business Studies, 41*(2), 178–184.

Chattopadhyay, S., & Agrawal, K. G. (1977). The dying Ulysses: middle aged middle managers in the Indian public sector. *Vikalpa, 2*(3), 135–150.

Chawla, V., & Guda, S. (2010). Individual spirituality at work and its relationship with job satisfaction, propensity to leave and job commitment: an exploratory study among sales professionals. *Journal of Human Values, 16*, 157–167.

Comer, P. D. (2008). Workplace spirituality and business ethics: insights from an Eastern spiritual tradition. *Journal of Business Ethics, 85*, 377–389.

Crant, J. M. (2000). Proactive behavior in organizations. *Journal of Management, 26*, 435–462.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391–418.

Cropanzano, R., Bowen, D. E., & Wgilliland, S. (2007). The management of organizational justice. *Academy of Management Perspectives, 21*(4), 34–48.

Dehler, G. E., & Welsh, M. A. (2003). The experience of work: spirituality and the new work-place. In R. A. Giacalone & C. L. Jurkiewicz (Eds.), *Handbook of workplace spirituality and organizational performance* (pp. 108–122). M.E. Sharpe: Armonk, NY.

Dent, E. B., Higgins, M. E., & Wharff, D. M. (2005). Spirituality and leadership: an empirical review of definitions, distinctions, and embedded assumptions. *The Leadership Quarterly, 16*(5), 625–653.

Donaldson, T., & Dunfee, T. (1994). Toward a unified conception of business ethics: integrative social contracts theory. *Academy of Management Review, 19*(2), 252–284.

Duchon, D., & Plowman, D. A. (2005). Nurturing the spirit at work: impact on unit performance. *The Leadership Quarterly, 16*(5), 807–834.

Dutt, P. (2018). The concept of Swadharma and effect on leadership. *SSRN Electronic Journal*.

Dutton, J. E., & Ashford, S. J. (1993). Selling issues to top management. *Academy of Management Review, 18*, 397–428.

Ehtesham. (2011). Relationship between organizational culture and performance management practices: A case of University in Pakistan. *Journal of Competitiveness*, 27–30.

Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology, 84*, 377–389.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Froh, J. J., Sefick, W. J., & Emmons, R. A. (2008). 'Counting blessings in early adolescents: an experimental study of gratitude and subjective well-being'. *Journal of School Psychology, 46*, 213–233.

Galen, M., & West, K. (1995). *Companies hit the road less travelled*. Business Week.

Galperin, B. L. (2003). Can workplace deviance be constructive? In A. Sagie, S. Stashevsky, & M. Koslowsky (Eds.), *Misbehaviour and dysfunctional attitudes in organizations* (pp. 154–170). New York: Palgrave Macmillan.

Galperin, B. L. (2012). Exploring the nomological network of workplace deviance: developing and validating a measure of constructive deviance. *Journal of Applied Social Psychology, 42*(12), 2988–3025.

Gardner, W. L., B. J Avail, F. May, Luthans, D. R., and F. Walumbwa: 2005, 'Can you see the real me?' A self-based model of authentic leader and follower development, The Leadership Quarterly 16(3), 343–372.

Garg, N. (2015). Employee engagement and individual differences: A study in Indian context. *IUP Journal of Organizational Behavior, 14*(1), 58.

Garg, N. (2017). Workplace spirituality and employee well-being: an empirical exploration. *Journal of Human Values, 23*(2), 1–19.

Garg, N. (2018). Promoting organizational performance in indian insurance industry: The roles of workplace spirituality and organizational citizenship behaviour. *Global Business Review, 21*(2), 1–16.

Garg, N. & Gera, S. (2019). Gratitude and leadership in higher education institutions: Exploring the mediating role of social intelligence among teachers. *Journal of Applied Research in Higher Education, Vol. ahead-of-print No. ahead-of-print*.

Garg, N., & Saxena, A. (2019). *Promoting constructive deviance as an antidote to organizational stress, analyzing workplace deviance in modern organizations, 139-158*. Ahmedabad, Gujrat: IGI Global.

Garg, N., Punia, B. K., & Jain, A. (2019). Workplace spirituality and job satisfaction: Exploring mediating effect of organization citizenship behaviour. *Vision- The Journal of Business Perspective, 23*(3), 287–296.

Gatling, A. R., Kang, H. A., & Kim, J. (2016). The effects of authentic leadership and organizational commitment on turnover intention. *Leadership & Organization Development Journal, 37*(2), 181–199.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: a simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Giacalone, R. A., & Jurkiewicz, C. L. (2004). Toward a science of workplace spirituality. In R. A. Giacalone & C. L. Jurkeiwicz (Eds.), *Handbook of workplace spirituality and organizational performance*. New York, NY: M.E. Sharpe.

Gabriela. (2013). Organizational injustice: Third parties reactions to mistreatment of employee. *Psicothema*, 215–222.

Grant, A. M. (2008). Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Psychology, 93*, 48–58.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (seventh ed.). Upper Saddle River, New Jersey: Prentice Hall.

Hart, D. W., & Brady, F. N. (2005). Spirituality and archetype in organizational life. *Business Ethics Quarterly, 15*(3), 409–428.

Harvey, P. (2000). *An introduction to Buddhist ethics*. Cambridge: Cambridge University Press.

Judge, W. Q. (1999). *The leader's shadow: exploring and developing executive character*. Thousand Oaks, CA: Jossey-Bass.

Jurkiewicz, C. L., & Giacalone, R. A. (2004). A values framework for measuring the impact of workplace spirituality on organizational performance. *Journal of Business Ethics, 49*(2), 129–142.

Karakas, F. (2010). Spirituality and performance in organizations: a literature review. *Journal of Business Ethics, 94*(1), 89–106.

Kelloway. (2010). Counter productive behavior as protest. *Human Resource Management Review*, 20–23.
Khari, C., & Sinha, S. (2017). Impact of workplace spirituality on knowledge sharing intention: A conceptual framework. *Journal of Human Values, 23*(1), 27–39.
Kinjerski, V., & Skrypnek, B. (2004). Defining spirit at work: finding common ground. *Journal of Organizational Change Management, 17*(1), 26–42.
Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
Kolodinsky, R. W., Giacalone, R. A., & Jurkiewicz, C. L. (2008). Workplace values and outcomes: exploring personal, organizational, and interactive workplace spirituality. *Journal of Business Ethics, 81*(2), 465–480.
Krishnakumar, S., & Neck, C. P. (2002). The "what", "why" and "how" of spirituality in the workplace. *Journal of Managerial Psychology, 17*(3), 153–164.
Lian, H., Ferris, D. L., & Brown, D. J. (2012). Does power distance exacerbate or mitigate the effects of abusive supervision? It depends on the outcome. *Journal of Applied Psychology, 97*(1), 107–123.
Madhav, P., & Kumar. (2016). Workplace spirituality and employee work attitudes: exploring gender as a moderator. *IOSR Journal of Business and Management, 18*(5), 1–10.
Malhotra, N. K., & Dash, S. (2011). *Marketing Research an Applied Orientation*. London: Pearson Publishing.
McMillan, D. W., & Chavis, D. M. (1986). Sense of community: a definition and theory. *Journal of Community Psychology, 14*, 6–23.
Milliman, J., Czaplewski, A. J., & Ferguson, J. (2003). Workplace spirituality and employee work attitudes: an exploratory empirical assessment. *Journal of Organizational Change Management, 16*(4), 426–447.
Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the moderating effects of negative reciprocity beliefs. *Journal of Applied Psychology, 92*(4), 1159–1168.
Morrison, E. W. (2006). Doing the job well: an investigation of pro-social rule breaking. *Journal of Management, 32*, 5–28.
Morrison, E. W., & Phelps, C. C. (1999). Taking charge at work: extra role efforts to initiate workplace change. *Academy of Management Journal, 42*(4), 403–419.
Mulla, Z. R., & Krishnan, V. R. (2007). Karma-Yoga: construct validation using value systems and emotional intelligence. *South Asian Journal of Management, 14*(4), 116–113.
Naito, T., Wangwanand, J., & Tani, M. (2005). Gratitude in university students in Japan and Thailand. *Journal of Cross-Cultural Psychology, 36*, 247–263.
Naylor, T., Willimon, W., & Osterberg, R. (1996). The search for community in the workplace. *Business and Society Review, 97*(1), 42–47.
Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal, 39*, 607–634.
Organ, D. W. (1988). *Organizational citizenship behavior: the good soldier syndrome*. Lexington, MA: Lexington Books.
Pandey, A., Gupta, R. K., & Arora, A. P. (2009). Spiritual climate of business organizations audits impact on customers' experience. *Journal of Business Ethics, 88*(2), 313–332.
Pandey, A., Gupta, R. K., & Kumar, P. (2016). Spiritual climate and its impact on learning in teams in business organizations. *Global Business Review, 17*(3S), 1S–14S.
Paula, B., & Sofia, G. (2003). Poverty, equity, human rights and health. *Bulletin of the World Health Organization, 81*, 539–545.
Pawar, B. S. (2008). Two approaches to workplace spirituality facilitation: a comparison and implications. *Organization Development Journal, 29*(6), 544–567.
Pawar, B. (2009). Workplace spirituality facilitation: a comprehensive model. *Journal of Business Ethics, 90*(3), 375–386.
Penney, L. M., & Spector, P. E. (2005). Job stress, incivility and counterproductive work behaviour: the moderating role of negative affectivity. *Journal of Organizational Behaviour, 26*, 777–796.
Petchsawang, P., & Duchon, D. (2009). Measuring workplace spirituality in an Asian context. *Human Resource Development International, 12*(4), 459–468.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.
Pradhan, R. K., & Jena, L. K. (2016). Workplace spirituality and organisational commitment: role of emotional intelligence among Indian banking professional. *Journal of Human Resource Management, 19*(1), 13–23.
Prasanna, & Madhavaiah. (2017). Effects of workplace spirituality towards counterproductive behavior in automobile manufacturing sector with reference to Tamilnadudistrict. *International Journal of Academic Research and Development, 2*(6), 375–379.

Radhakrishnan, S. (1926). *The Hindu view of life*. India: Harper Collins Publishers.
Radhakrishnan, S. (1970). *The Bhagavad Gita*. London: George Allen and Unwin.
Robbins, D. L., & Galperin, B. L. (2010). Constructive deviance: striving toward organizational change in healthcare. *Journal of Management and Marketing Research, 5*, 1–11.
Robinson, S. L., & Bennett, R. B. (1995). A typology of deviant workplace behaviors: a multi-dimensional scaling study. *Academy of Management Journal, 38*, 555–572.
Sanders III, J. E., Hopkins, W. E., & Geroy, G. D. (2003). From transactional to transcendental: toward and integrated theory of leadership. *Journal of Leadership and Organizational Studies in Ethnicity and Nationalism, 9*, 21–31.
Sengupta, S. S. (2010). Correlates of spiritual orientation and managerial effectiveness. *Indian Journal of Industrial Relations, 46*(1), 45–60.
Spreitzer, G.M., and S. Sonenshein: 2003, Positive deviance and extraordinary organizing. In K. Cameron, J. Dutton, & R. Quinn (Eds.), Positive organizational scholarship: foundations of a newdiscipline (pp. 207-224). San Francisco: Berrett-Kohler.
Sarkar, A. & Garg, N. (2020), "Peaceful workplace" only a myth? Examining the mediating role of psychological capital on spirituality and nonviolence behaviour at the workplace, *International Journal of Conflict Management, Vol. ahead-of-print No. ahead-of-print*. https://doi.org/10.1108/IJCMA-11-2019-0217
Sheep, M. L. (2006). Nurturing the whole person: The ethics of workplace spirituality in a society of organizations. *Journal of Business Ethics, 66*(4), 357–375.
Spreitzer, G. M., & Sonenshein, S. (2004). Toward the construct definition of positive deviance. *American Behavioral Scientist, 47*(6), 828–847.
Tepper, B. J. (2003). Organizational citizenship behavior and the spiritual employee. In R. A. Giacalone & C. L. Jurkiewicz (Eds.), *Handbook of workplace spirituality and organizational performance* (pp. 181–190). M.E. Sharpe: Armonk.
Tsang, J. (2007). Gratitude for small and large favors: a behavioral test. *Journal of Positive Psychology, 3*, 157–167.
Vadera, A. K., Pratt, M. G., & Mishra, P. (2013). Constructive deviance in organizations integrating and moving forward. *Journal of Management, 39*(5), 1221–1276.
Van Dyne, L., Cummings, L. L., & McLean Parks, J. (1995). Extra-role behaviors: in pursuit of construct and definitional clarity. In L. L. Cumings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 17, pp. 215–285).
Warren, D. (2003). Constructive and destructive deviance tn organizations. *Academy of Management Review, 28*(4), 622–632.
Wood, A. M., Joseph, S., & Maltby, J. (2009). Gratitude predicts psychological wellbeing above the Big Five facets. *Personality and Individual Differences, 46*, 443–447.
Yildiz, B., Alpkan, L., Ates, H., & Sezen, B. (2015). Determinants of constructive deviance: the mediator role of psychological ownership. *International Business Research, 8*(4), 107–121.
Zhang, X., & Bartol, K. (2010). Linking empowering leadership and employee creativity: the influence of psychological empowerment, intrinsic motivation and creative process engagement. *The Academy of Management Journal, 53*(1), 107–128.
Zohar, D., & Marshall, I. (2000). *SQ – spiritual intelligence: the ultimate intelligence*. London: Bloomsbury.

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

J. Korean Math. Soc. **57** (2020), No. 2, pp. 331–357
https://doi.org/10.4134/JKMS.j190051
pISSN: 0304-9914 / eISSN: 2234-3008

APPLICATIONS OF DIFFERENTIAL SUBORDINATIONS TO CERTAIN CLASSES OF STARLIKE FUNCTIONS

Shagun Banga and S. Sivaprasad Kumar

Abstract. Let p be an analytic function defined on the open unit disk $\mathbb{D}$. We obtain certain differential subordination implications such as $\psi(p) := p^{\lambda}(z)(\alpha+\beta p(z)+\gamma/p(z)+\delta z p'(z)/p^{j}(z)) \prec h(z)$ $(j=1,2)$ implies $p \prec q$, where h is given by $\psi(q)$ and q belongs to $\mathcal{P}$, by finding the conditions on α, β, γ, δ and λ. Further as an application of our derived results, we obtain sufficient conditions for normalized analytic function f to belong to various subclasses of starlike functions, or to satisfy $|\log(zf'(z)/f(z))| < 1$, $|(zf'(z)/f(z))^2 - 1| < 1$ and $zf'(z)/f(z)$ lying in the parabolic region $v^2 < 2u - 1$.

1. Introduction

The set of analytic functions f defined on the unit disk $\mathbb{D} = \{z : |z| < 1\}$ of the form $f(z) = z + a_2z^2 + a_3z^3 + \cdots$ is denoted by $\mathcal{A}$. Let $\mathcal{S}$ denote the subclass of $\mathcal{A}$ consisting of univalent functions. Let $\mathcal{A}_0$ be the class of all analytic functions $p(z)$, which does not vanish anywhere in $\mathbb{D}$, with the normalization $p(0) = 1$. Let $\mathcal{P}$ denote the subclass of $\mathcal{A}_0$ consisting of the functions p with the positive real part. This class is known as Carathéodory class of functions. For two analytic functions f and g, we have f subordinate to g, written as $f \prec g$, if there is a Schwarz function w such that $w(0) = 0$, $|w(z)| < 1$ and $f(z) = g(w(z))$. If g is univalent, then $f \prec g$ if and only if $f(0) = g(0)$ and $f(\mathbb{D}) \subseteq g(\mathbb{D})$. Ma and Minda [8] introduced the classes:

$$\mathcal{S}^*(\phi) = \left\{ f \in \mathcal{S} : \frac{zf'(z)}{f(z)} \prec \phi(z) \right\}$$

and

$$\mathcal{C}(\phi) = \left\{ f \in \mathcal{S} : 1 + \frac{zf''(z)}{f'(z)} \prec \phi(z) \right\},$$

Received January 17, 2019; Accepted August 26, 2019.
2010 *Mathematics Subject Classification.* 30C45, 30C80.
Key words and phrases. Carathéodory function, differential subordinations, minimum principle, exponential function, strongly starlike function, lemniscate of Bernoulli, Janowski starlike function.
The work presented here was supported by a Research Fellowship from the Department of Science and Technology, New Delhi.

©2020 Korean Mathematical Society

where ϕ is an analytic univalent function with positive real part in $\mathbb{D}$ such that $\phi(\mathbb{D})$ is symmetric with respect to the real axis and starlike with respect to $\phi(0) = 1$ with $\phi'(0) > 0$. For different choices of ϕ, $\mathcal{S}^*(\phi)$ reduces to well known classes. For example, when $\phi(z) := (1+Az)/(1+Bz)$, $\mathcal{S}^*((1+Az)/(1+Bz)) =: \mathcal{S}^*[A, B]$, is the class of Janowski starlike functions [5], where $-1 \leq B < A \leq 1$. For $A = 1$ and $B = -1$, this class reduces to the class of normalized starlike functions, $\mathcal{S}^*((1 + z)/(1 - z)) =: \mathcal{S}^*[-1, 1]$ and for $A = 1 - 2\nu$ $(0 \leq \nu < 1)$ and $B = -1$, Robertson [17] introduced the class of starlike functions of order ν, $\mathcal{S}^*((1 + (1 - 2\nu)z)/(1 - z)) =: \mathcal{S}^*(\nu)$. Note that $\mathcal{S}^*(0) = \mathcal{S}^*$, is the class of starlike functions. Another interesting class is the class of starlike functions of reciprocal order ν, in $\mathbb{D}$, which is given by

$$\operatorname{Re}\left(\frac{f(z)}{zf'(z)}\right) > \nu \quad (z \in \mathbb{D}).$$

For $\phi(z) := \sqrt{1 + z}$, Sokół and Stankiewicz [20] introduced the class of analytic functions associated with lemniscate of Bernoulli, $\mathcal{S}^*_L := \mathcal{S}^*(\sqrt{1 + z})$. Functions satisfying $|\log(zf'(z)/f(z))| < 1$, belongs to the class $\mathcal{S}^*_e :=\mathcal{S}^*(e^z)$, introduced by Mendiratta *et al.* [10]. The class of strongly starlike functions of order η is introduced in [1] and [22]. We denote it by

$$\mathcal{SS}^*(\eta) := \{f \in \mathcal{A} : |\arg(zf'(z)/f(z))| < \eta\pi/2\}, \quad (0 < \eta \leq 1).$$

It is obtained when $\phi(z) := ((1 + z)/(1 - z))^\eta$. The speciality of this class lies in the fact that it helps to study the function in terms of argument estimation. For $\eta = 1$, it reduces to the class of normalized starlike functions $\mathcal{S}^*$. For $\phi(z) := ((1 + cz)/(1 - z))^{(\eta_1+\eta_2)/2}$, Liu and Srivastava [7] introduced the class

$$\begin{aligned}\mathcal{SS}^*(\eta_1, \eta_2) = \{f \in \mathcal{A} : -\eta_2\pi/2 < |\arg(zf'(z)/f(z))| < \eta_1\pi/2\} \\ (0 < \eta_1, \eta_2 \leq 1),\end{aligned} \tag{1}$$

where

$$\eta = \frac{\eta_1 - \eta_2}{\eta_1 + \eta_2}, \qquad c = e^{\eta\pi i}. \tag{2}$$

As per the definition of the class $\mathcal{SS}^*(\eta_1, \eta_2)$ in (1), we give the name *oblique sector* function to the above defined function $((1 + cz)/(1 - z))^{(\eta_1+\eta_2)/2}$. Note that $\mathcal{SS}^*(\eta, \eta)$ reduces to the class $\mathcal{SS}^*(\eta)$. The class $S_P :=\mathcal{S}^*(\phi_{PAR}(z))$, introduced by Rønning [18], is the class of parabolic starlike functions, where

$$\phi_{PAR}(z) := 1 + \frac{2}{\pi^2}\left(\log\frac{1 + \sqrt{z}}{1 - \sqrt{z}}\right)^2 \quad \operatorname{Im}\sqrt{z} \geq 0,$$

consists of functions $f \in \mathcal{A}$, such that $\operatorname{Re}(zf'(z)/f(z)) > |zf'(z)/f(z) - 1|$. Let $\mathcal{N}(\kappa)$ be the subclass of $\mathcal{A}$ consisting of the functions $f(z)$ which satisfy the following

$$\operatorname{Re}\left(1 + \frac{zf''(z)}{f'(z)}\right) < \kappa \quad (\kappa > 1; z \in \mathbb{D}).$$

Uralegaddi *et al.* [24] investigated the class $\mathcal{N}(\kappa)$ for $1 < \kappa < 4/3$.

Let ψ be analytic in a domain $D \subset \mathbb{C}^2$ and h be univalent in D. Then an analytic function p is the solution of the first order differential subordination if

$$\psi(p(z), zp'(z)) \prec h(z), \quad (p(z), zp'(z)) \in D \quad \text{for } z \in \mathbb{D}. \tag{3}$$

Then the dominant of the differential subordination (3) is the univalent function q, if $p \prec q$ for all the solutions p of (3). The best dominant of all the dominants of differential subordination (3) is the function $\tilde{q}$ whenever $\tilde{q} \prec q$. Miller and Mocanu [11] discussed the general theory of the first-order differential subordinations. Motivated by this result, many authors [19], [21] established several generalizations of first order differential subordination. Recently, the authors [16] obtained some sufficient conditions for analytic functions in $\mathbb{D}$ to satisfy the subordination $p(z) \prec q(z)$ and in specific to have positive real part. Many authors [2], [3], [4], [7] and [13] have evolved this concept of finding conditions on the parameters involved in the first order differential subordination as given in (3) in order to prove $p(z) \prec ((1+z)/(1-z))^{\eta}$, $(0 < \eta \leq 1)$ by satisfying the condition $|\arg(\psi(p(z), zp'(z))| \leq \arg(h(z))$, after estimating the argument of $h(z)$.

In this paper, we obtain certain differential subordination implications by finding conditions on the parameters involved in it. Mainly, our results involve two admissible classes of analytic functions:

$$p^{\lambda}(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \prec h(z)$$

and

$$p^{\lambda}(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p^2(z)}\right) \prec h(z),$$

which implies $p(z) \prec q(z)$, where h is univalent and $q(z) \in \mathcal{P}$. Also, we study these classes in terms of argument estimation of the class for the oblique sector function q.

In Section 3, we also obtain sufficient conditions for functions belonging in $\mathcal{S}^*(\nu)$ $(\nu = 0, 1/2)$, $\mathcal{S}_e^*$, $\mathcal{S}_L^*$, $\mathcal{S}^*[A, B]$, $(-1 < B < A \leq 1)$ and S_P as an application of our derived results in Section 2. In some cases, our results become the general case of the results obtained by Ravichandran and Kumar [16], Nunokawa *et al.* [14], Obradowič and Tuneski [15] and Cho *et al.* [4]. We also extend the result obtained by Mocanu [12].

Miller and Mocanu [11] gave the following result which is required to prove our main results.

Lemma 1.1. *Let $q(z)$ be univalent in the unit disk $\mathbb{D}$ and θ and ϕ be analytic in a domain D containing $q(\mathbb{D})$ with $\phi(\omega) \neq 0$ when $\omega \in q(\mathbb{D})$. Set $Q(z) = zq'(z)\phi(q(z))$, $h(z) = \theta(q(z)) + Q(z)$. Suppose that* (i) *$Q(z)$ is starlike univalent in $\mathbb{D}$, and* (ii) $\mathrm{Re}\ (zh'(z)/Q(z)) > 0$ *for $z \in \mathbb{D}$. If $p(z)$ is analytic in $\mathbb{D}$ with $p(0) = q(0)$, $p(\mathbb{D}) \subset D$ and satisfies*

$$\theta(p(z)) + zp'(z)\phi(p(z)) \prec \theta(q(z)) + zq'(z)\phi(q(z)), \tag{4}$$

then $p(z) \prec q(z)$ and $q(z)$ is the best dominant.

2. Results involving differential subordination implications

We begin with the following main results:

Theorem 2.1. *Let λ be a real number and α, β, γ and $\delta(\neq 0)$ be complex numbers. Suppose $q(z) \in \mathcal{A}_0$ be univalent in $\mathbb{D}$ and satisfy the following conditions for $z \in \mathbb{D}$:*

(1) *$Q(z) := \delta z q'(z) q^{\lambda-1}(z)$ be starlike (univalent).*

(2) $$\operatorname{Re}\left(\frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta} q(z) + \frac{\gamma(\lambda-1)}{\delta q(z)} + (\lambda-1)\frac{zq'(z)}{q(z)} + \left(1 + \frac{zq''(z)}{q'(z)}\right)\right) > 0 =: \operatorname{Re}(H(z)).$$

If $p(z) \in \mathcal{A}_0$ satisfies

$$p^{\lambda}(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \prec q^{\lambda}(z)\left(\alpha + \beta q(z) + \frac{\gamma}{q(z)} + \delta\frac{zq'(z)}{q(z)}\right),$$

then $p(z) \prec q(z)$ and $q(z)$ is the best dominant.

Proof. Let $\theta(\omega) = \omega^{\lambda}(\alpha + \beta\omega + \gamma/\omega)$, $\omega \neq 0$ and $\phi(\omega) = \delta\omega^{\lambda-1}$. Thus, $\phi(\omega) \neq 0$ and $\theta(\omega)$, $\phi(\omega)$ are analytic in $\mathbb{C} - \{0\}$. Let the function $Q(z)$ and $h(z)$ be given by

$$Q(z) := zq'(z)\phi(q(z)) = \delta z q'(z) q^{\lambda-1}(z)$$

and

$$h(z) := \theta(q(z)) + Q(z) = q^{\lambda}(z)\left(\alpha + \beta q(z) + \frac{\gamma}{q(z)} + \delta\frac{zq'(z)}{q(z)}\right).$$

Thus, we have $Q(z)$ is starlike and $\operatorname{Re}(zh'(z)/Q(z))$ reduces to

$$\operatorname{Re}\left(\frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta} q(z) + \frac{\gamma(\lambda-1)}{\delta q(z)} + (\lambda-1)\frac{zq'(z)}{q(z)} + \left(1 + \frac{zq''(z)}{q'(z)}\right)\right) > 0. \tag{5}$$

On substituting $q(z)$ as $p(z)$ in $\theta(q(z)) + Q(z)$, we get

$$\theta(p(z)) + zp'(z)\phi(p(z)) \prec \theta(q(z)) + zq'(z)\phi(q(z)).$$

The result follows by an application of Lemma 1.1. □

Remark 2.2. (i) Suppose that the function f is analytic in a bounded domain D and continuous on $\overline{D}$.

(ii) Suppose that $u(x, y)$ is a real part of non-constant analytic function f on a bounded domain D and $u(x, y)$ is lower/upper bounded in D. If either of the conditions (i), (ii) hold, then the real part of an analytic function f attains its minimum/maximum value, respectively on the boundary of D, by Minimum/ Maximum Modulus Theorem.

As a consequence of the condition (ii) that has been used in this paper, we have to incorporate certain conditions on the parameters involved in the differential subordination implications in order to satisfy the two conditions of Lemma 1.1.

We give the following example to illustrate Remark 2.2.

Example 2.3. (i) Consider the function $f(z) = z/(1-z)$, $(z \in \mathbb{D})$. Now, to obtain the maximum/minimum value of the real part of $f(z)$ on the bounded domain $\mathbb{D}$, either of the above two conditions of Remark 2.2 should hold. Clearly, (i) fails as the function $f(z)$ is non continuous on $\overline{\mathbb{D}}$, so we apply (ii). For $z = x + iy$ $(x^2 + y^2 < 1)$, we have

$$\mathrm{Re}(f(z)) = \frac{x - x^2 - y^2}{(1-x)^2 + y^2} =: f(x, y).$$

The function $f(x, y)$ is unbounded when x tends to 1 and y tends or equal to 0. These are the only two possibilities for the function to be unbounded in $\mathbb{D}$. For this, we consider two cases:

Case 1: Let $y = 0$ in $f(x, y)$, then we get

$$f(x) = x/(1-x).$$

Clearly, $f(x)$ is unbounded when x tends to 1. As a result, we assume $x = 1-h$, $(h \to 0)$ in $f(x)$, we get

$$\lim_{h\to 0} f(1-h) = \lim_{h\to 0} \frac{1-h}{h} \to +\infty.$$

This shows that the real part of $f(z)$ is not upper bounded.

Case 2: Let $x = 1 - h$ in $f(x, y)$, $(h \to 0)$, we get

$$\lim_{h\to 0} f(1-h, y) = \lim_{h\to 0} \frac{-h - h^2 + 2h - y^2}{h^2 + y^2} \to -1, \quad \forall y.$$

Thus, in particular for those y such that $(1-h)^2 + y^2 < 1$ holds, $f(x, y)$ is bounded.

We conclude that the function is lower bounded but not upper bounded. Thus, we can only obtain minimum of real part of $f(z)$ by evaluating on the boundary of $\mathbb{D}$. For $z = e^{i\theta}$ $(\theta \in [-\pi, \pi])$, we get

$$\mathrm{Re}(f(e^{i\theta})) = -1/2.$$

We infer that minimum value of real part of $f(z)$ is $-1/2$ and maximum value can not be obtained by evaluating on the boundary of $\mathbb{D}$ as it fails to be upper bounded.

(ii) Consider the Koebe function $K(z) := z/(1-z)^2$. It maps the open unit disk $\mathbb{D}$ onto the entire complex plane minus the slit along the negative real axis from $w = -\infty$ to $w = -1/4$. Thus, real part of the Koebe function is unbounded, which can also be verified. For $z = x + iy$ $(x^2 + y^2 < 1)$, we have

$$\mathrm{Re}(K(z)) = \frac{x - 2x^2 + x^3 - 2y^2 + xy^2}{(1-x)^2 + y^2} =: f(x, y).$$

It is trivial to say that the real part of $K(z)$ could only be unbounded when z tends to 1. Equivalently, we can say $f(x, y)$ is unbounded when x tends to 1 and y tends to or equal to 0. This is a complex plane so x can tend to 1 from

all the directions, unlike on the real line where it can only tend along the real axis. Thus, we consider two cases:

Case 1: Let $y = 0$ in $f(x, y)$ and proceeding same as in the case 1 in part (i), we assume $x = 1 - h$, $(h \to 0)$ in $f(x)$, we get

$$\lim_{h\to 0} f(1-h) = \lim_{h\to 0} \frac{1-h}{h^2} \to +\infty.$$

Thus, $f(x, y)$ is not upper bounded in this case.

Case 2: Let $x = 1 - h$ in $f(x, y)$ $(h \to 0)$, we get

$$\lim_{h\to 0} f(1-h, y) = \frac{-1}{y^2} \to -\infty,\ y = 0.$$

Clearly, it is not lower bounded in this case.

Now, let $z = e^{i\theta}$ $(\theta \in [-\pi, \pi])$, we get

$$\mathrm{Re}(z/(1-z)^2) = -1/(4\sin^2\theta) := g(\theta).$$

We infer that the maximum value of $g(\theta)$ is $-1/4$, attained at $\theta = \pi/2$. Therefore, we conclude that $\max \mathrm{Re}(K(z)) = -1/4$, which is absurd as clearly from both the cases 1 and 2, we observe that real part of $K(z)$ is unbounded. Thus, we cannot evaluate the real part of $K(z)$ on the boundary of $\mathbb{D}$ to find the maximum value. This is due to the failure of both the conditions (i) and (ii) in Remark 2.2 for the Koebe function. Therefore, the maximum value of the real part of $K(z)$ is not attained on the boundary of $\mathbb{D}$.

Taking α, β, γ and δ to be real numbers in the above Theorem 2.1, we obtain the following results:

Corollary 2.4. *Let $0 \le \lambda \le 1$ and $\gamma\delta \le 0$. (i) For $0 \le \nu \le 1/2$, let $(\alpha + \nu\beta)/\delta \ge \nu/(2(1-\nu))$ and $1 + 2(1-\nu)\beta/\delta > 0$. (ii) For $1/2 \le \nu < 1$, let $\alpha + \nu\beta/\delta \ge (1-\nu)/(2\nu)$ and $2\beta/\delta > (\nu-1)/\nu^2$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$\begin{aligned} & p^\lambda(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \\ & \prec \left(\frac{1+(1-2\nu)z}{1-z}\right)^\lambda \left(\alpha + \beta\left(\frac{1+(1-2\nu)z}{1-z}\right)\right. \\ & \left. + \frac{\gamma(1-z)}{1+(1-2\nu)z} + \delta\left(\frac{(1-2\nu)z}{1+(1-2\nu)z} + \frac{z}{1-z}\right)\right) =: h(z), \end{aligned} \tag{6}$$

then $\mathrm{Re}\ p(z) > \nu$.

Proof. Let $q(z) = (1 + (1-2\nu)z)/(1-z)$ in Theorem 2.1. Then we have,

$$Q(z) = \frac{2(1-\nu)z}{(1-z)^{1+\lambda}(1+(1-2\nu)z)^{1-\lambda}},$$

and

$$\frac{zQ'(z)}{Q(z)} = 1 + (1+\lambda)\frac{z}{1-z} - (1-\lambda)\frac{(1-2\nu)z}{1+(1-2\nu)z} =: K(z).$$

If $z = e^{i\theta}$, where $\theta \in [-\pi, \pi]$, we have

$$\operatorname{Re}(K(z)) = \left(\frac{1-\lambda}{2}\right)\left(\frac{\nu(1-\nu)}{\nu^2 + (1-2\nu)\cos^2(\theta/2)}\right). \tag{7}$$

Since $1-\lambda$, $\nu(1-\nu)$, $\nu^2 + (1-2\nu)\cos^2(\theta/2) \geq 0$, it follows that Q is starlike (univalent) in $\mathbb{D}$. Also, from condition (2) of Theorem 2.1, we have

$$\begin{aligned} H(z) := \frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}\left(\frac{1+(1-2\nu)z}{1-z}\right) + \frac{\gamma(\lambda-1)}{\delta}\left(\frac{1-z}{1+(1-2\nu)z}\right) \\ + (\lambda-1)\left(\frac{(1-2\nu)z}{1+(1-2\nu)z} + \frac{z}{1-z}\right) + \left(1 + \frac{2z}{1-z}\right). \end{aligned}$$

For $z = e^{i\theta}$ $(\theta \in [-\pi, \pi])$, real part of $H(z)$ reduces to

$$1 + \frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}\nu + \frac{\gamma(\lambda-1)}{\delta}A(\theta) - \frac{\lambda+1}{2} + (\lambda-1)(1-2\nu)B(\theta) =: L(\theta),$$

where $A(\theta) = (\nu\sin^2(\theta/2))/(\nu^2 + (1-2\nu)\cos^2(\theta/2))$ and $B(\theta) = (1-2\nu+\cos\theta)/(2(1-2\nu+\cos\theta+2\nu(\nu-\cos\theta)))$. Also, considering Remark 2.2, we need to have

$$\gamma/\delta < 1/2 \quad \text{and} \quad 1 + 2\beta(1-\nu)/\delta > 0, \tag{8}$$

which is trivial as the given conditions $\gamma\delta \leq 0$ and $2\beta/\delta > (\nu-1)/\nu^2$ imply the conditions in the equation (8), respectively. Now, to complete the proof it suffices to show that $L(\theta) \geq 0$ for the conditions given in the hypothesis. For this, we consider the following cases:

Case 1: Consider $0 \leq \nu \leq 1/2$, then $(1-2\nu) \geq 0$. As $0 \leq \lambda \leq 1$ and $\gamma\delta \leq 0$, we take into account the minimum value of $A(\theta) = 0$ and the maximum value of $B(\theta) = 1/(2(1-\nu))$, that are attained at $\theta = 0$ by the second derivative test. Thus, we get

$$L(\theta) \geq \lambda\left(\frac{\alpha}{\delta} + \frac{\beta\nu}{\delta} + \frac{1-2\nu}{2(1-\nu)} - \frac{1}{2}\right) + \frac{1}{2} + \frac{\beta\nu}{\delta} - \frac{1-2\nu}{2(1-\nu)} \geq 0,$$

which is possible when $(\alpha+\beta\nu)/\delta \geq \nu/(2(1-\nu))$.

Case 2: Consider $1/2 \leq \nu < 1$ then $(1-2\nu) \leq 0$. Similarly, taking the range of λ and $\gamma\delta$ into consideration we take into account the minimum value of $B(\theta) = -1/(2\nu)$ which is attained at $\theta = \pi$ by the second derivative test and let $A(\theta)$ to be same as in the case 1. Thus, we get

$$L(\theta) \geq \lambda\left(\frac{\alpha}{\delta} + \frac{\beta\nu}{\delta} - \frac{1-2\nu}{2\nu} - \frac{1}{2}\right) + \frac{1}{2} + \frac{\beta\nu}{\delta} + \frac{1-2\nu}{2\nu} \geq 0,$$

which is possible when $(\alpha+\beta\nu)/\delta \geq (1-\nu)/(2\nu)$ and $\beta/\delta > (\nu-1)/2\nu^2$. With this, we complete the proof. □

We obtain the relation between the class of starlike of reciprocal order ν and $\mathcal{N}(\kappa)$ in the following corollary:

Corollary 2.5. *Let $f \in \mathcal{A}$. The function f is starlike of reciprocal order ν if*

(a) $f \in \mathcal{N}((3-\nu)/2\nu)$, $0 < \nu \leq 3/4$.
(b) $f \in \mathcal{N}(\nu/2(1-\nu))$, $3/4 \leq \nu < 1$.

Proof. Let $p(z) = f(z)/(zf'(z))$, $\beta = \lambda = 0$, $\gamma = -\delta = 1$ and $\alpha = \min((\nu - 1)/(2\nu); \nu/(2(\nu-1)))$ in Corollary 2.4, we get, if f satisfies

$$1 + \frac{zf''(z)}{f'(z)} \prec \frac{1+2(\nu-1)z}{1+(1-2\nu)z} - \frac{z}{1-z} =: S(z), \tag{9}$$

then $\operatorname{Re}(f(z)/(zf'(z))) > \nu$. To complete the proof it suffices to show that $\operatorname{Re}(S(z)) < (3-\nu)/(2\nu)$ for $0 < \nu \leq 3/4$ and $\operatorname{Re}(S(z)) < \nu/(2(1-\nu))$ for $3/4 \leq \nu < 1$. Since, $S(z)$ is upper bounded, evaluating it on the boundary of $\mathbb{D}$, we obtain

$$\operatorname{Re}(S(e^{i\theta})) = \frac{1 - 6\nu + 4\nu^2 + \cos(\theta)}{-2 + 4\nu - 4\nu^2 - 2\cos(\theta) + 4\nu\cos(\theta)} + \frac{1}{2} =: g(\theta).$$

A calculation shows that $g''(\theta)_{\theta=\pi} = (1-\nu)(4\nu-3)/(4\nu^3)$ and $g''(\theta)_{\theta=0} = \nu(4\nu-3)/(4(\nu-1)^3)$.

(a) For $0 < \nu \leq 3/4$, $\max g(\theta) = (3-\nu)/(2\nu)$, attained at $\theta = \pi$ and we obtain

$$\operatorname{Re}\left(1 + \frac{zf''(z)}{f'(z)}\right) < \frac{3-\nu}{2\nu},$$

equivalently, $f \in \mathcal{N}((3-\nu)/(2\nu))$. This completes the proof for part (a).

(b) For $3/4 \leq \nu < 1$, $\max g(\theta) = \nu/(2(1-\nu))$, attained at $\theta = 0$ and we obtain

$$\operatorname{Re}\left(1 + \frac{zf''(z)}{f'(z)}\right) < \frac{\nu}{(2(1-\nu))},$$

equivalently, $f \in \mathcal{N}(\nu/2(1-\nu))$. This completes the proof for part (b). □

Taking $\nu = 0$ in Corollary 2.4, we get the following result:

Corollary 2.6. *Let $1 + 2\beta/\delta > 0$, $\gamma\delta \leq 0$, $\alpha\delta \geq 0$ and $0 \leq \lambda \leq 1$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^\lambda(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \prec \left(\frac{1+z}{1-z}\right)^\lambda \left(\alpha + \beta\left(\frac{1+z}{1-z}\right) + \gamma\left(\frac{1-z}{1+z}\right) + \frac{2\delta z}{1-z^2}\right),$$

then $\operatorname{Re} p(z) > 0$.

Taking $\nu = 1/2$ in Corollary 2.4, we get the following result:

Corollary 2.7. *Let $1 + \beta/\delta > 0$, $\gamma\delta \leq 0$, $-1 + (2\alpha+\beta)/\delta \geq 0$ and $0 \leq \lambda \leq 1$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^\lambda(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right)$$

$$\prec \left(\frac{1}{1-z}\right)^{\lambda}\left(\alpha+\beta\frac{1}{1-z}+\gamma(1-z)+\frac{\delta z}{1-z}\right),$$

then $\operatorname{Re} p(z) > 1/2$.

Considering oblique sector function q in Theorem 2.1, which is defined as

$$q(z)=\left(\frac{1+cz}{1-z}\right)^{\frac{\eta_1+\eta_2}{2}} \qquad (0<\eta_1,\eta_2\le 1),$$

where c and η are as defined in the equation (2). We have $\operatorname{Re}(zQ'(z)/Q(z)) > 0$ for given η and λ from [7, Theorem 2, p. 6], where Q is as defined in the condition (1) of Theorem 2.1. Therefore, $Q(z)$ is starlike (univalent) in $\mathbb{D}$. Also, from condition (2) of Theorem 2.1, we have

$$H(z)=\left(\frac{\alpha\lambda}{\delta}+\frac{\beta(\lambda+1)}{\delta}\left(\frac{1+cz}{1-z}\right)^{\frac{\eta_1+\eta_2}{2}}+\frac{\gamma(\lambda-1)}{\delta}\left(\frac{1-z}{1+cz}\right)^{\frac{\eta_1+\eta_2}{2}}+\frac{zQ'(z)}{Q(z)}\right).$$

In view of the fact that $q(z)\in\mathcal{P}$ implies $1/q(z)\in\mathcal{P}$, we get

$$\operatorname{Re}(H(z)) > \alpha\lambda/\delta \ge 0,$$

provided $\beta(\lambda+1)/\delta$, $\gamma(\lambda-1)/\delta$, $\alpha\lambda/\delta \ge 0$. Therefore, both the conditions of Theorem 2.1 get satisfied and we get the result as follows:

Corollary 2.8. *Let* $\alpha\lambda/\delta\ge 0$, $\beta(\lambda+1)/\delta\ge 0$, $\gamma(\lambda-1)/\delta\ge 0$, $0<\eta_1,\eta_2\le 1$ *and* $|\lambda|\le 2/(\eta_1+\eta_2)$. *If* $p(z)\in\mathcal{A}_0$ *satisfies*

$$\begin{aligned} &p^{\lambda}(z)\left(\alpha+\beta p(z)+\frac{\gamma}{p(z)}+\delta\frac{zp'(z)}{p(z)}\right)\\ &\prec \left(\frac{1+cz}{1-z}\right)^{\frac{(\eta_1+\eta_2)\lambda}{2}}\left(\alpha+\beta\left(\frac{1+cz}{1-z}\right)^{\frac{\eta_1+\eta_2}{2}}+\gamma\left(\frac{1-z}{1+cz}\right)^{\frac{\eta_1+\eta_2}{2}}\right.\\ &\left.+\frac{\eta_1+\eta_2}{2}\left(\frac{(1+c)z}{(1+cz)(1-z)}\right)\right), \end{aligned}$$

then $p(z)\prec((1+cz)/(1-z))^{(\eta_1+\eta_2)/2}$.

Remark 2.9. Taking $\alpha=\gamma=0$ and $\delta=1$, Corollary 2.8 is the result obtained by [7].

Letting $\eta_1=\eta_2$ and $c=1$ in Corollary 2.8, we have the following result:

Corollary 2.10. *Let* $\alpha\lambda/\delta\ge 0$, $\beta(\lambda+1)/\delta\ge 0$, $\gamma(\lambda-1)/\delta\ge 0$, $0<\eta\le 1$ *and* $|\lambda|\le 1/\eta$. *If* $p(z)\in\mathcal{A}_0$ *satisfies*

$$\begin{aligned} p^{\lambda}(z)\left(\alpha+\beta p(z)+\frac{\gamma}{p(z)}+\delta\frac{zp'(z)}{p(z)}\right) &\prec \left(\frac{1+z}{1-z}\right)^{\eta\lambda}\left(\alpha+\beta\left(\frac{1+z}{1-z}\right)^{\eta}\right.\\ &\left.+\gamma\left(\frac{1-z}{1+z}\right)^{\eta}+\frac{2\delta\eta z}{1-z^2}\right)=:h(z), \end{aligned} \tag{10}$$

then $p(z) \prec ((1+z)/(1-z))^{\eta}$.

Taking $\lambda = 1$ and $\gamma = 0$ in Corollary 2.10, we have the following result:

Corollary 2.11. *Let* $\alpha\delta$, $\beta\delta \geq 0$. *If* $p(z) \in \mathcal{A}_0$ *and*

$$\alpha p(z) + \beta p(z)^2 + \delta z p'(z) \prec \alpha \left(\frac{1+z}{1-z}\right)^{\eta} + \beta \left(\frac{1+z}{1-z}\right)^{2\eta} + \frac{2\delta\eta z}{1-z^2} \left(\frac{1+z}{1-z}\right)^{\eta},$$

then $|\arg p(z)| < \eta\pi/2$.

Remark 2.12. Corollary 2.11 is the result obtained by Ravichandran and Kumar [16] for $\alpha\delta$, $\beta\delta > 0$.

Taking $\eta = 1$ in Corollary 2.10, we get the following result:

Corollary 2.13. *Let* $\alpha\lambda/\delta \geq 0$, $\beta(\lambda+1)/\delta \geq 0$, $\gamma(\lambda-1)/\delta \geq 0$ *and* $|\lambda| \leq 1$. *If* $p(z) \in \mathcal{A}_0$ *satisfies*

$$p^{\lambda}(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \prec \left(\frac{1+z}{1-z}\right)^{\lambda}\left(\alpha + \beta\left(\frac{1+z}{1-z}\right) + \gamma\left(\frac{1-z}{1+z}\right) + \frac{2\delta z}{1-z^2}\right),$$

then $\operatorname{Re} p(z) > 0$.

The argument estimation of the function h as given by the right hand side of the subordination (10), gives the reformulation of Corollary 2.10 with $\gamma = 0$ and $0 \leq \lambda \leq 1/\eta$ as follows:

Corollary 2.14. *Let* α, $\beta \geq 0$ *and* δ *be a positive real number. If* $p(z) \in \mathcal{A}_0$ *satisfies*

$$\left|\arg\left(p^{\lambda}(z)\left(\alpha + \beta p(z) + \delta\frac{zp'(z)}{p(z)}\right)\right)\right| < \frac{\pi}{2}\zeta, \tag{11}$$

where (i) $\zeta = \eta\lambda$, *whenever* $0 \leq \eta \leq 1/2$, (ii) $\zeta = \eta\lambda + 1/2$, *whenever* $1/2 \leq \eta \leq 1$ *and* $\delta\eta \geq \alpha$, (iii) $\zeta = \eta\lambda + 2/\pi \tan^{-1}(\delta\eta/\alpha)$, *whenever* $1/2 \leq \eta \leq 1$ *and* $\alpha \geq \delta\eta$, *then*

$$|\arg p(z)| < \frac{\pi}{2}\eta.$$

Proof. Here, $h(z)$ is given by

$$h(z) := \left(\frac{1+z}{1-z}\right)^{\eta\lambda}\left(\alpha + \beta\left(\frac{1+z}{1-z}\right)^{\eta} + \frac{2\delta\eta z}{1-z^2}\right).$$

Consider

$$\begin{aligned} h(e^{i\theta}) &= \left(i\cot\frac{\theta}{2}\right)^{\eta\lambda}\left(\alpha + \beta\left(i\cot\frac{\theta}{2}\right)^{\eta} + i\frac{\delta\eta}{\sin\theta}\right) \\ &= \left|\cot\frac{\theta}{2}\right|^{\eta\lambda} e^{\pm i\eta\lambda\pi/2}\left(\alpha + \beta\left|\cot\frac{\theta}{2}\right|^{\eta} e^{\pm i\eta\pi/2} + i\frac{\delta\eta}{\sin\theta}\right). \end{aligned}$$

Note that the '+' sign comes for $0 < \theta < \pi$ and the '$-$' sign comes for $-\pi < \theta < 0$. Also, we observe that the real and the imaginary part of $h(e^{i\theta})$ is an even and odd function of θ, respectively. Thus, we will consider $0 < \theta < \pi$. Then, we have the following

$$\begin{aligned}\arg h(e^{i\theta}) &= \frac{\pi}{2}\eta\lambda + \arg\left(\alpha + \beta\left|\cot\frac{\theta}{2}\right|^{n} e^{i\eta\pi/2} + i\frac{\delta\eta}{\sin\theta}\right)\\ &= \frac{\pi}{2}\eta\lambda + \tan^{-1}\left(\frac{\beta|\cot(\theta/2)|^{\eta}\sin(\eta\pi/2) + \delta\eta/\sin(\theta)}{\alpha + \beta|\cot(\theta/2)|^{\eta}\cos(\eta\pi/2)}\right)\\ &\geq \frac{\pi}{2}\eta\lambda + \tan^{-1}\left(\frac{\beta s^{\eta}\sin(\eta\pi/2) + \delta\eta}{\alpha + \beta s^{\eta}\cos(\eta\pi/2)}\right) =: \frac{\pi}{2}\eta\lambda + \tan^{-1} g(s) \qquad (12)\end{aligned}$$

for α, $\beta \geq 0$, $\delta > 0$ and where $s = |\cot\theta/2|$, $(s_1 \leq s \leq s_2)$. Note that $s_1 \to 0$ and $s_2 \to \infty$. To complete the proof, it suffices to show that $\arg(he^{i\theta}) \geq \zeta\pi/2$, which equivalently implies $|\arg(p^{\lambda}(z)(\alpha + \beta p(z) + \delta z p'(z)/p(z)))| < \arg(h(z))$ which further yields $p(z) \prec ((1+z)/(1-z))^{\eta}$ as $p^{\lambda}(z)(\alpha + \beta p(z) + \delta z p'(z)/p(z)) \prec h(z)$ implies $p(z) \prec ((1+z)/(1-z))^{\eta}$ from Corollary 2.10. For this we consider two cases:

Case 1: Let $0 < \eta \leq 1/2$, then

$$g(s) \geq \frac{\beta s^{\eta}\sin(\eta\pi/2)}{\alpha + \beta s^{\eta}\cos(\eta\pi/2)} \geq \frac{\beta s_1^{\eta}\sin(\eta\pi/2)}{\alpha + \beta s_1^{\eta}\cos(\eta\pi/2)} \approx 0.$$

This completes the proof for the mentioned range of η in this case.

Case 2: Let $1/2 \leq \eta \leq 1$, then

$$g(s) \geq \frac{\beta s^{\eta}\cos(\eta\pi/2) + \delta\eta}{\alpha + \beta s^{\eta}\cos(\eta\pi/2)} =: l(s).$$

When $\delta\eta \geq \alpha$, $l(s)$ attains its minimum value at $s = s_2$. Thus, we obtain

$$l(s) \geq \frac{\beta s_2^{\eta}\sin(\eta\pi/2) + \delta\eta}{\alpha + \beta s_2^{\eta}\cos(\eta\pi/2)} \approx 1.$$

When $\alpha \geq \delta\eta$, $l(s)$ attains its minimum value at $s = s_1$. Thus, we obtain

$$l(s) \geq \frac{\beta s_1^{\eta}\sin(\eta\pi/2) + \delta\eta}{\alpha + \beta s_1^{\eta}\cos(\eta\pi/2)} \approx \frac{\delta\eta}{\alpha}.$$

This completes the proof for the mentioned range of η in this case. Thus, both the cases yield the desired condition $\arg h(e^{i\theta}) \geq \zeta\pi/2$ and hence the result. □

Remark 2.15. For $1/2 \leq \eta \leq 1$ and taking $\beta = 0$ in Corollary 2.14, we get the result obtained by Cho *et al.* [4], when restricting its range of η.

Corollary 2.16. *Let $\delta > 0$, $|\lambda\mu| \leq 1$ and $\alpha\lambda \geq -(A+B)$, where $A = \beta(\lambda+1)/e^{|\mu|} \geq 0$ and $B = \gamma(\lambda-1)/e^{|\mu|} \geq 0$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^{\lambda}(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \prec e^{\lambda\mu z}\left(\alpha + \beta e^{\mu z} + \gamma e^{-\mu z} + \delta\mu z\right),$$

then $p(z) \prec e^{\mu z}$, where μ is a non-zero real number such that $|\mu| \leq 1$.

Proof. This result follows from Theorem 2.1 by taking $q(z) = e^{\mu z}$. Then $Q(z) = \delta\mu z e^{\lambda\mu z}$ and by taking $z = x + iy$ $(x^2 + y^2 < 1)$, we get

$$\operatorname{Re}\left(\frac{zQ'(z)}{Q(z)}\right) = 1 + \lambda\mu x.$$

A simple computation shows that $Q(z)$ is starlike (univalent) in $\mathbb{D}$ whenever $|\lambda\mu| \leq 1$. Also, in view of the fact that $\operatorname{Re}(e^{\mu z})$, $\operatorname{Re}(e^{-\mu z}) > 1/e^{|\mu|}$, it follows that

$$\begin{aligned}
&\operatorname{Re}\left(\frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}q(z) + \frac{\gamma(\lambda-1)}{\delta}\frac{1}{q(z)} + (\lambda-1)\frac{zq'(z)}{q(z)} + \left(1 + \frac{zq''(z)}{q'(z)}\right)\right) \\
&= \operatorname{Re}\left(1 + \lambda\mu z + \frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}e^{\mu z} + \frac{\gamma(\lambda-1)}{\delta}e^{-\mu z}\right) \\
&> 1 + \lambda\mu x + \frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}\frac{1}{e^{|\mu|}} + \frac{\gamma(\lambda-1)}{\delta}\frac{1}{e^{|\mu|}} =: S(x)
\end{aligned}$$

for A, $B > 0$ and $\delta \geq 0$. For the given range of $|\lambda\mu|$, we consider two cases:

Case 1: Let $-1 \leq \lambda\mu \leq 0$, then $1 + \lambda\mu x > 1 + \lambda\mu \geq 0$ and we obtain

$$S(x) > \frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}\frac{1}{e^{|\mu|}} + \frac{\gamma(\lambda-1)}{\delta}\frac{1}{e^{|\mu|}}. \tag{13}$$

Case 2: Let $0 \leq \lambda\mu \leq 1$, then $1 + \lambda\mu x > 1 - \lambda\mu \geq 0$ and we obtain the same equation (13) as in the case 1.

Further $S(x) \geq 0$ if $\alpha\lambda \geq (\gamma - \beta - \lambda(\beta + \gamma))/e^{|\mu|}$, which is trivial from the conditions in the hypothesis. Thus, both the conditions of Theorem 2.1 get satisfied and hence the result. □

Taking $\mu = 1$ in Corollary 2.16, we obtain:

Corollary 2.17. *Let $\delta > 0$, $|\lambda| < 1$ and $\alpha\lambda \geq -(A + B)$, where $A = \beta(\lambda + 1)/e \geq 0$ and $B = \gamma(\lambda - 1)/e \geq 0$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^{\lambda}(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \prec e^{\lambda z}\left(\alpha + \beta e^{z} + \gamma e^{-z} + \delta z\right),$$

then $p(z) \prec e^{z}$.

Corollary 2.18. *Let $-2 \leq \lambda \leq 2$, $\gamma(\lambda-1) \geq 0$, $\beta(\lambda+1) \geq 0$, $\delta > \max(0; \sqrt{2}\gamma)$, $-(2\sqrt{2}\gamma + \delta)/4 < \alpha \leq -3\gamma/(2\sqrt{2})$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$\begin{aligned}
p^{\lambda}(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p(z)}\right) \prec\ & (1+z)^{\lambda/2}\left(\alpha + \beta\sqrt{1+z}\right. \\
& \left. + \frac{\gamma}{\sqrt{1+z}} + \frac{\delta z}{2(1+z)}\right),
\end{aligned}$$

then $p(z) \prec \sqrt{1+z}$.

Proof. Now, to achieve the desired result we apply Theorem 2.1 for $q(z) = \sqrt{1+z}$ and we get $Q(z) = \delta z(1+z)^{\lambda/2-1}/2$. Then

$$\frac{zQ'(z)}{Q(z)} = 1 + \left(\frac{\lambda}{2} - 1\right)\frac{z}{1+z} =: K(z).$$

In view of the fact that $K(z)$ is bounded only if $\lambda \leq 2$, we evaluate it on the boundary. Therefore, $z = e^{i\theta}$ ($\theta \in [-\pi, \pi]$) yields

$$\mathrm{Re}(K(e^{i\theta})) = \frac{\lambda+2}{4} \geq 0$$

for $\lambda \geq -2$. Clearly Q is starlike (univalent) in $\mathbb{D}$. Also, we have

(14)

$$\begin{aligned}
&\mathrm{Re}\left(\frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}q(z) + \frac{\gamma(\lambda-1)}{\delta}\frac{1}{q(z)} + (\lambda-1)\frac{zq'(z)}{q(z)} + \left(1 + \frac{zq''(z)}{q'(z)}\right)\right)\\
&= \mathrm{Re}\left(1 + \frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}\sqrt{1+z} + \frac{\gamma(\lambda-1)}{\delta}\frac{1}{\sqrt{1+z}} + \left(\frac{\lambda}{2} - 1\right)\frac{z}{1+z}\right)\\
&\geq 1 + \frac{\alpha\lambda}{\delta} + \frac{\gamma(\lambda-1)}{\delta}\frac{1}{\sqrt{2}} + \frac{\lambda}{4} - \frac{1}{2}\\
&= \lambda\left(\frac{1}{4} + \frac{\alpha}{\delta} + \frac{\gamma}{\sqrt{2}\delta}\right) + \frac{1}{2} - \frac{\gamma}{\sqrt{2}\delta} =: \lambda(r) + s,
\end{aligned}$$

since $\mathrm{Re}(\sqrt{1+z}) > 0$, $\mathrm{Re}(1/\sqrt{1+z}) > 1/\sqrt{2}$, $\beta(\lambda+1)/\delta$ and $\gamma(\lambda-1)/\delta \geq 0$. Now, to complete the proof we need to find the conditions on the parameters such that $\lambda(r) + s \geq 0$. For this, it suffices to show that

$$\lambda \geq \frac{\frac{\gamma}{\sqrt{2}\delta} - \frac{1}{2}}{\frac{\alpha}{\delta} + \frac{\gamma}{\sqrt{2}\delta} + \frac{1}{4}}, \tag{15}$$

which is a valid expression when $\alpha > -(2\sqrt{2}\gamma + \delta)/4$ and $\delta > 0$. Also, from the inequality $-2 \leq \lambda \leq 2$ we have $\alpha \leq -3\gamma/(2\sqrt{2})$ and thus, equation (15) holds true. And the derived range of α is meaningful only if $\delta \geq \sqrt{2}\gamma$. This completes the proof. □

The following lemmas will be needed to prove some further results.

Lemma 2.19. *For $-1 \leq B < A \leq 1$, the function $f(z) = (1+Az)/(1+Bz)$ satisfies $\min \mathrm{Re} f(z) = (1-A)/(1-B)$ and $\min \mathrm{Re}(1/f(z)) = (1+B)/(1+A)$.*

Proof. Consider

$$\mathrm{Re}(f(e^{i\theta})) = \frac{1 + (A+B)\cos\theta + AB}{1 + A^2 + 2A\cos\theta} =: K(\theta).$$

$K(\theta)$ attains its minimum value at $\theta = \pi$ by the second derivative test and the minimum value is given by

$$K(\pi) = \frac{1-A}{1-B}. \tag{16}$$

Now, consider $f(-z) = (1 - Az)/(1 - Bz)$. In view of the fact that the image of unit disk under $f(z)$ is same as $f(-z)$, from equation (16), we get

$$\min \operatorname{Re}\left(\frac{1+Bz}{1+Az}\right) = \frac{1+B}{1+A}$$

as $-A$ playing the role as B and $-B$ as A. This completes the proof. □

With the same technique we can also find the minimum and maximum value of the real part of the functions $(1-Bz)/(1+Bz)$ and $z(A-B)/((1+Az)(1+Bz))$, respectively as given in the following lemmas:

Lemma 2.20. *For $-1 \leq B < 1$, the minimum of real part of the function $f(z) = (1 - Bz)/(1 + Bz)$ is $(1 - |B|)/(1 + |B|)$.*

Lemma 2.21. *For $-1 < B < A \leq 1$. Consider the function $f(z) = (A - B)z/((1+Az)(1+Bz))$, then $\max \operatorname{Re} f(z) = (A-B)/(1+A)(1+B)$, whenever $(1 + AB)(1 - A)(1 - B) > 8AB$ and $\min \operatorname{Re} f(z) = (B - A)/((1 - A)(1 - B))$, whenever $(1 + AB)(1 + A)(1 + B) > 8AB$.*

Corollary 2.22. *Let $0 \leq \lambda \leq 1$ and $\gamma\delta \leq 0$. For $-1 < B < A \leq 1$, if $(1 + AB)(1 - A)(1 - B) > 8AB$, we assume $\alpha/\delta \geq (B - A)/((1 + A)(1 + B))$ and $\beta/\delta \geq (A - B)(1 - B)/((1 - A^2)(1 + B))$. Suppose $p(z) \in \mathcal{A}_0$ satisfies*

$$p^{\lambda}(z)\left(\alpha+\beta p(z)+\frac{\gamma}{p(z)}+\delta\frac{zp'(z)}{p(z)}\right) \prec \left(\frac{1+Az}{1+Bz}\right)^{\lambda}\left(\alpha+\beta\left(\frac{1+Az}{1+Bz}\right)\right.$$
$$\left.+\gamma\left(\frac{1+Bz}{1+Az}\right)+\delta\frac{(A-B)z}{(1+Az)(1+Bz)}\right),$$

then $p(z) \prec (1 + Az)/(1 + Bz)$.

Proof. The result is followed by taking $q(z) = (1+Az)/(1+Bz)$ in Theorem 2.1. Then, we have

$$Q(z) = \frac{\delta z(A-B)}{(1+Az)^{1-\lambda}(1+Bz)^{1+\lambda}},$$

and

$$\frac{zQ'(z)}{Q(z)} = 1 + (\lambda-1)\frac{Az}{1+Az} - (1+\lambda)\frac{Bz}{1+Bz} =: K(z).$$

As per Remark 2.2 for $-1 \leq \lambda \leq 1$, $K(e^{i\theta}) \geq 0$. Thus, clearly $Q(z)$ is starlike (univalent) in $\mathbb{D}$. Also, we have

(17)

$$\operatorname{Re}\left(\frac{\alpha\lambda}{\delta}+\frac{\beta(\lambda+1)}{\delta}q(z)+\frac{\gamma(\lambda-1)}{\delta}\frac{1}{q(z)}+(\lambda-1)\frac{zq'(z)}{q(z)}+\left(1+\frac{zq''(z)}{q'(z)}\right)\right)$$
$$= \operatorname{Re}\left(\frac{\alpha\lambda}{\delta}+\frac{\beta(\lambda+1)}{\delta}\left(\frac{1+Az}{1+Bz}\right)+\frac{\gamma(\lambda-1)}{\delta}\left(\frac{1+Bz}{1+Az}\right)\right.$$
$$\left.+(\lambda-1)\frac{(A-B)z}{(1+Az)(1+Bz)}+\frac{1-Bz}{1+Bz}\right) =: S(z).$$

Now, Lemmas 2.19, 2.20 and 2.21 yield

$$\begin{aligned} S(z) &\geq \frac{\alpha\lambda}{\delta} + \frac{\beta(\lambda+1)}{\delta}\left(\frac{1-A}{1-B}\right) + \frac{\gamma(\lambda-1)}{\delta}\left(\frac{1+B}{1+A}\right) \\ &\quad + (\lambda-1)\left(\frac{A-B}{(1+A)(1+B)}\right) + \left(\frac{1-|B|}{1+|B|}\right) \\ &= \lambda\left(\frac{\alpha}{\delta} + \frac{A-B}{(1+A)(1+B)}\right) + \frac{\beta(\lambda+1)}{\delta}\left(\frac{1-A}{1-B}\right) \\ &\quad - \left(\frac{A-B}{(1+A)(1+B)}\right) + \frac{1-|B|}{1+|B|} + \frac{\gamma(\lambda-1)}{\delta}\left(\frac{1+B}{1+A}\right) \end{aligned} \tag{18}$$

for $\beta\delta \geq 0$ and $\gamma\delta \leq 0$. To complete the proof it suffices to prove the second condition of Theorem 2.1. For this, we need to have equation (18) greater than or equal to 0, which is possible when $0 \leq \lambda \leq 1$, $\alpha/\delta \geq (B-A)/(1+A)(1+B)$ and $\beta/\delta \geq (A-B)(1-B)/((1-A^2)(1+B))$. Hence the result follows immediately. □

With $\lambda = 1$ and $q(z) = \phi_{PAR}(z) := 1 + (2/\pi^2)(\log((1+\sqrt{z})/(1-\sqrt{z})))^2$ in Theorem 2.1, we have the following result:

Corollary 2.23. *Let $\beta/\delta \geq \max(0; -\alpha/\delta)$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$\begin{aligned} &\alpha p(z) + \beta p(z)^2 + \delta z p'(z) \\ &\prec \left(1 + \frac{2}{\pi^2}\left(\log\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)^2\right)\left(\alpha + \beta\left(1 + \frac{2}{\pi^2}\left(\log\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)^2\right)\right. \\ &\quad \left. + \frac{4\delta}{\pi^2}\frac{\sqrt{z}}{1-z}\log\left(\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)\right), \end{aligned}$$

then $p(z) \prec 1 + (2/\pi^2)(\log((1+\sqrt{z})/(1-\sqrt{z})))^2$.

Proof. As we know that $\phi_{PAR}(z)$ is a convex function, thus by Alexander's theorem we have $Q(z) = \delta z q'(z)$ is starlike (univalent) in $\mathbb{D}$. Also, from condition (2) of Theorem 2.1, we have

$$\operatorname{Re}(H(z)) = \operatorname{Re}\left(\frac{\alpha}{\delta} + \frac{2\beta}{\delta}\left(1 + \frac{2}{\pi^2}\left(\log\left(\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)\right)^2\right) + 1 + \frac{zq''(z)}{q'(z)}\right). \tag{19}$$

Evaluating the equation (19) on $z = e^{i\theta}$ ($\theta \in [-\pi, \pi]$), we obtain

$$\begin{aligned} \operatorname{Re}(H(e^{i\theta})) &\geq \operatorname{Re}\left(\frac{\alpha}{\delta} + \frac{2\beta}{\delta}\left(1 + \frac{2}{\pi^2}(\log\left(i\cot(\theta/4)\right)^2)\right)\right) \\ &= \frac{\alpha}{\delta} + \frac{2\beta}{\delta}\left(\frac{1}{2} + \frac{2}{\pi^2}\log(\cot(\theta/4))^2\right) =: \frac{\alpha}{\delta} + \frac{2\beta}{\delta}g(\theta). \end{aligned} \tag{20}$$

For $\theta = \pi$, simple calculation shows that

$$\begin{aligned} g''(\theta) = {} & \csc^2(\theta/4)\log(\cot(\theta/4))/(4\pi^2) + (\csc^2(\theta/4)\sec(\theta/4)^2)/(4\pi^2) \\ & - (\log(\cot(\theta/4)))\sec^2(\theta/4)/(4\pi^2) > 0. \end{aligned}$$

Thus, $\mathrm{Re}(H(e^{i\theta})) \geq 0$ from equation (20), on substituting $g(\pi) = 1/2$, the minimum value of $g(\theta)$. Hence, both the conditions of Theorem 2.1 get satisfied and therefore the result follows. □

Theorem 2.24. *Let α, β, γ and $(0 \neq)\delta$ be complex numbers and λ be a real number. Let $q(z) \in \mathcal{A}_0$ be univalent in $\mathbb{D}$ and satisfy the following conditions for $z \in \mathbb{D}$:*

(1) *$Q(z) := \delta z q'(z) q^{\lambda-2}(z)$ be starlike (univalent),*

(2) *$\mathrm{Re}(H(z)) > 0$, where*

$$H(z) = \frac{\gamma(\lambda-1)}{\delta} + \frac{\alpha\lambda}{\delta} q(z) + \frac{\beta(\lambda+1)}{\delta} q^2(z) + (\lambda-2)\frac{zq'(z)}{q(z)} + \left(1 + \frac{zq''(z)}{q'(z)}\right).$$

If $p(z) \in \mathcal{A}_0$ satisfies

$$p^\lambda(z)\left(\alpha + \beta p(z) + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p^2(z)}\right) \prec q^\lambda(z)\left(\alpha + \beta q(z) + \frac{\gamma}{q(z)} + \delta\frac{zq'(z)}{q^2(z)}\right),$$

then $p(z) \prec q(z)$ and $q(z)$ is the best dominant.

The proof of this theorem is similar to that of Theorem 2.1 and therefore omitted.

Remark 2.25. Let $\lambda = \alpha = \gamma = \beta = 0$ and $\delta = 1$ in Theorem 2.24, we have the following result of Ravichandran and Kumar [16].

Corollary 2.26. *Let $q(z) \in \mathcal{A}_0$ be univalent in $\mathbb{D}$. Let $zq'(z)/q(z)^2$ be starlike. If $p(z) \in \mathcal{A}_0$ satisfies*

$$\frac{zp'(z)}{p(z)^2} \prec \frac{zq'(z)}{q(z)^2},$$

then $p(z) \prec q(z)$. The dominant q is the best dominant.

We take α, β, γ and δ to be real numbers, for proving all the following results. For the next result, we further assume $\beta = 0$ and let $q(z)$ be the function defined by $q(z) = (1+(1-2\nu)z)/(1-z)$ for $0 \leq \nu < 1$ in Theorem 2.24, then we obtain function Q as follows:

$$Q(z) = \frac{2(1-\alpha)z}{(1-z)^\lambda(1+(1-2\nu)z)^{2-\lambda}}.$$

Proceeding as in Corollary 2.4, we have $Q(z)$ starlike (univalent) in $\mathbb{D}$ for $1 \leq \lambda \leq 2$. If we take $z = e^{i\theta}$ $(\theta \in [-\pi, \pi])$, we have

$$\begin{aligned} &\mathrm{Re}\left(\frac{\gamma(\lambda-1)}{\delta} + \frac{\alpha\lambda}{\delta} q(z) + (\lambda-2)\frac{zq'(z)}{q(z)} + 1 + \frac{zq''(z)}{q'(z)}\right) \\ &= 1 + \frac{\gamma(\lambda-1)}{\delta} + \frac{\alpha\lambda}{\delta}\nu + (\lambda-2)(1-2\nu)B(\theta) - \frac{\lambda}{2} =: S(\theta), \end{aligned} \tag{21}$$

where $B(\theta)$ is as defined in the proof of Corollary 2.4 and let $1+2\alpha(1-\nu)/\delta > 0$ as per Remark 2.2. In order to apply Theorem 2.24, we need to find the range of the parameters such that $S(\theta) \geq 0$. For this we consider two cases:

Case 1: Consider $0 \leq \nu \leq 1/2$, then $(1-2\nu) \geq 0$. Since $1 \leq \lambda \leq 2$, we take into account the maximum value of $B(\theta)$. Thus, we get

$$S(\theta) \geq \lambda \left(\frac{\gamma}{\delta} + \frac{\alpha\nu}{\delta} + \frac{1-2\nu}{2(1-\nu)} - \frac{1}{2} \right) + 1 - \frac{\gamma}{\delta} - \frac{1-2\nu}{1-\nu} =: \lambda(r) + s.$$

It suffices to find the conditions on the parameters for which $\lambda(r) + s \geq 0$. For this, either (i) we show

$$\lambda \geq \frac{\frac{\gamma}{\delta} + \frac{1-2\nu}{1-\nu} - 1}{\frac{\gamma}{\delta} + \frac{\alpha\nu}{\delta} + \frac{1-2\nu}{2(1-\nu)} - \frac{1}{2}}, \tag{22}$$

which is a valid expression only if $(\gamma + \alpha\nu)/\delta > \nu/(2(1-\nu))$. Since $1 \leq \lambda$, the inequality (22) holds if $1 + 2\alpha(1-\nu)/\delta > 0$, or (ii) let $r = 0$, simple calculation yields $S(\theta) \geq 0$.

Case 2: Consider $1/2 \leq \nu < 1$, then $(1-2\nu) \leq 0$. Therefore, we take into account the minimum value of $B(\theta)$. Thus, we get

$$S(\theta) \geq \lambda \left(\frac{\gamma}{\delta} + \frac{\alpha\nu}{\delta} - \frac{1-2\nu}{2\nu} - \frac{1}{2} \right) + 1 - \frac{\gamma}{\delta} + \frac{1-2\nu}{\nu}.$$

Again, proceeding as in the case 1, we assume $(\gamma + \alpha\nu)/\delta \geq (1-\nu)/(2\nu)$ and $\alpha/\delta > -1/(2(1-\nu))$. Thus, $\alpha/\delta \geq \max(-1/(2(1-\nu)); (\nu-1)/2\nu^2) = (\nu-1)/(2\nu^2)$. Hence the result follows as:

Corollary 2.27. *Let $0 \leq \nu < 1$ and $1 \leq \lambda \leq 2$. (i) For $0 \leq \nu \leq 1/2$, let $1 + 2(1-\nu)\alpha/\delta > 0$ and $(\gamma + \nu\alpha)/\delta \geq \nu/(2(1-\nu))$. (ii) For $1/2 \leq \nu < 1$, let $2\alpha/\delta > (\nu-1)/\nu^2$ and $(\gamma + \nu\alpha)/\delta \geq (1-\nu)/(2\nu)$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^\lambda(z) \left(\alpha + \frac{\gamma}{p(z)} + \delta \frac{zp'(z)}{p^2(z)} \right) \prec \left(\frac{1 + (1-2\nu)z}{1-z} \right)^\lambda \left(\alpha + \frac{\gamma(1-z)}{1 + (1-2\nu)z} + \frac{2\delta(1-\nu)z}{(1 + (1-2\nu z))^2} \right),$$

then $\operatorname{Re} p(z) > \nu$.

Taking $\nu = 0$ in the above Corollary 2.27, we have the following result:

Corollary 2.28. *Let $1 + 2\alpha/\delta > 0$, $\gamma\delta \geq 0$ and $1 \leq \lambda \leq 2$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^\lambda(z) \left(\alpha + \frac{\gamma}{p(z)} + \delta \frac{zp'(z)}{p^2(z)} \right) \prec \left(\frac{1+z}{1-z} \right)^\lambda \left(\alpha + \gamma \frac{1-z}{1+z} + \frac{2\delta z}{(1+z)^2} \right),$$

then $\operatorname{Re} p(z) > 0$.

Taking $\nu = 1/2$ in Corollary 2.27, we obtain the following result:

Corollary 2.29. *Let $1 + \alpha/\delta > 0$, $(2\gamma + \alpha)/\delta \geq 1$ and $1 \leq \lambda \leq 2$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^\lambda(z) \left(\alpha + \frac{\gamma}{p(z)} + \delta \frac{zp'(z)}{p^2(z)} \right) \prec \left(\frac{1}{1-z} \right)^\lambda (\alpha + \gamma(1-z) + \delta z),$$

then Re $p(z) > 1/2$.

Remark 2.30. By taking $\alpha = \gamma = 0$ and $\lambda = 1$ in Corollary 2.6, or by taking $\lambda = 2$, $\gamma = 0$ and $\alpha = \beta$ in Corollary 2.28, we have the following result of Nunokawa *et al*. [14]:

Corollary 2.31. *Let* $1 + 2\beta/\delta > 0$. *If* $p(z) \in \mathcal{A}_0$ *satisfies*

$$\beta p(z)^2 + \delta z p'(z) \prec \beta \left(\frac{1+z}{1-z}\right)^2 + \frac{2\delta z}{(1-z)^2},$$

then Re $p(z) > 0$.

Corollary 2.32. *If* $p(z) \in \mathcal{A}_0$ *satisfies*

$$p(z) + \frac{zp'(z)}{p(z)} \prec \mathcal{R}(z),$$

where $\mathcal{R}$ *is the open door mapping, then* $p(z) \prec (1+z)/(1-z)$.

Proof. Let $\lambda = \gamma = \alpha = 0$ and $\beta = \delta = 1$ in Corollary 2.6 or by assuming $\lambda = \alpha = \delta = 1$ and $\gamma = 0$ in Corollary 2.28, we get

$$p(z) + \frac{zp'(z)}{p(z)} \prec \frac{1+z}{1-z} + \frac{2z}{1-z^2} =: \mathcal{R}(z).$$

This completes the proof. □

Remark 2.33. Corollary 2.32 is the result obtained by Nunokawa *et al*. [14].

For the following result, we one again assume $\beta = 0$ in Theorem 2.24.

Corollary 2.34. *Let* $\alpha\lambda/\delta \geq 0$, $\gamma(\lambda-1)/\delta \geq 0$, $0 < \eta \leq 1$ *and* $-1 < \eta\lambda \leq 2$. *If* $p(z) \in \mathcal{A}_0$ *satisfies*

$$\begin{aligned} & p^\lambda(z)\left(\alpha + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p^2(z)}\right) \\ & \prec \left(\frac{1+z}{1-z}\right)^{\eta\lambda}\left(\alpha + \gamma\left(\frac{1-z}{1+z}\right)^{\eta} + \frac{2\delta\eta z}{(1-z)^{1-\eta}(1+z)^{1+\eta}}\right) \\ & =: h(z), \end{aligned} \tag{23}$$

then $p(z) \prec ((1+z)/(1-z))^\eta$.

Proof. Let $q(z) = ((1+z)/(1-z))^\eta$ in Theorem 2.24, then $Q(z)$ is given by

$$Q(z) = \frac{2\delta\eta z}{1-z^2}\left(\frac{1+z}{1-z}\right)^{\eta(\lambda-1)}$$

and

$$\frac{zQ'(z)}{Q(z)} = 1 + \frac{2z^2}{1-z^2} + \frac{2\eta(\lambda-1)z}{1-z^2} =: K(z).$$

Since $K(z)$ is lower bounded for $-1 < \eta\lambda \leq 2$, we evaluate $K(z)$ on the boundary of $\mathbb{D}$ which implies Re $K(e^{i\theta}) \geq 0$. Thus, $Q(z)$ is starlike (univalent) in $\mathbb{D}$. Also, from condition (2) of Theorem 2.24, we have

$$\mathrm{Re}(H(z)) = \mathrm{Re}\left(1 + \frac{\gamma(\lambda-1)}{\delta} + \frac{\alpha\lambda}{\delta}\left(\frac{1+z}{1-z}\right)^{\eta} + \frac{2z(z+(\lambda-1)\eta)}{1-z^2}\right).$$

Some computation shows that for $z = e^{i\theta}$ $(\theta \in [-\pi, \pi])$, we obtain

$$\mathrm{Re}(H(e^{i\theta})) = \frac{\gamma(\lambda-1)}{\delta} \geq 0,$$

when $\alpha\lambda/\delta$, $\gamma(\lambda-1)/\delta \geq 0$. □

Taking $\eta = 1$ in Corollary 2.34, we obtain the following result:

Corollary 2.35. *Let $\alpha\lambda/\delta$, $\gamma(\lambda-1)/\delta \geq 0$ and $-1 < \lambda \leq 2$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^{\lambda}(z)\left(\alpha + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p^2(z)}\right) \prec \left(\frac{1+z}{1-z}\right)^{\lambda}\left(\alpha + \gamma\left(\frac{1-z}{1+z}\right) + \frac{2\delta z}{(1+z)^2}\right),$$

then Re $p(z) > 0$.

Here, we derive the argument relation between

$$p^{\lambda}(z)\left(\alpha + \gamma/p(z) + \delta z p'(z)/p^2(z)\right)$$

and $h(z)$, as defined by the equation (23), such that the subordination (23) holds for $0 < \eta \leq 1$. We assume $-1 < \lambda \leq 0$.

Corollary 2.36. *Let $\alpha \geq -\delta > 0$ and $\gamma \geq 0$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$\left|\arg\left(p^{\lambda}(z)\left(\alpha + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p^2(z)}\right)\right)\right| < \frac{\pi}{2}\zeta, \tag{24}$$

where

(i) $\zeta = -\eta\lambda$ *whenever* $0 < \eta \leq 1/2$,

(ii) $\zeta = -\eta\lambda + (2/\pi)\tan^{-1}(-\delta\eta\cos(\eta\pi/2)/\alpha)$ *whenever* $1/2 \leq \eta \leq 1$, *then*

$$|\arg p(z)| < \frac{\pi}{2}\eta.$$

Proof. A calculation shows that

$$\begin{aligned} h(e^{i\theta}) &= (i\cot\theta/2)^{\eta\lambda}\left(\alpha + \gamma\,(i\cot\theta/2)^{-\eta} + i\frac{\delta\eta}{\sin\theta}(i\cot\theta/2)^{-\eta}\right) \\ &= |\cot\theta/2|^{\eta\lambda}e^{\pm i\eta\lambda\pi/2}\Big(\alpha + \gamma|\cot\theta/2|^{-\eta}e^{\mp i\eta\pi/2} \\ &\quad + \frac{\delta\eta}{\sin\theta}|\cot\theta/2|^{-\eta}e^{i\pi/2(1\mp\eta)}\Big). \end{aligned}$$

As the '+' sign comes for $0 < \theta < \pi$ and the '−' sign comes for $-\pi < \theta < 0$. Also, we observe that the real and the imaginary part of $h(e^{i\theta})$ is an even and

odd function of θ, respectively. Thus, we will consider here $-\pi < \theta < 0$. Then, we have the following

$$\begin{aligned}\arg h(e^{i\theta}) &= -\frac{\pi}{2}\eta\lambda + \tan^{-1}\left(\frac{\gamma|\cot(\theta/2)|^{-\eta}\sin(\eta\pi/2) + \frac{\delta\eta}{\sin\theta}\cos(\eta\pi/2)}{\alpha + \gamma|\cot(\theta/2)|^{-\eta}\cos(\eta\pi/2) - \frac{\delta\eta}{\sin\theta}\sin(\eta\pi/2)}\right)\\ &\geq -\frac{\pi}{2}\eta\lambda + \tan^{-1}\left(\frac{\gamma s^{-\eta}\sin(\eta\pi/2) - \delta\eta\cos(\eta\pi/2)}{\alpha + \gamma s^{-\eta}\cos(\eta\pi/2)}\right)\\ &=: -\frac{\pi}{2}\eta\lambda + \tan^{-1}(g(s))\end{aligned}$$

for $\alpha \geq -\delta >$, $\gamma \geq 0$ and where $s = |\cot(\theta/2)|$, $(s_1 \leq s \leq s_2)$, $s_1 \to 0$ and $s_2 \to \infty$. Now, we consider two cases:

Case 1: Let $0 < \eta \leq 1/2$, then

$$g(s) \geq \left(\frac{\gamma s^{-\eta}\sin(\eta\pi/2)}{\alpha + \gamma s^{-\eta}\cos(\eta\pi/2)}\right) \geq \left(\frac{\gamma s_2^{-\eta}\sin(\eta\pi/2)}{\alpha + \gamma s_2^{-\eta}\cos(\eta\pi/2)}\right) \approx 0.$$

Case 2: Let $1/2 \leq \eta \leq 1$, then

$$g(s) \geq \left(\frac{\gamma s^{-\eta}\cos(\eta\pi/2) + \delta\eta\cos(\eta\pi/2)}{\alpha + \gamma s^{-\eta}\cos(\eta\pi/2)}\right) =: l(s)$$

as $\alpha \geq -\delta$, then

$$l(s) \geq \left(\frac{\gamma s_2^{-\eta}\cos(\eta\pi/2) + \delta\eta\cos(\eta\pi/2)}{\alpha + \gamma s_2^{-\eta}\cos(\eta\pi/2)}\right) \approx \frac{-\delta\eta\cos(\eta\pi/2)}{\alpha}.$$

Thus from both the cases, we get $\arg(h(e^{i\theta})) \geq \zeta\pi/2$, where ζ is as given in the hypothesis. We observe that condition (24) concludes that the subordination (23) holds. Also, the hypothesis of Corollary 2.34 gets satisfied, as a result we get

$$p(z) \prec ((1+z)/(1-z))^{\eta},$$

equivalently $|\arg(p(z))| < \eta\pi/2$. This completes the proof. □

The next result follows from Theorem 2.24 by taking $q(z) = e^{\mu z}$. Again assume $\beta = 0$.

Corollary 2.37. *Let $|(\lambda-1)\mu| \leq 1$, $\alpha\lambda/\delta \geq 0$ and $\gamma(\lambda-1)/\delta \geq 0$. If $p(z) \in \mathcal{A}_0$ satisfies*

$$p^{\lambda}(z)\left(\alpha + \frac{\gamma}{p(z)} + \delta\frac{zp'(z)}{p^2(z)}\right) \prec e^{\mu\lambda z}\left(\alpha + \frac{\gamma}{e^{\mu z}} + \delta\frac{\mu z}{e^{\mu z}}\right),$$

then $p(z) \prec e^{\mu z}$, where μ is a non-zero real number such that $|\mu| \leq 1$.

The proof of this corollary is on the similar lines of the proof of Corollary 2.16 and therefore omitted.

Taking $\mu = 1$ in the above Corollary 2.37, we obtain the following result:

Corollary 2.38. *Let* $|\lambda-1|\leq 1$, $\alpha\lambda/\delta\geq 0$ *and* $\gamma(\lambda-1)\geq 0$. *If* $p(z)\in\mathcal{A}_0$ *satisfies*

$$p^{\lambda}(z)\left(\alpha+\frac{\gamma}{p(z)}+\delta\frac{zp'(z)}{p^2(z)}\right)\prec e^{\lambda z}\left(\alpha+\gamma e^{-z}+\delta z e^{-z}\right),$$

then $p(z)\prec e^{z}$.

Corollary 2.39. *Let* $-1\leq\lambda\leq 3$, $\alpha\lambda/\delta\geq 0$, $\beta(\lambda+1)/\delta\geq 0$ *and* $-1/4<\gamma/\delta\leq 0$. *If* $p(z)\in\mathcal{A}_0$ *satisfies*

$$p^{\lambda}(z)\left(\alpha+\beta p(z)+\frac{\gamma}{p(z)}+\delta\frac{zp'(z)}{p^2(z)}\right)\prec(1+z)^{\lambda/2}\left(\alpha+\beta\sqrt{1+z}+\frac{\gamma}{\sqrt{1+z}}+\frac{\delta z}{2(1+z)^{3/2}}\right),$$

then $p(z)\prec\sqrt{1+z}$.

Proof. The result follows from Theorem 2.24 by taking $q(z)=\sqrt{1+z}$.

Clearly, $Q(z)=\delta(1+z)^{(\lambda-3)/2}z/2$ is starlike (univalent) in $z\in\mathbb{D}$ on the similar lines of the proof of Corollary 2.18. Also, we have

(25)

$$\begin{aligned}&\operatorname{Re}\left(1+\frac{\gamma(\lambda-1)}{\delta}+\frac{\alpha\lambda}{\delta}q(z)+\frac{\beta(\lambda+1)}{\delta}q^2(z)+(\lambda-2)\frac{zq'(z)}{q(z)}+\frac{zq''(z)}{q'(z)}\right)\\&=\operatorname{Re}\left(1+\frac{\gamma(\lambda-1)}{\delta}+\frac{\alpha\lambda}{\delta}\sqrt{1+z}+\frac{\beta(\lambda+1)}{\delta}(1+z)+\frac{\lambda}{4}+(\lambda-3)\frac{z}{2(1+z)}\right)\\&\geq 1+\frac{\gamma(\lambda-1)}{\delta}+\frac{\lambda-3}{4}=\lambda\left(\frac{\gamma}{\delta}+\frac{1}{4}\right)-\frac{\gamma}{\delta}+\frac{1}{4}\geq 0,\end{aligned}$$

by proceeding as in the proof of Corollary 2.18 for the given range of constants in the hypothesis. And hence the result. □

Taking $\beta=0$ in Theorem 2.24, we have the following result:

Corollary 2.40. *Let* $\alpha\delta\geq 0$ *and* $0\leq\lambda\leq 2$. *For* $-1<B<A\leq 1$, *if* $(1+AB)(1-A)(1-B)>8AB$, *let* $(B-A)/((1+A)(1+B))\leq\gamma/\delta\leq 2(B-A)/((1+A)(1+B))$. *Suppose* $p(z)\in\mathcal{A}_0$ *satisfies*

$$p^{\lambda}(z)\left(\alpha+\frac{\gamma}{p(z)}+\delta\frac{zp'(z)}{p^2(z)}\right)\prec\left(\frac{1+Az}{1+Bz}\right)^{\lambda}\left(\alpha+\gamma\left(\frac{1+Bz}{1+Az}\right)+\delta\frac{(A-B)z}{(1+Az)^2}\right),$$

then $p(z)\prec(1+Az)/(1+Bz)$.

Proof. The result is followed by taking $q(z)=(1+Az)/(1+Bz)$ in Theorem 2.24. Then, we have

$$Q(z)=\frac{\delta z(A-B)}{(1+Az)^{2-\lambda}(1-Bz)^{\lambda}},$$

and

$$\frac{zQ'(z)}{Q(z)} = 1 + (\lambda - 2)\frac{Az}{1+Az} - \frac{\lambda Bz}{1+Bz}, \quad 0 \leq \lambda \leq 2, \tag{26}$$

which is clearly bounded below for the mentioned range of λ. Thus, computing equation (26) on the boundary of $\mathbb{D}$, we get $Q(z)$ is starlike (univalent) for the given range of λ. Also, from condition (2) of Theorem 2.24, we have $\mathrm{Re}(H(z))$ as follows:

$$\mathrm{Re}\left(\frac{\gamma(\lambda-1)}{\delta}+\frac{\alpha\lambda}{\delta}\left(\frac{1+Az}{1+Bz}\right)+(\lambda-2)\left(\frac{A-Bz}{(1+Az)(1+Bz)}\right)+\left(\frac{1-Bz}{1+Bz}\right)\right). \tag{27}$$

Now, by using Lemmas 2.19, 2.20 and 2.21, we obtain

$$\begin{aligned}\mathrm{Re}(H(z)) &\geq \frac{\gamma(\lambda-1)}{\delta} + \frac{\alpha\lambda}{\delta}\left(\frac{1-A}{1-B}\right) + \left(\frac{1-|B|}{1+|B|}\right) \\ &\quad + (\lambda-2)\left(\frac{A-B}{(1+A)(1+B)}\right) \quad (\alpha\delta \geq 0) \\ &= \lambda\left(\frac{\gamma}{\delta} + \frac{A-B}{(1+A)(1+B)}\right) + \frac{\alpha\lambda}{\delta}\left(\frac{1-A}{1-B}\right) + \frac{1-|B|}{1+|B|} \\ &\quad - \left(\frac{\gamma}{\delta} + \frac{2(A-B)}{(1+A)(1+B)}\right). \end{aligned} \tag{28}$$

To achieve the desired result it is required to show that equation (28) becomes greater than or equal to 0, which is possible when $(B-A)/((1+A)(1+B)) \leq \gamma/\delta \leq (2(B-A))/((1+A)(1+B))$. This completes the proof. □

3. Sufficient conditions

On substituting $p(z) = zf'(z)/f(z)$ in Corollaries 2.6, 2.13, 2.28 and 2.35, respectively, we have the following example.

Example 3.1. The following are sufficient conditions for $f \in \mathcal{S}^*$.

(i) Let (a) $1 + 2\beta/\delta > 0$, $\gamma\delta \leq 0$, $\alpha\delta \geq 0$ and $0 \leq \lambda \leq 1$, or (b) $\alpha\lambda/\delta \geq 0$, $\beta(\lambda+1)/\delta \geq 0$, $\gamma(\lambda-1)/\delta \geq 0$ and $|\lambda| \leq 1$. If $f \in \mathcal{A}$ satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha + \gamma\frac{f(z)}{zf'(z)} + (\beta-\delta)\frac{zf'(z)}{f(z)} + \delta\left(1 + \frac{zf''(z)}{f'(z)}\right)\right)$$
$$\prec \left(\frac{1+z}{1-z}\right)^{\lambda}\left(\alpha + \beta\left(\frac{1+z}{1-z}\right) + \gamma\left(\frac{1-z}{1+z}\right) + \frac{2\delta z}{1-z^2}\right).$$

(ii) Let (a) $1+2\alpha/\delta > 0$, $\gamma\delta \geq 0$ and $1 \leq \lambda \leq 2$, or (b) $\alpha\lambda/\delta$, $\gamma(\lambda-1)/\delta \geq 0$ and $-1 \leq \lambda \leq 2$. If $f(z) \in \mathcal{A}$ satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha + \gamma\frac{f(z)}{zf'(z)} + \delta\left(\frac{1 + zf''(z)/f'(z)}{zf'(z)/f(z)} - 1\right)\right)$$
$$\prec \left(\frac{1+z}{1-z}\right)^{\lambda}\left(\alpha + \gamma\frac{1-z}{1+z} + \frac{2\delta z}{(1+z)^2}\right).$$

Remark 3.2. Let $\alpha = \gamma = 0$ and $\lambda = 1$ in Example 3.1(i)(a). Then we have the following result of Ravichandran and Kumar [16]:

Corollary 3.3. *Let* $1 + 2\beta/\delta > 0$. *If* $f(z) \in \mathcal{A}$ *and*

$$\frac{zf'(z)}{f(z)}\left((\beta-\delta)\frac{zf'(z)}{f(z)} + \delta\left(1 + \frac{zf''(z)}{f'(z)}\right)\right) \prec \beta\left(\frac{1+z}{1-z}\right)^2 + \frac{2\delta z}{(1-z)^2},$$

then $f(z) \in \mathcal{S}^*$.

Remark 3.4. Let $\lambda = \alpha = \gamma = 0$ and $\delta = 1$ in Example 3.1(ii)(b). Then we have the following result of Obradowič and Tuneski [15]:

Corollary 3.5. *If* $f(z) \in \mathcal{S}$ *and*

$$\frac{1 + zf''(z)/f'(z)}{zf'(z)/f(z)} \prec 1 + \frac{2z}{(1+z)^2},$$

then $f(z) \in \mathcal{S}^*$.

Kaplan [6] introduced the close-to-convex class $\mathcal{CC}$ as follows:

$$\mathcal{CC} = \left\{ f \in \mathcal{A} : \operatorname{Re}\left(\frac{f'(z)}{g'(z)}\right) > 0 \right\},$$

where $g(z)$ is convex (univalent) function in $\mathbb{D}$.

Remark 3.6. Let $p(z) = 2f'(z)/(2+z)$, $\alpha = \gamma = 0$ and $\lambda = 1$ in Corollary 2.6, we get the sufficient condition for f to belong to the class $\mathcal{CC}$, as given below:

Example 3.7. Let $1 + 2\beta/\delta > 0$. If $f \in \mathcal{A}$ satisfies

$$\frac{2f'(z)}{(2+z)^2}\left(2\beta f'(z) - \delta z\right) + \frac{2\delta z f''(z)}{2+z} \prec \beta\left(\frac{1+z}{1-z}\right)^2 + \frac{2\delta z}{(1-z)^2},$$

then $f \in \mathcal{CC}$.

As an application of Corollaries 2.17 and 2.38, we have the following example:

Example 3.8. Let $f \in \mathcal{A}$. Then following are sufficient conditions for $f \in \mathcal{S}_e^*$.

(i) Let $\delta > 0$, $\alpha\lambda \geq -(A+B)$, where $A = \beta(\lambda+1)/e \geq 0$, $B = \gamma(\lambda-1)/e \geq 0$ and $|\lambda| \leq 1$. If f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha + \gamma\frac{f(z)}{zf'(z)} + (\beta-\delta)\frac{zf'(z)}{f(z)} + \delta\left(1 + \frac{zf''(z)}{f'(z)}\right)\right) \prec e^{\lambda z}\left(\gamma e^{-z} + \alpha + \beta e^{z} + \delta z\right).$$

(ii) Let $|\lambda - 1| \leq 1$, $\alpha\lambda/\delta \geq 0$ and $\gamma(\lambda-1) \geq 0$. If f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha + \beta\frac{zf'(z)}{f(z)} + \gamma\frac{f(z)}{zf'(z)} + \delta\left(\frac{1 + zf''(z)/f'(z)}{zf'(z)/f(z)} - 1\right)\right) \prec e^{\lambda z}\left(\alpha + \beta e^{z} + \gamma e^{-z} + \delta z e^{-z}\right).$$

Taking $p(z) = zf'(z)/f(z)$ in Corollaries 2.7 and 2.29, respectively, we have the following example:

Example 3.9. Let $f \in \mathcal{A}$. Then we have the following sufficient conditions for $f \in \mathcal{S}^*(1/2)$.

(i) Let $1 + \beta/\delta > 0$, $\gamma\delta \leq 0$, $-1 + (2\alpha + \beta)/\delta \geq 0$ and $0 \leq \lambda \leq 1$ and f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda} \left(\alpha + \gamma \frac{f(z)}{zf'(z)} + (\beta - \delta)\frac{zf'(z)}{f(z)} + \delta\left(1 + \frac{zf''(z)}{f'(z)}\right)\right)$$
$$\prec \left(\frac{1}{1-z}\right)^{\lambda} \left(\alpha + \beta\frac{1}{1-z} + \gamma(1-z) + \frac{\delta z}{1-z}\right).$$

(ii) Let $1 + \alpha/\delta > 0$, $(2\gamma + \alpha)/\delta \geq 1$ and $1 \leq \lambda \leq 2$ and f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda} \left(\alpha + \gamma \frac{f(z)}{zf'(z)} + \delta\left(\frac{1 + zf''(z)/f'(z)}{zf'(z)/f(z)} - 1\right)\right)$$
$$\prec \left(\frac{1}{1-z}\right)^{\lambda} (\alpha + (1-z)\gamma + \delta z).$$

Remark 3.10. Let $\lambda = \alpha = \gamma = 0$, $\delta = \beta = 1$ and $\lambda = \alpha = \delta = 1$, $\gamma = 0$, respectively in Example 3.9(i) and (ii), we obtain the following known result of Marx [9] and Strohhäcker [23].

Corollary 3.11. *A convex function is starlike of order* $1/2$.

Definition. A function $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$ is said to be ρ-convex in $\mathbb{D}$ if it is analytic and $f(z)f'(z)/z \neq 0$ and also it satisfies

$$\operatorname{Re}\left(\rho\left(1 + \frac{zf''(z)}{f'(z)}\right) + (1-\rho)\frac{zf'(z)}{f(z)}\right) > 0. \tag{29}$$

The set of all such functions is denoted by ρ-CV.

Mocanu [12] gave the following result for ρ-convex functions.

Corollary 3.12. *Let ρ be an arbitrary real number, and suppose that $f(z)$ is ρ-convex. If $\rho \geq 1$, then $f(z)$ is convex.*

Taking $\lambda = \alpha = \gamma = 0$, $\beta = 1$ and $\delta = \rho$ in Example 3.9(i), we get the following result:

Corollary 3.13. *Let $1/\rho \geq 1$ and if f satisfies*

$$(1-\rho)\frac{zf'(z)}{f(z)} + \rho\left(1 + \frac{zf''(z)}{f'(z)}\right) \prec \frac{1+\rho z}{1-z}, \tag{30}$$

then $\dfrac{zf'(z)}{f(z)} \prec \dfrac{1}{1-z}$.

The above Definition, Corollaries 3.11, 3.12 and 3.13 yield the following result:

Corollary 3.14. *A ρ-convex function is starlike of order $1/2$ for $\rho > 0$.*

As an application of Corollaries 2.18 and 2.39, respectively, we have the following example.

Example 3.15. Let $f \in \mathcal{A}$. Then following are sufficient conditions for $f \in \mathcal{S}_L^*$.

(i) Let $-2 \le \lambda \le 2$, $\gamma(\lambda-1) \ge 0$, $\beta(\lambda+1) \ge 0$, $\delta > \max(0, \sqrt{2}\gamma)$ and $-(2\sqrt{2}\gamma+\delta)/4 < \alpha \le -3\gamma/(2\sqrt{2})$ and f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha+\gamma\frac{f(z)}{zf'(z)}+(\beta-\delta)\frac{zf'(z)}{f(z)}+\delta\left(1+\frac{zf''(z)}{f'(z)}\right)\right)$$
$$\prec (1+z)^{\lambda/2}\left(\frac{\gamma}{\sqrt{1+z}}+\alpha+\beta\sqrt{1+z}+\delta\left(\frac{z}{2(1+z)}\right)\right).$$

(ii) Let $-1 \le \lambda \le 3$, $\alpha\lambda/\delta \ge 0$, $\beta(\lambda+1)/\delta \ge 0$ and $-1/4 < \gamma/\delta \le 0$ and f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha+\beta\frac{zf'(z)}{f(z)}+\gamma\frac{f(z)}{zf'(z)}+\delta\left(\frac{1+zf''(z)/f'(z)}{zf'(z)/f(z)}-1\right)\right)$$
$$\prec (1+z)^{\lambda/2}\left(\alpha+\beta\sqrt{1+z}+\frac{\gamma}{\sqrt{1+z}}+\frac{\delta z}{2(1+z)^{3/2}}\right).$$

Let $p(z) = zf'(z)/f(z)$ in Corollaries 2.22 and 2.40. Then we have the following example.

Example 3.16. Let $f \in \mathcal{A}$. Then following are sufficient conditions for $f \in \mathcal{S}^*[A,B]$, $-1 < B < A \le 1$.

(i) Let $0 \ne f \in \mathcal{S}$. Let $0 \le \lambda \le 1$. If $(1+AB)(1-A)(1-B) > 8AB$, let $\alpha/\delta \ge (B-A)/((1+A)(1+B))$, $\beta/\delta \ge (A-B)(1-B)/((1-A^2)(1+B))$. If f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha+\gamma\frac{f(z)}{zf'(z)}+(\beta-\delta)\frac{zf'(z)}{f(z)}+\delta\left(1+\frac{zf''(z)}{f'(z)}\right)\right)$$
$$\prec \left(\frac{1+Az}{1+Bz}\right)^{\lambda}\left(\alpha+(\beta-\delta)\left(\frac{1+Az}{1+Bz}\right)+\delta\frac{(A-B)z}{(1+Az)(1+Bz)}\right).$$

(ii) Let $\alpha\delta \ge 0$, and $0 \le \lambda \le 2$. If $(1+AB)(1-A)(1-B) > 8AB$, let $(B-A)/((1+A)(1+B)) \le \gamma/\delta \le 2(B-A)/((1+A)(1+B))$. If f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)^{\lambda}\left(\alpha+\gamma\frac{f(z)}{zf'(z)}+\delta\left(\frac{1+zf''(z)/f'(z)}{zf'(z)/f(z)}-1\right)\right)$$
$$\prec \left(\frac{1+Az}{1+Bz}\right)^{\lambda}\left(\alpha+\gamma\left(\frac{1+Bz}{1+Az}\right)+\delta\left(\frac{(A-B)z}{(1+Az)^2}\right)\right).$$

From Corollary 2.23, we get the following example by taking

$$p(z) = zf'(z)/f(z).$$

Example 3.17. Let $f \in \mathcal{A}$. Then following are sufficient conditions for $f \in \mathcal{S}_P$.

(i) Let $\beta/\delta \geq \max(0; -\alpha/\delta)$. If f satisfies

$$\left(\frac{zf'(z)}{f(z)}\right)\left(\alpha + (\beta-\delta)\frac{zf'(z)}{f(z)} + \delta\left(1 + \frac{zf''(z)}{f'(z)}\right)\right)$$
$$\prec \left(1 + \frac{2}{\pi^2}\left(\log\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)^2\right)\left(\alpha + \beta\left(1 + \frac{2}{\pi^2}\left(\log\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)^2\right)\right.$$
$$\left. + \frac{4\delta}{\pi^2}\frac{\sqrt{z}}{1-z}\log\left(\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)\right).$$

Remark 3.18. As an application of the above results and by Noshiro-Warschawski, an analytic function f is univalent if we substitute $p(z) = f'(z)$ in Corollaries 2.6, 2.28, 2.7, 2.29, 2.18, 2.39, 2.23, 2.17, 2.38, 2.13 and 2.35, respectively.

References

[1] D. A. Brannan and W. E. Kirwan, *On some classes of bounded univalent functions*, J. Lond. Math. Soc. (2) **1** (1969), 431–443. https://doi.org/10.1112/jlms/s2-1.1.431

[2] N. E. Cho and J. A. Kim, *Angular estimations of certain analytic functions*, J. Korean Math. Soc. **34** (1997), no. 2, 427–436.

[3] ———, *On a sufficient condition and an angular estimation for Φ-like functions*, Taiwanese J. Math. **2** (1998), no. 4, 397–403. https://doi.org/10.11650/twjm/1500407012

[4] N. E. Cho, H. J. Lee, J. H. Park, and R. Srivastava, *Some applications of the first-order differential subordinations*, Filomat **30** (2016), no. 6, 1465–1474. https://doi.org/10.2298/FIL1606465C

[5] W. Janowski, *Extremal problems for a family of functions with positive real part and for some related families*, Ann. Polon. Math. **23** (1970/1971), 159–177. https://doi.org/10.4064/ap-23-2-159-177

[6] W. Kaplan, *Close-to-convex schlicht functions*, Michigan Math. J. **1** (1952), 169–185 (1953). http://projecteuclid.org/euclid.mmj/1028988895

[7] J.-L. Liu and R. Srivastava, *The order of starlikeness of some classes of strongly starlike functions*, Quaest. Math. **41** (2018), no. 5, 707–718. https://doi.org/10.2989/16073606.2017.1398194

[8] W. C. Ma and D. Minda, *A unified treatment of some special classes of univalent functions*, in Proceedings of the Conference on Complex Analysis (Tianjin, 1992), 157–169, Conf. Proc. Lecture Notes Anal., I, Int. Press, Cambridge, MA, 1994.

[9] A. Marx, *Untersuchungen über schlichte Abbildungen*, Math. Ann. **107** (1933), no. 1, 40–67.

[10] R. Mendiratta, S. Nagpal, and V. Ravichandran, *On a subclass of strongly starlike functions associated with exponential function*, Bull. Malays. Math. Sci. Soc. **38** (2015), no. 1, 365–386. https://doi.org/10.1007/s40840-014-0026-8

[11] S. S. Miller and P. T. Mocanu, *Differential subordinations*, Monographs and Textbooks in Pure and Applied Mathematics, **225**, Marcel Dekker, Inc., New York, 2000.

[12] P. T. Mocanu, *Une propriété de convexité généralisée dans la théorie de la représentation conforme*, Mathematica (Cluj) **11 (34)** (1969), 127–133.

[13] M. Nunokawa, *On the order of strongly starlikeness of strongly convex functions*, Proc. Japan Acad. Ser. A Math. Sci. **69** (1993), no. 7, 234–237. http://projecteuclid.org/euclid.pja/1195511343

[14] M. Nunokawa, S. Owa, N. Takahashi, and H. Saitoh, *Sufficient conditions for Carathéodory functions*, Indian J. Pure Appl. Math. **33** (2002), no. 9, 1385–1390.
[15] M. Obradowič and N. Tuneski, *On the starlike criteria defined by Silverman*, Zeszyty Nauk. Politech. Rzeszowskiej Mat. No. 24 (2000), 59–64 (2001).
[16] V. Ravichandran and S. S. Kumar, *On sufficient conditions for starlikeness*, Southeast Asian Bull. Math. **29** (2005), no. 4, 773–783.
[17] M. S. Robertson, *Certain classes of starlike functions*, Michigan Math. J. **32** (1985), no. 2, 135–140. https://doi.org/10.1307/mmj/1029003181
[18] F. Rønning, *Uniformly convex functions and a corresponding class of starlike functions*, Proc. Amer. Math. Soc. **118** (1993), no. 1, 189–196. https://doi.org/10.2307/2160026
[19] K. Sharma and V. Ravichandran, *Applications of subordination theory to starlike functions*, Bull. Iranian Math. Soc. **42** (2016), no. 3, 761–777.
[20] J. Sokół and J. Stankiewicz, *Radius of convexity of some subclasses of strongly starlike functions*, Zeszyty Nauk. Politech. Rzeszowskiej Mat. No. 19 (1996), 101–105.
[21] H. M. Srivastava and S. S. Eker, *Some applications of a subordination theorem for a class of analytic functions*, Appl. Math. Lett. **21** (2008), no. 4, 394–399. https://doi.org/10.1016/j.aml.2007.02.032
[22] J. Stankiewicz, *Quelques problèmes extrémaux dans les classes des fonctions α-angulairement étoilées*, Ann. Univ. Mariae Curie-Skłodowska Sect. A **20** (1966), 59–75 (1971).
[23] E. Strohhäcker, *Beiträge zur Theorie der schlichten Funktionen*, Math. Z. **37** (1933), no. 1, 356–380.
[24] B. A. Uralegaddi, M. D. Ganigi, and S. M. Sarangi, *Univalent functions with positive coefficients*, Tamkang J. Math. **25** (1994), no. 3, 225–230.

Shagun Banga
Department of Applied Mathematics
Delhi Technological University
Delhi-110042, India
Email address: shagun05banga@gmail.com

S. Sivaprasad Kumar
Department of Applied Mathematics
Delhi Technological University
Delhi-110042, India
Email address: spkumar@dce.ac.in

ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse20922020.pdf
https://doi.org/10.30534/ijatcse/2020/20922020

Autism Detection using r-fMRI: Subspace Approximation and CNN Based Approach

Daya Gupta[1*], Ishani Vij[2*], Mansi Gupta[3*]
[1]Delhi Technological University, India, dgupta@dce.ac.in
[2]Delhi Technological University, India, ishanivij11@gmail.com
[3]Delhi Technological University, India, mansigupta_bt2k16@dtu.ac.in

ABSTRACT

Autism spectrum disorder (ASD) is a heterogeneous disorder that causes impaired social interactions and altered behavioral patterns. It is currently diagnosed by assessing the developmental screenings, followed by a diagnostic evaluation of an individual using standard screening and diagnostic tools. However, a more concrete diagnostic method is required to get hold of the underlying cause of the disorder to ensure better treatment and prevention of the disorder. Recent research on ASD has shown resting state functional magnetic resonance imaging (r-fMRI) as a useful tool for classification of ASD and neurotypical subjects. However, due to the large dimensionality of fMRI scans, directly applying classification methods to the data results in high computational cost. In order to avoid the curse of dimensionality , we propose a combined multistage SAAK transform and CNN based approach which selects the most relevant and discriminative features without losing any information and uses them for a more accurate and less computationally expensive classification of ASD subjects from typically developing controls. A classification accuracy of 74.55% is achieved using the proposed method. We show that the performance of the proposed approach of classification using ABIDE dataset is comparable to that of standalone CNN while being less computationally intensive.

Key words : ABIDE, Autism Spectrum Disorder, Convolutional Neural Network (CNN), Karhunen- Loeve Transform (KLT), subspace approximation

1. INTRODUCTION

Autism Spectrum Disorder is a neurological and developmental disorder that affects the overall cognitive, social, emotional and physical health of an individual. It is characterised by a broad range of conditions, due to which the disease is referred to as a spectrum. These impact the nervous system and cause uneasy social interactions, difficulty in communication, restricted and repetitive behaviours and frequent sensory overloads. It is found to occur in more than 1% of the global population. Currently there is no single notion regarding the underlying causes of the disorder. Medical professionals diagnose ASD by analysing the behaviour and development of the child using a range of screening and diagnostic tools like Autism Diagnosis Interview- Revised [1], Childhood Autism Rating Scale (CARS) [2]. The reliability of these diagnostic methods remains questionable as they are prone to misdiagnosis [3]. Moreover, these methods give no information about the biological aspects of the disorder. To get a clearer idea about the underlying brain activity of autistic individuals, an inclination towards exploring the brain variance between healthy and diseased populations based on neuroimaging data has been observed.

However, given the high dimensionality of neuroimaging data, directly applying machine learning models for classification turns out to be computationally expensive. C. Jay Kuo and Yueru Chen proposed a multi-stage Subspace approximation with augmented kernels (SAAK) transform method for feature extraction. In this work, we borrow the SAAK transform method from [4] and propose a combined SAAK transform and CNN based approach for classifying the subjects as autistic and neurotypical (healthy) individuals. This approach reduces the computational cost without compromising on the accuracy of classification. We consider the functional magnetic resonance imaging (r-fMRI) scans of the subjects and the associated Blood Oxygen Level Dependent (BOLD) time series signals to build connectivity matrices, which act as input to the model. Each stage of SAAK transform reduces the dimensionality of the feature space. It also nullifies the rectification loss due to the ReLU layer by augmenting a negative kernel to all the kernels obtained. Here we point out a major difference between the ReLU layer in CNN and that in SAAK transform.

*The authors contributed equally to this work.

Traditionally CNNs have given state-of-the-art results in image analysis problems. Multiple convolution and pooling layers enable repeated extraction and processing of features till one gets the final features, which exhibit high robustness and do not show changes with transformations in the input. The feature extraction step imposes constraints when dealing with a large sample size and high dimensionality data such as ABIDE dataset. SAAK transform acts as a computationally less expensive alternative for extracting relevant high level features from the data.

The rest of this paper is organised as follows: In section 2 we discuss the methods implemented so far in this field, followed by a thorough explanation of the proposed approach in Section 3. Section 4 and 5 contain the experimental details and a brief description of the experiments conducted in this study. Section 6 and 7 present the classification results, followed by a comprehensive discussion. Concluding remarks are presented in section 8.

2. LITERATURE SURVEY

Ample of studies have used neuroimaging data obtained using magnetic resonance imaging (MRI) for the classification of ASD. With the advent of MRI as a medical imaging technique, researchers have been able to gain deeper knowledge about the pathology of the disorder [5] [6] [7]. It generates detailed images of the internal organs and tissues, along with their working using strong magnetic fields and radio waves. There are two variants of MRI: structural MRI and functional MRI. Structural MRI is a technique for examining the anatomy and pathology of the brain. The studies for ASD detection conducted using structural MRI [8] [9] [10] used features like volume, surface and thickness as parameters for classification. Apart from structural alterations, autistic and normal individuals also exhibit differences in underlying brain connectivity [11] [12] [13]. The second technique of MRI, called functional MRI, is used to cap these differences. In functional MRI, the level of interaction between the different components of the brain, which may or may not be physically connected are measured. The brain activity is examined by detecting changes in the Blood Oxygen Level Dependent (BOLD) signals. It relies on the fact that cerebral blood flow and neuronal activities are coupled i.e. the blood flow to a brain region increases when that part of the brain is in use [14].

Various studies have used pairwise correlations between different regions of interest (ROIs) as features to define connectivity matrices for the subjects [13] [15] and then implemented supervised learning methods like SVM [16] and leave-one-out classifier [17], as well as CNN [18] for classification. Subbaraju et al. (2017) achieved an accuracy of 77.3% using a combined spatial filter and SVM classifier. Their method selects the most discriminative features using a spatial filter by orthogonally projecting the connectivity matrices of BOLD signals of ASD and control individuals, and feeds them to an SVM classifier. Another work by Sarah et al. (2019) used a graph based classification where they viewed the brain structure as a graph, with the ROIs as nodes and BOLD signals as edges of the graph, and used graph signal processing to obtain features related to the frequency content of the signals. A decision tree trained using these features for 452 subjects gave a classification accuracy of 73.5%. Studies that used fMRI data collected from a single site have achieved accuracy as high as 97% (Just et al., 2014). Plitt et al. (2015) compared the classifications of various machine learning algorithms on the scans of 178 ASD and IQ matched neurotypical males, and achieved a peak accuracy of 76.67% with ridge regression method. Relatively better classification is performed when sample size is small, typically constrained to around 100 participants (Arbabshirani et al., 2016), while it experiences a considerable drop when the sample size is large or in case of multi- site data. (Nielson et al., 2013).

In this study, we use data collected at multiple sites and from people of varied demographic conditions for a more robust study.

3. PROPOSED APPROACH

Here we present our strategy to adopt a combined subspace approximation with augmented kernel (SAAK) transform and CNN based approach for classification. Our data consists of 4-dimensional (time- series) fMRI scans of the subjects. Different parcellations of the brain, called regions of interest (ROI), are defined by various atlases. These atlases were used to obtain association matrices for the subjects [19]. We start by building a brain connectivity matrix for each subject. A connectivity matrix helps to quantitatively visualise the strength of functional connections. It reflects the pairwise functional connectivity between voxels of different regions of the brain and is created using the averaged Blood Oxygen Level Dependent (BOLD) time-series signals. These signals are extracted from the regions of interest defined by an atlas. The connectivity between ROIs is estimated using correlation between the time- series signals. We used three different methods for calculating the entries of the matrices: partial correlation, Pearson correlation and tangent space embedding [20]. An illustration of this approach is shown in Figure 1.

3.1 Subspace Approximation

The connectivity matrices obtained so far are of size $N \times N$, where N is the number of regions the brain is divided into, giving us N^2 features corresponding to each data item. We

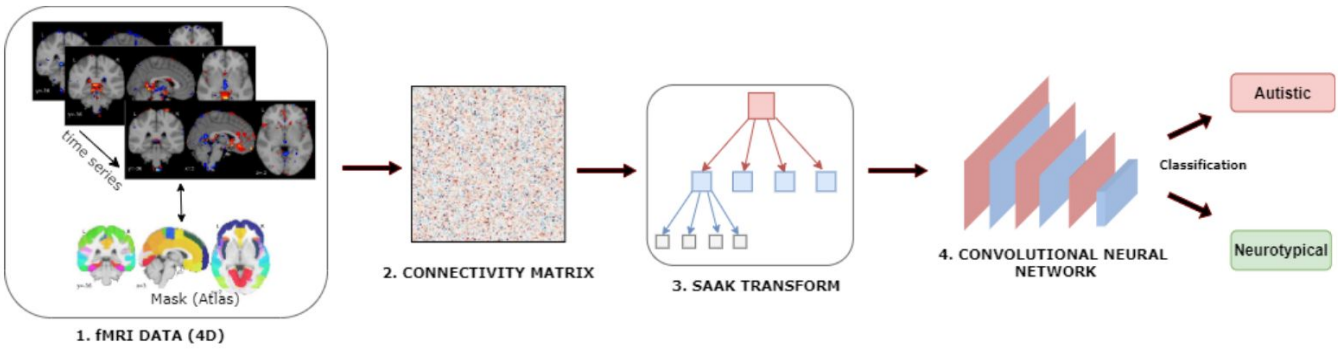


Figure 1: Illustration of the proposed approach

seek to reduce the feature space while retaining as much information as possible. This is done by transforming the data to a linear subspace with an orthogonal basis using KLT in multiple stages.

First the connectivity matrices are scaled to have dimensions that are the n^{th} power of 2, for some $n \in I^+$. These scaled connectivity matrices act as the input to the multistage SAAK transform. We then obtain the KLT basis functions, b_k, for the input vectors $f \in R_n$. The requirement for multiple stages arises as it is practically infeasible to conduct KLT on the entire input matrix due to its large size , as explained below.

3.2 Multiple Stages

For the determination of the KL transform of a feature set, it is first required to calculate the covariance matrix. Let $X=[x_1 x_2 \ldots x_N]^T$ denote the initial feature matrix of a subject. The covariance matrix of X is computed as:

$$C_X = \frac{1}{N}\sum_{i=1}^{N}(x_i x_i^{'}) \tag{1}$$

Now, the eigenvectors, V, of the covariance matrix are arranged into rows to form the orthogonal transformation matrix, T.

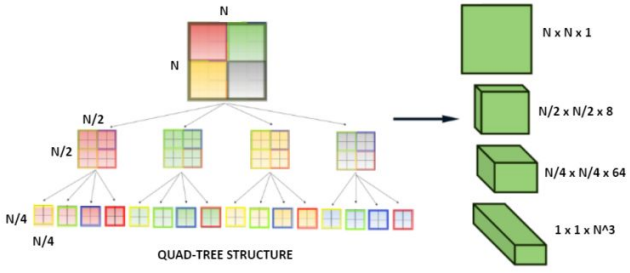


Figure 2: Multistage SAAK Transform quad-tree structure

Thus, the transformation matrices in KLT are dependent on the feature set, more precisely on its covariance matrix. For a feature set with dimension $N \times N$, the matrix will be of the order N^4. In case of large values of N, the transformation will require non- trivial computations, suggesting the infeasibility of application. Consequently, large sized connectivity matrices are transformed into a smaller subspace by recursively decomposing them into smaller parts. The application of KLT at each level brings down the dimensions of the feature set by a factor of two.

Hence, we decompose the matrices into four smaller non-overlapping parts in each stage, eventually forming a quad-tree structure. Each node of this tree is half the size of its parent node, the leaf nodes being of the size 2×2. KLT is applied at each level of the tree to obtain the basis functions.. All the stages of SAAK transform are cascaded in a bottom up manner. A sign-confusion problem occurs due to the recursive application of KLT [18]. Therefore, a ReLU activation function is introduced between two successive transforms. Figure 2 shows the formation of quad tree structure in multistage saak transform, useful for transforming images with dimensions $(N \times N \times 1)$ into dimensions $(1 \times M \times M)$, where M denotes N^3.

3.3 Kernel Augmentation

The addition of the ReLU operation between the KLT transforms is accompanied by an increased rectification loss. We attempt to bring down the rectification loss by using kernel augmentation [4]. We generate $(N-1)$ pairs of AC kernels, in addition to a DC kernel, using the KLT basis functions obtained above. We retain the projections on the DC kernels and pass the AC projections through the ReLU activation function. The DC vector is represented as

$$a_0 = \frac{1}{N}(1, 1, \ldots, 1)^T \tag{2}$$

Each AC kernel pair consists of a positive kernel, a_{2k-1}, augmented with its negative vector, a_{2k}. These are generated using the basis functions, b_k, according to as:

$$a_{2k-1} = b_k, \quad a_{2k} = -b_k, \quad k=1,2,\ldots,N-1 \tag{3}$$

The input vectors, *f*, are projected onto these kernels to yield a projection vector

$$p = (p_0, p_1, \ldots, p_{(2N-1)})^T \quad \in R^{2N-1} \tag{4}$$

where $p_k = a_k^T f$ represents the projection of *f* on kernel a_k. The above process representing a single stage SAAK transform is repeated till signed KLT coefficients of dimensions $1 \times 1 \times 2^{3t}$ are obtained, where $2^t \times 2^t$ is the initial size of the connectivity matrix. All the projection vectors except the DC projection p_0 are passed through the ReLU function and we obtain the SAAK coefficients, *o,* as:

$$o = (o_0, o_1, \ldots, o_{(2N-1)})^T \quad \in R^{2N-1} \tag{5}$$

where, $o_0 = p_0$ (DC projection retained) and for *k = 1*, ..., *N*:

$$o_{2k-1} = p_{2k-1} \quad and \quad o_{2k} = 0 \quad , if \ p_{2k-1} > 0 \tag{6}$$

$$o_{2k-1} = 0 \quad and \quad o_{2k} = p_{2k} \quad , if \ p_{2k} > 0 \tag{7}$$

Unlike CNN where the activation of input signals introduces non- linearity in the output, we preserve all the convolution input by augmenting the kernels with their negatives. This way, there will be a positive/negative correlation pair of each input vector with kernel pair (a_{2k-1}, a_{2k}). ReLU will allow one correlation from the pair to pass through it and will block the other one.

3.4 Feature Selection

Finally, we calculate and compare the F- score of the coefficients to select the most important features. The discriminative power of a feature is directly proportional to its F-test score, hence features with highest F-test scores are chosen to form the final feature vectors. F- score is computed as:

$$F = \frac{larger\ sample\ variance}{smaller\ sample\ variance} \tag{8}$$

The larger sample variance (LSV) and smaller sample variance (SSV) can be written respectively as:

$$LSV = \frac{\sum_{i=1}^{K} n_i (\overline{X}_i - \overline{X})^2}{(K-1)} \tag{9}$$

where, $\overline{X}$ is the mean of the entire data, $\overline{X}_i$ is the mean of the i^{th} classification group, n_i represents the number of observations in the i^{th} group, and *K* is the total number of groups.

$$SSV = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2}{(N-K)} \tag{10}$$

where, X_{ij} is the j^{th} observation in the i^{th} group and *N* is the total size of the sample.

In this study, for the CC200 atlas, the initial size of the connectivity matrix is 200 x 200. The number of features was reduced to 7294 after the multistage SAAK transform. After calculating the F-score for all the features, we chose 1000 top features with the highest F- scores.

3.5 CNN

Our CNN architecture, represented in Figure 3, consists of three convolution layers interspersed with pooling layers followed by a fully connected layer which computes disease probability and is subsequently used for classification. The model has a batch size of 32 and was trained for 20 epochs. To avoid overfitting the data, the learning rate of CNN was set to 0.001 and momentum to 0.9.

Note that Figure 3 is only for architectural representational purposes. The actual CNN used is one dimensional in nature and takes as input the output of SAAK transform.

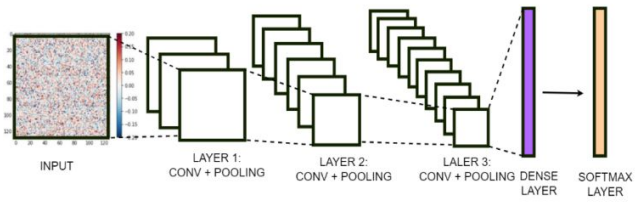


Figure 3: Convolutional neural network architecture

4. MATERIALS AND METHODS

4.1 Data and participants

The data used for this study was downloaded from the Autism Brain Imaging Data Exchange (ABIDE) project (http://fcon_1000.projects.nitrc.org/indi/abide/). The data comprises of rs-fMRI scans of subjects, along with phenotypic information such as subject identity number, age, sex, assessment scores (IQ test, social interaction, communication etc.) for each subject.

Due to the difference in conditions (eye status during scan i.e. opened/ closed, taken stimulants 24 hours prior to scan, etc) while sampling the data, and the variety in demographic conditions of the subjects (age, sex, handedness, etc) the dataset shows a significant extent of heterogeneity. Hence, making use of the phenotypic information available to us, we

define the participants of this study as the subjects who met all the following criteria:

- Eye status during rest scan- open
- Quality control (manual assessment)

Table 1 shows the demographic details of the 609 participants.

Table 1: Participant classification

Gender	Individual Orientation	
	ASD	***Neurotypival***
Male	252	252
Female	36	69
Age (<=20)	242	270
Age (>20)	46	51

4.2 Preprocessing

The ABIDE consortium shared a preprocessed version of the data as a part of the Preprocessed Connectomes Project (PCP). Out of the four available pipelines, we chose the data which was pre-processed according to the C-PAC (Configurable Pipeline for the Analysis of Connectomes) pipeline. The data obtained was corrected by reslicing for head-motion, intensity normalisation (with global mean = 1000), motion realignment, and nuisance signal removal (due to physiological processes like heartbeat and respiration). The data were transformed from the individual native space to the Montreal Neurological Institute (MNI) space. Global signal regression, however, is not used in this study. Steps involved in data processing are shown in Figure 4.

Figure 4: Data preprocessing workflow

5. EXPERIMENTS

The following experiments were carried out in this study:

1. Classification Model: Experiments were conducted to determine the best classifier for classification of data before applying SAAK transform as well as that for linear SAAK transform output. SAAK transform output was classified using linear support vector classifier (SVC), Dense Net and CNN.
2. Atlas for time-series extraction: Atlas for time-series extraction: Out of the seven atlases used by PCP for ROI time-series extraction, we consider the following five atlases for the input data and compare the classification performances: Harvard-Oxford (HO), Craddock 200 (CC200), Eickhoff-Zilles (EZ), Talaraich and Tournoux (TT) and Dosenbach 160 (DOZ160).
3. Type of connectivity matrix: Connectivity matrix for a subject is estimated based on its time-series data. Three kinds of connectivity measures: partial correlation, correlation and tangent space embedding were used to construct the connectivity matrix and the classification ability of the model corresponding to each of these was analysed .
4. Time demand of CNN vs proposed approach: Computational efficiency of CNN and proposed combined SAAK and CNN approach was evaluated and compared.

6. RESULTS

6.1 Performance comparison of proposed approach and other approaches

Performance of the proposed approach was compared with various machine learning and deep learning classification models present in the literature. Table 2 shows the cross-validation accuracy for all the atlases used in the PCP for ROI time series extraction. These scores are corresponding to the partial correlation connectivity matrices. Five classification approaches viz. Decision tree, SVC, dense net, CNN and SAAK + CNN were evaluated. CNN achieved classification accuracy of 75.23 percent, marginally higher than accuracy of 74.55 percent obtained using the proposed approach.

Table 2: Performance of classifiers with different atlases

Brain Atlas	Classifier Accuracy (%)				
	Decision Tree	***SVC***	***Dense Net***	***CNN***	***SAAK + CNN***
HO	62.12	64.35	67.71	71.92	71.26
CC200	64.55	65.78	69.20	75.23	**74.55**
EZ	64.98	65.21	64.88	72.67	71.98
TT	60.23	60.47	62.55	69.04	68.39
DOZ160	62.63	59.71	68.23	70.95	70.31

Table 3: Performance of different correlation measures

Correlation Measure	Classifier Accuracy (%)		
	SAAK+SVC	***SAAK + Dense Net***	***SAAK+CNN***
Partial	68.84	71.62	**74.55**
Pearson	66.29	69.13	70.61
Tangent	68.51	70.02	74.46

Table 3 presents classification accuracy corresponding to the different correlation measures used for constructing the correlation matrix, using the CC200 atlas.

6.2. Computational Efficiency

All experiments were conducted in python. on a system with windows 10 operating system, core i7 processor and 8 GB RAM. Execution time of both CNN and proposed approach are as follows:

- CNN : 7 minutes 23 sec
- SAAK + CNN : 3 min 9 sec

6.3 Filtering of Signals

Band pass filtered and non band pass filtered signals were evaluated and the observed accuracy of signals after band pass filtering was found to be marginally better.

- Band pass filtering (0.01 - 0.073 Hz): 74.55% accuracy
- Non band pass filtered : 74.39% accuracy

7. DISCUSSIONS

We carried out a detailed experimental evaluation of our proposed approach on the ABIDE dataset and observed that classification using the proposed method gave results comparable to other state-of-the-art methods for large data sets while being more computationally effective.

7.1 Impact of connectivity parameterisation

We computed the connectivity matrices of subjects to use them as input to the model. The connectivity matrices were established using three techniques - Pearson correlation, partial correlation and tangent space embedding (which is a combination of both), and the performance of the model in each case was studied. The best accuracy was obtained using a partial correlation based network. Only a marginal difference in accuracy was seen between the partial correlation and tangent space methods. The tangent space method has a strong mathematical foundation and is being explored in various studies. This result can be backed up by the findings of previous studies on correlation methods which suggest the Pearson method only reflects the marginal association between network nodes and does not capture the true or direct functional connection between them. For example, a large correlation between a pair of nodes can appear due to their common connections to a third-party node, even if the two nodes are not directly connected. [21], [22]. Figure 5 shows the connectivity matrices for the three types of connectivity measures: pearson, partial and tangent.

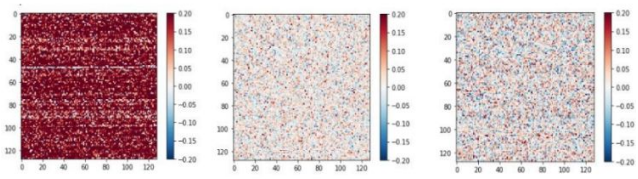


Figure 5: Pearson, partial and tangent correlation matrices

7.2. Impact of parcellation scheme on prediction accuracy

BOLD signals inside the brain are averaged at parcel level. Therefore, both coordinates of the brain regions defined and the size of regions have an effect on the signal content. Hence, we analyse the impact of parcellation scheme on the accuracy of prediction by considering five atlases.
We evaluated the atlases against the proposed method as well as some other state-of-the-art methods on the same dataset. We found out that the SAAK transform classification gave better results than all other methods tested. Further, ROIs defined using the functional CC200 atlas outperformed all other atlases shown in the table. This shows that classification performance is indeed influenced by the atlas selected.

7.3 Impact of frequency of signals

The extracted time- series signals were band-pass filtered to obtain signals in the frequency band 0.01Hz - 0.073Hz (Slow 4 and Slow 5 bands). We implemented our approach with both filtered and non- filtered signals. The filtered signals gave a slightly better discrimination performance over the non- filtered ones. However, we could not establish a clear preference between the two.

7.4 Computational efficiency: SAAK vs CNN

Convolutional neural networks (CNN) and SAAK transform both employ the ReLU activation function. However, there is a fundamental difference in the way these techniques determine their filter weights. In CNN, a supervised learning approach is used to determine the weights. The mapping between input and their corresponding output labels is used to update the weights iteratively. Usually, back-propagation of the loss helps update weights, thereby reducing the classification error. On the other hand, multi-stage SAAK transform uses the second-order statistics of the input vector to select the filter weights. While CNNs usually require a large number of iterations, both input vector, and associated labels to learn the weights, SAAK transform is a feed-forward single-pass process. Using just the input vector, SAAK transform effectively deduces the filter weights in a single pass. This makes SAAK transform a simple and a faster alternative to CNN.

8. CONCLUSION

As of present, no scientific method is used in industry for diagnosis and subsequent management of ASD. Recent years have shown a rise in ASD research [23], [24] but this has been limited by the high dimensionality of images. In this paper, we proposed a SAAK transform based classifier to identify ASD affected individuals from healthy individuals. We analysed the functional connectomes using different atlases, and selected the best parcellation scheme. We also showed that our proposed approach achieved classification accuracy comparable to the existing machine- learning state of the art approach, while being significantly more efficient computationally.

REFERENCES

1. O. Tadevosyan-Leyfer, M. Dowd, R. Mankoski, B. Winklosky, S. Putnam, L. McGrath, H. Tager-Flusberg, and S. E. Folstein, "**A principal components analysis of the autism diagnostic interview-revised**," Journal of the American Academy of Child & Adolescent Psychiatry, vol. 42, no. 7, pp. 864–872, 2003.
2. M. E. Van Bourgondien, L. M. Marcus, and E. Schopler, "**Comparison of dsm-iii-r and childhood autism rating scale diagnoses of autism**," Journal of Autism and Developmental Disorders, vol. 22, no. 4, pp. 493–506, 1992. https://doi.org/10.1007/BF01046324
3. T. Charman and K. Gotham, "**Measurement issues: Screening and diagnostic instruments for autism spectrum disorders–lessons from research and practise**," Child and adolescent mental health, vol. 18, no. 1, pp. 52– 63, 2013.
4. C.-C. J. Kuo and Y. Chen, "**On data-driven saak transform**," Journal of Visual Communication and Image Representation, vol. 50, pp. 237–246, 2018. https://doi.org/10.1016/j.jvcir.2017.11.023
5. B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "**Functional connectivity in the motor cortex of resting human brain using echoplanar mri**," Magnetic resonance in medicine, vol. 34, no. 4, pp. 537– 541, 1995. https://doi.org/10.1002/mrm.1910340409
6. T. P. Gabrielsen, J. S. Anderson, K. G. Stephenson, J. Beck, J. B. King, R. Kellems, D. N. Top, N. C. Russell, E. Anderberg, R. A. Lundwall, et al., "**Functional mri connectivity of children with autism and low verbal and cognitive performance**," Molecular autism, vol. 9, no. 1, p. 67, 2018. https://doi.org/10.1186/s13229-018-0248-y
7. D. P. Kennedy, E. Redcay, and E. Courchesne, "**Failing to deactivate: resting functional abnormalities in autism**," Proceedings of the National Academy of Sciences, vol. 103, no. 21, pp. 8275–8280, 2006.
8. Y. Jiao, R. Chen, X. Ke, K. Chu, Z. Lu, and E. H. Herskovits, "**Predictive models of autism spectrum disorder based on brain regional cortical thickness**," Neuroimage, vol. 50, no. 2, pp. 589–599, 2010. https://doi.org/10.1016/j.neuroimage.2009.12.047
9. R.Abitha and Dr S. Mary Vennila, " **CBARG-Cultural Based Optimized Association Rule Generation Method to Improve Knowledge Discovery in Autism Spectrum Disorder**," International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, no. 6, pp. 3327-22 https://doi.org/10.30534/ijatcse/2019/104862019
10. C. Ecker, V. Rocha-Rego, P. Johnston, J. Mourao-Miranda, A. Marquand, E. M. Daly, M. J. Brammer, C. Murphy, D. G. Murphy, M. A. Consortium, et al., "**Investigating the predictive value of whole-brain structural mr scans in autism: a pattern classification approach**," Neuroimage, vol. 49, no. 1, pp. 44–56, 2010.
11. D. P. Kennedy and E. Courchesne, "**The intrinsic functional organization of the brain is altered in autism**," Neuroimage, vol. 39, no. 4, pp. 1877– 1885, 2008. https://doi.org/10.1016/j.neuroimage.2007.10.052
12. V. L. Cherkassky, R. K. Kana, T. A. Keller, and M. A. Just, "**Functional connectivity in a baseline resting-state network in autism**," Neuroreport, vol. 17, no. 16, pp. 1687–1690, 2006.
13. C. Ecker, A. Marquand, J. Mourao-Miranda, P. Johnston, E. M. Daly, ˜ M. J. Brammer, S. Maltezos, C. M. Murphy, D. Robertson, S. C. Williams, et al., "**Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach**," Journal of Neuroscience, vol. 30, no. 32, pp. 10612–10623, 2010.
14. O. B. Paulson, S. G. Hasselbalch, E. Rostrup, G. M. Knudsen, and D. Pelligrino, "**Cerebral blood flow response to functional activation**," Journal of Cerebral Blood Flow & Metabolism, vol. 30, no. 1, pp. 2–14, 2010. https://doi.org/10.1038/jcbfm.2009.188
15. F. Zhao, H. Zhang, I. Rekik, Z. An, and D. Shen, "**Diagnosis of autism spectrum disorders using multi-level high-order functional networks derived from resting-state functional mri**," Frontiers in human neuroscience, vol. 12, p. 184, 2018.
16. H. Chen, X. Duan, F. Liu, F. Lu, X. Ma, Y. Zhang, L. Q. Uddin, and H. Chen, "**Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity—a multicenter**

study," Progress in Neuro-Psychopharmacology and Biological Psychiatry, vol. 64, pp. 1–9, 2016. https://doi.org/10.1016/j.pnpbp.2015.06.014

17. J. A. Nielsen, B. A. Zielinski, P. T. Fletcher, A. L. Alexander, N. Lange, E. D. Bigler, J. E. Lainhart, and J. S. Anderson, "**Multisite functional connectivity mri classification of autism: Abide results**," Frontiers in human neuroscience, vol. 7, p. 599, 2013. https://doi.org/10.3389/fnhum.2013.00599
18. C.-C. J. Kuo, "**The CNN as a guided multilayer recos transform [lecture notes]**," IEEE signal processing magazine, vol. 34, no. 3, pp. 81–89, 2017. https://doi.org/10.1109/MSP.2017.2671158
19. C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, et al., "**The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives**," Neuroinformatics, vol. 41, 2013.
20. G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, and B. Thirion, "**Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling**," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 200–208, Springer, 2010. https://doi.org/10.1007/978-3-642-15705-9_25
21. S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "**Network modelling methods for fmri,**" Neuroimage, vol. 54, no. 2, pp. 875–891, 2011.
22. Y. Wang, J. Kang, P. B. Kemmer, and Y. Guo, "**An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation**," Frontiers in neuroscience, vol. 10, pp. 123, 2016. https://doi.org/10.3389/fnins.2016.00123
23. Norshuhani Zamin, Norita Md. Norwawi, Noreen Izza Arshad and Mohamad Amirul Asraf Ismail, "**Jawi Alphabot: A Jawi Teaching Robot for Children with Autism**," International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, no. 1.4, pp. 363-367, 2019 https://doi.org/10.30534/ijatcse/2019/5581.42019
24. Mohamed A. Saleh, N. Marbukhari and Habibah Hashim, "**A Deep Learning Approach in Robot-Assisted Behavioral Therapy for Autistic Children**", International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, no. 1.6, pp. 437-443, 2019 https://doi.org/10.30534/ijatcse/2019/6381.62019

Channel Characterization for Devices in a Turbulent Diffusive Environment: A Mobile Molecular Communication Approach

Nilay Pandey, *Graduate Student Member, IEEE*, Sandeep Joshi, *Member, IEEE*, Ranjan K. Mallik, *Fellow, IEEE*, and Brejesh Lall, *Member, IEEE*

***Abstract*—Mobile molecular communication among nano-devices has potential applications in diverse fields such as targeted drug delivery systems, designing and deploying nano-sensors and nano-actuators for detection, treatment of a large set of diseases, and monitoring of environmental conditions. In this paper, we consider a mobile molecular communication system where the fluid medium has a fully developed homogeneous turbulence, and both the transmit and the receive nano-devices are mobile. The considered molecular communication system uses information molecules to transmit information from a transmit nano-device to a receive nano-device. To fully characterize and describe this mobile molecular communication channel, we derive its statistical properties like mean, correlation function, and distribution functions. We then analyze the performance of the system in terms of its symbol error probability and channel capacity. Numerical results corroborate the derived analytical findings.**

***Index Terms*—Lagrangian method, mobile molecular communication, mobile nano-devices, symbol error probability, time-varying channel, turbulence.**

I. Introduction

RECENT advancements in the field of biology and nanotechnology have led to a tremendous interest in the design and analysis of nanonetworks consisting of biological or man-made functional components. The nano-devices forming the functional components are characterized by their small sizes and limited computational capabilities. Nanonetworks also refer to communication based on transceivers and messages that are of nano scales, even though the communication between the nano-devices is taking place in macro scale environments, such as pheromone communications and movement of pollutants in the atmosphere [1]. Using traditional electromagnetic (EM) communication in such environments is challenging. For such applications, molecular communication (MC) has emerged as a promising, nature-inspired paradigm of communication research where messages are transmitted through an exchange of signaling molecules [2], [3].

Nano-devices forming a network communicate by modulating certain properties of the transmitted signaling molecules. These properties include concentration [4], type [5], number [6], time of release [7]–[9], and position [10]. The performance of such timing channels can be further improved by introducing diversity in the channel by the use of multiple signaling molecules [11]. The information and communication theory in the case of EM communication is well explored in the context of communication between wireless devices for different fading scenarios [12], and the generality of the underlying mathematical reasoning and theories developed for such systems allows us to use them for designing and analyzing MC systems. However, despite these similarities, there are certain aspects of MC that are fundamentally different from EM communication, including the fact that in MC, the symbol duration is usually a random variable (RV), leading to out of order arrival of the information molecules at the receiver, the lack of a general channel impulse response (CIR), and the difficulty in finding the optimum symbol distribution [13]. These issues make the MC channel analysis difficult and require certain simplifying assumptions to make the analysis tractable [2], [9].

A majority of the works in the existing literature have considered MC channels with fixed transmitter and receiver configurations, but similar to traditional wireless channels many potential application areas require the deployment of nano-devices (we use the term devices and nano-devices interchangeably to describe the communicating units which are mobile in a nano-network), that are mobile with respect to each other. One application is a targeted drug delivery system, where a group of nano-devices work together to carry out operations like sensing the target sites and releasing the drug molecules [14]. Other examples of long-term applications of nano-devices are designing and deploying nano-sensors

Manuscript received September 16, 2019; revised February 9, 2020 and April 10, 2020; accepted April 14, 2020. Date of publication April 20, 2020; date of current version May 14, 2020. This work was supported in part by the Ministry of Electronics and Information Technology (MeiTY), Government of India, through the Visvesvaraya Ph.D. Scheme, being implemented by Digital India Corporation, and in part by the Science and Engineering Research Board, a Statutory Body of the Department of Science and Technology, Government of India, under the J. C. Bose Fellowship. The associate editor coordinating the review of this article and approving it for publication was N. Yang. *(Corresponding author: Nilay Pandey.)*

Nilay Pandey is with the Bharti School of Telecommunication Technology and Management, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: nilaypandey03@gmail.com).

Sandeep Joshi is with the Mobile Communications Research Group, Samsung Research and Development Institute India (Bangalore), Bengaluru 560037, India (e-mail: sandeep85joshi@gmail.com).

Ranjan K. Mallik and Brejesh Lall are with the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: rkmallik@ee.iitd.ernet.in; brejesh@ee.iitd.ernet.in).

Digital Object Identifier 10.1109/TMBMC.2020.2988761

2332-7804 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

and nano-actuators for detection, treatment of a large set of diseases, monitoring of environmental conditions, and sensing of structural faults [15].

The mobile nano-devices employed in such networks are required to have the capability to directly communicate with each other without the involvement of a central coordination unit similar to the concept of device-to-device communication in wireless communications [16]. The channel statistics for these communicating devices in a nano-network vary with time due to the random mobility of the communicating devices; hence, the description of these channel statistics is essential to characterize the network completely and to establish a reliable communication link. The success probability is one such important performance metric which describes the probability of successful information transmission between the devices. A mathematical framework describing the channel statistics and the performance characterization of these devices using stochastic analysis provides the design flexibility and adequate system description.

Several works in the recent past have considered the problem of mobile molecular communication. In [17], the authors considered clock synchronization between a fixed transmitter and a mobile receiver. In MC systems, the channel can be characterized in terms of the CIR, which is defined as the expected number of molecules received at the receiver over a specific time after the simultaneous release of information molecules at the transmitter. In [18], the authors presented a training-based CIR estimation for such a channel. These works assumed a CIR that changes slowly with time as the mobile devices were assumed to move slowly relative to each other. If the movement is too fast, it can lead to symbol transpositions where the random arrival times of the molecules result in the order of arrival of two or more information molecules being changed [19]. The authors here proposed the use of positional-distance codes to mitigate this problem. Mobile nano-devices which exchange information upon collision with each other through Förster resonance energy transfer (FRET) were considered in [20]. The authors in [21] established the mathematical framework required for analyzing mobile MC systems. They extended their work in [22] and provided stochastic modeling for time-variant MC channels where both the transmitter as well as the receiver can be mobile. A time-variant MC channel employing absorbing receiver was considered in [23]. In almost all of these works, the fluid medium was either assumed to be stationary or having a uniform flow. However, the most common forms of flow are either laminar or turbulent. In fact, majority of the fluid-flows present in nature and encountered in engineering applications are turbulent, examples being the movement of pollutants in the atmosphere, transport of organic material in oceans, pumping of particle suspension through pipes, blood flow in certain regions [24]. Also, an average uniform flow assumption is more justifiable in turbulent flows, where the velocity profile is much flatter compared to the parabolic profile in case of laminar flows, as long as the flow is occurring away from the boundary.

Some recent papers have explored turbulence in the context of MC [25], [26]. The authors in [27] proposed modulating information bits into stable vortex ring structures in order to minimize the inter-symbol interference (ISI) in long range MC. The assumption of a linear combination of sequential molecular signals in case of normal diffusion also holds good when turbulence is introduced in the channel [25], indicating that the linear signal processing techniques developed for traditional communication system can also be used in case of MC with turbulent channels.

The fact that the channels with turbulent flows are highly common in nature and engineering problems has motivated us to analyze channels with turbulent flow from the mobile molecular communication perspective, which, to the best our knowledge, has not been considered in the existing literature. Specifically, the main contributions of this paper are as follows:

- Considering a diffusive MC channel with fully developed turbulence (FDT), we use the Lagrangian approach to derive the turbulent relative diffusivity of this channel.
- The movement of a particle in a turbulent flow is due to fluctuations in the surrounding fluid. Assuming a short-range correlation of the fluctuating forces, we apply a Brownian motion model to describe the turbulent dispersion and derive the time varying CIR.
- To characterize and describe this turbulent channel, we derive the statistical properties like mean, correlation function, and distribution functions.
- The considered system is analyzed in terms of symbol error probability (SEP) and capacity considering a one shot communication scheme.

Note that the turbulent channel considered in this paper refers to certain regions of macro scale environments that occur in applications such as pheromone communications, movement of pollutants in the atmosphere, pumping of particle suspension through pipes, and blood flow in certain regions [27]–[29]; however, the communication that is taking place in our case is still based on transceivers and messages that are of nano/micro scales, and therefore the system is in line with the definition of nanonetworks [1], and we use the same term in our case. Apart from this, we have considered a one shot communication scheme to keep the analysis tractable and focus on characterizing the turbulent diffusive channel so as to provide a basis for further studies.

The remainder of the paper is organized as follows. Section II discusses the system model. We develop the statistics for this mobile MC channel with turbulence in Section III. Section IV derives the performance metrics, SEP, and capacity of the considered system for the case of one shot communication. In Section V, we provide numerical results, and finally, Section VI concludes the paper.

II. System Model

We consider the relative motion of the transmit and the receive nano-devices in a three-dimensional unbounded fluid medium having fully developed, stationary, homogeneous turbulence as shown in Fig. 1. The transmit and the receive nano-devices are assumed to be released in the fluid medium simultaneously, at an infinitesimal distance apart, and have identical spherical shapes. To model their relative motion in the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

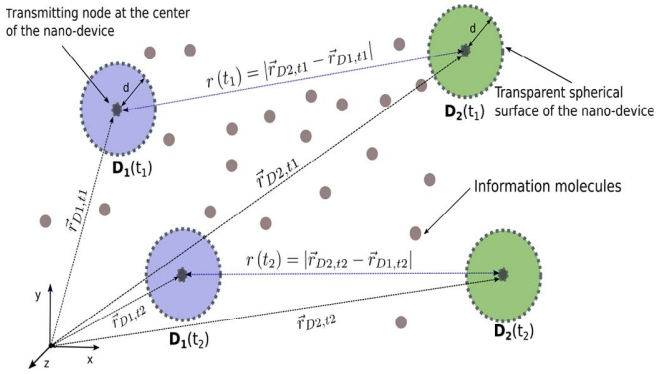


Fig. 1. Representation of the considered system model. The mobile transmit and receive nano-devices are represented as D_1 and D_2, respectively, and are spheres with the same radius. The positions at two different time instances t_1 and t_2 are represented respectively as $D_1(t_1)$ and $D_1(t_2)$ for the transmit nano-device and $D_2(t_1)$ and $D_2(t_2)$ for the receive nano-device. Their instantaneous separation at time t_n is $r(t_n)$.

case of homogeneous turbulence, it is enough to characterize the statistics of the instantaneous relative separation between the two nano-devices. We make the following assumptions regarding our system model:

- The fluid environment has a fully developed homogeneous turbulence, such that the viscous drag on the nano-devices as well as the information molecules is negligible, and the relative movements are determined by the properties of the turbulence. Note that the nano-devices as well as the information molecules are small enough (smaller, or of the order of the Kolmogorov scale, see Remark 3) to satisfy the size-independent turbulent diffusion.
- Both the nano-devices are spheres with same radius, and can release a puff of information molecules at their center to convey a message, and their surfaces act as transparent receivers.
- The fluid medium is dilute with respect to the information molecules, such that the motion of the information molecules is independent of their concentration.

It is to be noted that these are standard assumptions made in the literature to make the system analysis more tractable [2], [9]. Also, the turbulence is assumed to occur away from any surface, such that any boundary effect is not present.

Remark 1: The term homogeneous means that the average turbulent characteristics are independent of the position in the fluid, i.e., invariant to axis translation. If the field is also invariant under axis rotations, then it is isotropic, and is stationary if the average characteristics are invariant under a translation in time. Stationary, isotropic, and homogeneous statistics is a frequently made assumption while analyzing turbulence [24].

A. Turbulent Relative Diffusivity

The motion of the nano-devices in turbulent flows is dominated by the random forces acting on them, and because of these fluctuating forces, the transmit and the receive nano-devices drift apart. Let $\mathbf{r}(t) = \mathbf{r}_{rx}(t) - \mathbf{r}_{tx}(t)$ be the displacement between the two nano-devices at time t, given they are released at $t = 0$. In the Lagrangian method, the relative distance $r(t) = |\mathbf{r}(t)|$ plays an important role in modeling the turbulent relative diffusion. Given that t lies in the inertial range of time scales [30], the probability density function (PDF) of $r(t)$ in spherical coordinates is given as the solution of the differential equation

$$\frac{\partial f_{\mathbf{r}(t)}(\mathbf{r})}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 K(r)\frac{\partial f_{\mathbf{r}(t)}(\mathbf{r})}{\partial r}\right), \tag{1}$$

where $K(r)$ is a scalar called eddy diffusivity given as $K(r) = k_0\varepsilon^{1/3}r^{4/3}$, with k_0 being a dimensionless constant and ε the mean energy dissipation [30]. For infinitesimal initial separation, and considerably larger travel time than the integral time scale, a solution of (1) that is spherically symmetric is given as [31]

$$f_{r(t)}(r) = \frac{Ar^2}{\left(k_0\varepsilon^{1/3}t\right)^{9/2}}\exp\left(-\frac{9r^{2/3}}{4k_0\varepsilon^{1/3}t}\right), \tag{2}$$

where $A = 2187/560\sqrt{\pi}$ is the normalizing constant.

Richardson, in his popular paper [32], proposed that this distance between the nano-devices undergoing turbulent diffusion can be used to fully characterize the turbulent dispersion, and based on the empirical analysis of atmospheric dispersion data, reached to the conclusion that the turbulent relative diffusivity D should be given as the rate of change of the mean squared distance $\bar{r}^2(t)$ [33] as

$$D = \frac{1}{2}\frac{d}{dt}\left(\bar{r}^2(t)\right). \tag{3}$$

Here, $\bar{r}^2(t)$ can be obtained using (2) as

$$\begin{aligned}\bar{r}^2(t) &= \int_0^\infty r^2 f_{r(t)}(r)\mathrm{d}r \\ &= \frac{A}{\left(k_0\varepsilon^{1/3}t\right)^{9/2}}\int_0^\infty r^4\exp\left(-\frac{9r^{2/3}}{4k_0\varepsilon^{1/3}t}\right)\mathrm{d}r \\ &= \frac{1144}{81}\frac{\left(k_0\varepsilon^{1/3}t\right)^{15/2}}{\left(k_0\varepsilon^{1/3}t\right)^{9/2}} \approx 14k_0^3\varepsilon t^3,\end{aligned} \tag{4}$$

which is in agreement with Richardson's famous scaling law. Using (4) in (3), we get the turbulent relative diffusivity D as

$$D = 21k_0^3\varepsilon t^2. \tag{5}$$

Remark 2: Turbulent motion occurs over a very wide range of scales, both in length (eddy sizes) and time. For example, in case of clouds, the large scale can be of the order of kilometers and may persist over very long periods of time, while the small scales are of the order of millimeters. The largest eddies account for most of the transport of momentum and energy. The size (L) of the largest eddies is only limited by the physical boundaries of the flow. L is referred to as the integral length scale. The smallest length scales (μ) existing in a turbulent flow are those where the kinetic energy is dissipated into heat. These length scales L and μ have the associated turn-over times t_L and t_μ, respectively. $t_\mu \leq t \leq t_L$ is referred to as the inertial time scale.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

Remark 3: In MC channels with fully developed homogeneous turbulence, this time dependent turbulent relative diffusivity completely characterizes the relative position of the diffusing nano-devices at a given time *t*, given that they are released at an infinitesimal distance apart at $t = 0$. Also, in the case of FDT, the fluctuating accelerations, are independent of the sizes and densities if the nano-devices and the information molecules are smaller, or of the order of Kolmogorov scale [34].

B. Channel Impulse Response

The turbulent flow is characterized by the random fluctuating forces acting on the nano-devices and their relative position changes with time. Let τ represent the relative time elapsed since the time of transmission. In our case, the CIR, which is the probability that an information molecule released at the center of the transmitting nano-device at time $t \geq 0$ is present within the volume of the receiving nano-device at a given time $t + \tau$, is dependent on the exact time *t* at which the transmission occurs. Given the relative distance *r* between the centers of the two nano-devices at the time of transmission ($\tau = 0$), the time varying CIR (for $\tau \geq 0$) can be written as [28], [35]

$$h(t,\tau) = \frac{V_{Rx}}{(4\pi D\tau)^{3/2}} \exp\left(\frac{-r^2(t)}{4D\tau}\right), \tag{6}$$

where, *r* is the relative separation between the transmitting and receiving nano-devices at the time of transmission and *D* is the turbulent relative diffusivity that characterizes the relative motion between the information molecules and the receiving nano-device.

For our analysis, we assume that the nano-devices as well as the information molecules are small enough to satisfy the size independent turbulent diffusion (see Remark 1), and thus, the value of *D* can be obtained by replacing *t* by τ in (5). Since the relative separation between the nano-devices is dependent on the time at which transmission occurs, the CIR in our case is dependent on two time parameters, *t*, the time when the transmission occurs, and τ, the time elapsed since transmission (fixed *t*) over which the receiver observes the transmitted molecules. Using (5) and (6), the CIR $h(t,\tau)$ for the diffusive mobile molecular channel with FDT is expressed as

$$h(t,\tau) = \frac{V_{Rx}}{(6\pi g\varepsilon\tau^3)^{3/2}} \exp\left(-\frac{r^2(t)}{6g\varepsilon\tau^3}\right), \tag{7}$$

where *r*(*t*) has distribution as given by (2) and $g = 14k_0^3$.

Remark 4: For the inter-particle distances that are within the inertial range, the properties of turbulence can be characterized using only one parameter: the mean energy dissipation rate ϵ. Also, in the case of FDT, the particle acceleration is independent of the viscous drag, and depends only upon the fluctuating velocities of the surrounding fluid. Thus, the particle experiences small impulsive forces, like those experienced by a particle undergoing Brownian motion. Therefore, the motion of a fluid particle in FDT mimics a Markov process [36].

III. System Analysis

In this section, we take an approach similar to [22] to derive the statistical properties of the channel which are useful in fully characterizing the considered system model and give insights into the design of mobile MC systems.

A. Mean of the Channel

The mean of the channel gives the average number of information molecules traversing the medium between the transmit and receive devices. The mean of the time varying CIR $h(t,\tau)$ for a fixed *t* can be written as

$$\begin{aligned} m(t) = \mathbf{E}(h(t,\tau)) &= \int_0^\infty h(t,\tau)|_{r(t)=r} f_{r(t)}(r)\mathrm{d}r \\ &= \int_0^\infty \frac{V_{Rx}}{(6\pi g\varepsilon\tau^3)^{3/2}} \exp\left(-\frac{r^2(t)}{6g\varepsilon\tau^3}\right)\Bigg|_{r(t)=r} \\ &\quad \times \frac{Ar^2}{\left(k_0\varepsilon^{1/3}t\right)^{9/2}} \exp\left(-\frac{9r^{2/3}}{4k_0\varepsilon^{1/3}t}\right)\mathrm{d}r, \end{aligned} \tag{8}$$

where $\mathbf{E}(\cdot)$ denotes the statistical average. Substituting $r^2 = y$ in (8) we get

$$\begin{aligned} m(t) &= \frac{V_{Rx}A}{2(6\pi g\varepsilon\tau^3)^{3/2}\left(k_0\varepsilon^{1/3}t\right)^{9/2}} \\ &\quad \times \int_0^\infty y^{1/2} \exp\left(-\frac{y}{6g\varepsilon\tau^3}\right) \exp\left(-\frac{9y^{1/3}}{4k_0\varepsilon^{1/3}t}\right)\mathrm{d}y. \end{aligned} \tag{9}$$

Making a further substitution $y/(6g\varepsilon\tau^3) = x$ gives

$$\begin{aligned} m(t) &= \frac{V_{Rx}A(6g\varepsilon\tau)^{3/2}}{2(6\pi g\varepsilon\tau^3)^{3/2}\left(k_0\varepsilon^{1/3}t\right)^{9/2}} \\ &\quad \times \int_0^\infty x^{1/2}\exp(-x) \exp\left(-\frac{9(6g)^{1/3}\tau}{4k_0 t}x^{1/3}\right)\mathrm{d}x \\ &= \frac{V_{Rx}A}{2(\pi\tau^2)^{3/2}\left(k_0\varepsilon^{1/3}t\right)^{9/2}} \\ &\quad \times \int_0^\infty \exp(-x)x^{1/2} \exp\left(-\frac{a_0\tau}{t}x^{1/3}\right)\mathrm{d}x, \end{aligned} \tag{10}$$

where $a_0 = (9(6g)^{1/3})/(4k_0)$. We solve the integral of the type $\int_0^\infty \exp(-x)f(x)\mathrm{d}x$ in (10) using the Gauss-Laguerre technique resulting in

$$\begin{aligned} m(t) &= \frac{V_{Rx}A}{2(\pi\tau^2)^{3/2}\left(k_0\varepsilon^{1/3}t\right)^{9/2}} \\ &\quad \times \sum_{i=1}^{n} w_i x_i^{1/2} \exp\left(-\frac{a_0\tau}{t}x_i^{1/3}\right), \end{aligned} \tag{11}$$

where x_i represents the roots of the Laguerre polynomial $L_n(x)$ [37], and w_i is the weight given by

$$w_i = \frac{x_i}{(n+1)^2[L_{n+1}(x_i)]^2}. \tag{12}$$

Note that the mean *m*(*t*) depends on both the time parameters *t* and τ. This is obvious as with increasing *t*, the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

devices move away from each other. Also, in case of turbulent diffusion, the inter-particle separation changes much faster with time as compared to simple diffusion. This dependence of the mean on t makes $h(t,\tau)$ a non-stationary process.

B. Complementary Cumulative Distribution Function

The complementary cumulative distribution function (CCDF) is an important statistical property which also gives the success probability. The success probability is the probability that the impulse response of the channel exceeds a minimum threshold value; this minimum value can be so chosen that it ensures that the maximum number of information molecules reach the receive nano-device from the transmit nano-device. We can also use the CCDF to derive the outage probability which is the cumulative distribution function (CDF) of the CIR, and is the complement of the CCDF, i.e., $\Pr(h(t,\tau) < x)$.

The CCDF of the impulse response can be given as

$$\begin{aligned} P_c &= \Pr(h(t,\tau) > x) \\ &= \Pr\left(\frac{V_{Rx}}{(6\pi g\varepsilon\tau^3)^{3/2}} \exp\left(-\frac{r^2(t)}{6g\varepsilon\tau^3}\right) > x \right) \\ &= \Pr\left(r(t) < \sqrt{6g\varepsilon\tau^3}\left(\ln\left(\frac{\phi}{x}\right)\right)^{1/2} \right), \end{aligned} \tag{13}$$

where, $\phi = V_{Rx}/(6\pi g\varepsilon\tau^3)^{3/2}$. Using (2) and (13), we get

$$P_c = \int_0^k \frac{Ar^2}{\left(k_0\varepsilon^{1/3}t\right)^{9/2}} \exp\left(-\frac{9r^{2/3}}{4k_0\varepsilon^{1/3}t}\right) \mathrm{d}t. \tag{14}$$

Solving the integral in (14) and further simplification results in the expression for the CCDF as

$$\begin{aligned} P_c = \frac{Ar^2}{\left(k_0\varepsilon^{1/3}\right)^{9/2}} \Bigg\{ &\exp\left(-\frac{A_1}{k}\right)\left(\frac{4A_1^2 + 10A_1k + 15k^2}{4A_1^{3/2}k^{5/2}}\right) \\ &+ \frac{15\sqrt{\pi}\mathrm{erfc}\left(\sqrt{A_1/k}\right)}{8A_1^{7/2}} \Bigg\}, \end{aligned} \tag{15}$$

where $k = \sqrt{6g\varepsilon\tau^3}(\ln(\frac{\phi}{x}))^{1/2}$, $A_1 = 9r^{2/3}/4k_0\varepsilon^{1/3}$, and $\mathrm{erfc}(\cdot)$ denotes the complementary error function.

C. Auto Correlation Function

The auto correlation function (ACF) gives the correlation between the behavior of the nano-devices at two different time instants say t_1 and t_2. The ACF is a useful statistical parameter which is used to compute the coherence time of the channel which is given as the time for which there is no change in the turbulent scenario of the considered medium similar to the case of wireless communication where the coherence time is measured in terms of the channel fading. The ACF is expressed as

$$\begin{aligned} \phi(t_1,t_2) &= \mathbf{E}[h(t_1,\tau)h(t_2,\tau)] \\ &= \int_0^\infty \int_0^\infty h(t_1,\tau)|_{r(t)=r_1} \times h(t_2,\tau)|_{r(t)=r_2} \\ &\quad \times f_{r(t_1)r(t_2)}(r_1,r_2)\mathrm{d}r_1\mathrm{d}r_2. \end{aligned} \tag{16}$$

For distances larger than the Kolmogorov length scale, the turbulent diffusion follows a Markov's model (see Remark 2), such that $f_{r(t_1),r(t_2)}(r_1,r_2) = f_{r(t_1)}(r_1)f_{r(t_2)}(r_2|r_1)$, and we can write $f_{r(t_2)}(r_2|r_1,r_0) = f_{r(t_2)}(r_2|r_1)$, where

$$f_{r(t_2)}(r_2/r_1) = \frac{Ar^2}{\left(k_0\varepsilon^{1/3}t_2\right)^{9/2}} \exp\left(-\frac{9r^{2/3}}{4k_0\varepsilon^{1/3}t_2}\right). \tag{17}$$

Using (16) and (17), we can express the ACF $\phi(t_1,t_2)$ as

$$\begin{aligned} \phi(t_1,t_2) = \frac{V_{Rx}^2A^2}{(6\pi g\varepsilon\tau^3)^3\left(k_0\varepsilon^{1/3}t_1t_2\right)^9} &\times \int_0^\infty \int_0^\infty \exp\left(-\frac{r_1^2}{6g\varepsilon\tau^3}\right)\exp\left(-\frac{9r_1^{2/3}}{4k_0\varepsilon^{1/3}t_1}\right) r_1^2\mathrm{d}r_1 \\ &\times \exp\left(-\frac{r_2^2}{6g\varepsilon\tau^3}\right)\exp\left(-\frac{9r_2^{2/3}}{4k_0\varepsilon^{1/3}t_2}\right) r_2^2\mathrm{d}r_2. \end{aligned} \tag{18}$$

Upon further simplification, we get

$$\begin{aligned} \phi(t_1,t_2) = \frac{9\,V_{Rx}^2A^2}{4(6\pi g\varepsilon\tau^3)^3\left(k_0\varepsilon^{1/3}t_1t_2\right)^9} &\times \int_0^\infty \int_0^\infty x^{7/2}\exp\left(-bx^3 - c_1x\right) \\ &\times y^{7/2}\exp\left(-by^3 - c_2y\right)\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{19}$$

where, $b = 1/6\pi g\varepsilon\tau^3$, $c_1 = 9/4k_0\varepsilon^{1/3}t_1$, and $c_2 = 9/4k_0\varepsilon^{1/3}t_2$.

In this paper, (19) is solved numerically, however, integrals of the form $\int_0^\infty x^{7/2}\exp(-cx - dx^3)\mathrm{d}x$ occuring in (19) can be solved to give closed form expressions in terms of Airy functions *Ai*(*x*) and *Bi*(*x*) [37], and their derivatives. However, the resulting expressions are not very tractable and complicated. These functions are commonly encountered in optics and quantum mechanics and are given as

$$Ai(x) = \frac{1}{\pi}\sqrt{\frac{x}{3}}K_{1/3}\left(\frac{2}{3}x^{3/2}\right), \tag{20a}$$

$$Bi(x) = \sqrt{\frac{x}{3}}\left[I_{-1/3}\left(\frac{2}{3}x^{3/2}\right) + I_{1/3}\left(\frac{2}{3}x^{3/2}\right)\right], \tag{20b}$$

where $I_v(x)$ denotes the modified Bessel function of the first kind and order v and $K_u(x)$ denotes the modified Bessel function of the second kind and order u.

D. Probablity Density Function

The PDF of the impulse response is derived from the CDF which is given as

$$F_{h(t,\tau)}(x) = \Pr(h(t,\tau) \le x). \tag{21}$$

Simplifying similar to (13), we get

$$\begin{aligned} F_{h(t,\tau)}(x) = \frac{Ar^2}{\left(k_0\varepsilon^{1/3}\right)^{9/2}} \Bigg\{ &-\exp\left(-\frac{A_1}{k}\right) \\ &\times \left(\frac{4A_1^2 + 10A_1k + 15k^2}{4A_1^{3/2}k^{5/2}}\right) \\ &+ \frac{15\sqrt{\pi}\mathrm{erf}\left(\sqrt{A_1/k}\right)}{8A_1^{7/2}} \Bigg\}, \end{aligned} \tag{22}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

where $k = \sqrt{6g\varepsilon\tau^3}(ln(\frac{\Phi}{x}))^{1/2}$, $A_1 = 9r^{2/3}/4k_0\varepsilon^{1/3}$, and *erf*$(\cdot)$ denotes the error function.

The PDF is obtained by the differentiation of the CDF and is given as (23) shown at the bottom of the page.

IV. Performance Analysis

In this section, we present the performance analysis of the considered system model in terms of the SEP and the capacity. We consider a one shot communication between the transmit and the receive nano-devices. The detection mechanism is a single sample detection, where the receive nano-device simply counts the number of information molecules present within the receiver volume V_{Rx} at a time τ_s after the release of the molecules. Based on this number, the receive nano-device makes a decision regarding the transmitted symbol.

A. SEP Analysis

Let the input and the output alphabets for an *M*-ary communication be $\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$ and $\mathcal{Y} = \{y_1, y_2 \ldots, y_M\}$, respectively. To send a message $x_i \in \mathcal{X}$, N_i molecules are released simultaneously at the center of the transmit nano-devices at time *t*. The receive nano-device counts the number of information molecules *R* present within the receiver volume V_{Rx} at a time $t + \tau_s$. For simplicity, we assume that $N_1 < N_2 \ldots < N_M$. Let *X* and *Y* be the channel input and output, respectively. The SEP for the one shot communication scheme is given as

$$P_{\mathrm{e}} = \sum_{i=1}^{M} p_i \Pr\{err|X = x_i\}, \tag{24}$$

where, p_i is the a priori probability of sending the symbol x_i and $\Pr\{err|X = x_i\}$ denotes the probability of error when symbol x_i is transmitted and is given as

$$\begin{aligned}\Pr\{err|X = x_i\} = &\sum_{r=0}^{\eta_{i-1}} \binom{N_i}{r} h(t,\tau_s)(1-h(t,\tau_s))^{N_i-r} \\ &+ \sum_{r=\eta_i}^{N_i} \binom{N_i}{r} h(t,\tau_s)(1-h(t,\tau_s))^{N_i-r},\end{aligned} \tag{25}$$

where η_i is the decision threshold for x_i. For $N = N_i$, the number of observed information molecules *R* follows a binomial distribution with parameters N_i and $h(t,\tau_s)$. Under the assumption $N_1 < N_2 \ldots < N_M$, the decision threshold $\eta_i \in \{\eta_1, \eta_2, \ldots, \eta_{M-1}\}$ for maximum likelihood (ML) decision rule can be obtained using a pairwise comparison of the probability distributions and is given as $\eta_i = \lfloor \eta_i' \rceil$, where $\lfloor \cdot \rceil$ represents the nearest integer operator, and η_i' satisfies

$$\frac{\Gamma(N_{i+1} - \eta_i' + 1)\Gamma(N_i + 1)}{\Gamma(N_i - \eta_i' + 1)\Gamma(N_{i+1} + 1)} = \frac{p_{i+1}}{p_i}(1 - h(t,\tau_s))^{N_{i+1}-N_i}, \tag{26}$$

with $\eta_0 = 0$ and $\eta_M = N_M$.

The above equations related to the SEP are evaluated numerically in this paper. However, given the large number of transmitted molecules for each of the symbols, and the fact that the probability of an individual molecule being inside the receiver volume at the sampling time $t + \tau_s$, i.e., $h(t,\tau_s)$ is very small (see Fig. 3), we can use a Poisson's distribution with mean $\lambda_i = N_i h(t,\tau_s)$ to accurately approximate the distribution of *R*. With the Poisson's approximation, $\Pr\{err|X = x_i\}$ becomes

$$\Pr\{err|X = x_i\}|_{poi} = 1 - \exp(-\lambda_i)\left(\sum_{r=\eta_{i-1}}^{\eta_i} \frac{\lambda_i^r}{r!}\right), \tag{27}$$

where, the threshold η_i, under the assumption $N_1 < N_2 \ldots < N_M$, is given as $\eta_i = \lfloor \eta_i' \rceil$, where

$$\eta_i' = \frac{\ln(p_i/p_{i+1}) + (\lambda_{i+1} - \lambda_i)}{\ln(N_{i+1}/N_i)}, \tag{28}$$

with $\eta_0 = 0$ and $\eta_M = N_M$.

B. Capacity Analysis

We consider an *M*-ary modulation scheme as before, with input and output alphabets given by $\mathcal{X}, \mathcal{Y}$. We denote the conditional probability $\Pr(Y = y_j|X = x_i)$ by *p*(*j*|*i*). The mutual information between input *X* and output *Y* within a slot is given as

$$I(X;Y) = \sum_{j=1}^{M}\sum_{i=1}^{M} p_i p(j|i) \log_2 \frac{p(j|i)}{\sum_{i=1}^{M} p_i p(j|i)}, \tag{29}$$

where $\{p_i\}$ is the set of a priori probabilities and

$$p(j|i) = \sum_{r=\eta_{j-1}}^{\eta_j} \binom{N_i}{r} h(t,\tau_s)(1-h(t,\tau_s))^{N_i-r}, \tag{30}$$

and with the Poisson's approximation

$$p(j|i)|_{Poi} = \exp(-\lambda_i)\left(\sum_{r=\eta_{j-1}}^{\eta_j} \frac{\lambda_i^r}{r!}\right). \tag{31}$$

The channel capacity is given as

$$C = \max_{\{p_i\}} I(X;Y), \tag{32}$$

$$f_{h(t,\tau)}(x) = \frac{Ar^2}{\left(k_0\varepsilon^{1/3}\right)^{9/2}} \exp\left(\frac{-A_1}{k}\right) \frac{\left(6g\varepsilon\tau^3\right)}{4x\left(k^5 A_1^3\right)^{1/2}} \left\{ \left(\frac{10A_1 + 45k^2\left(6g\varepsilon\tau^3\right)^{-3/2}}{2k}\right) + \frac{15}{4} + \frac{\left(4A_1^2 + 10A_1 k + 15k^3\left(6g\varepsilon\tau^3\right)^{-3/2}\right)(2A_1 - 5)}{4k^3} \right\} \tag{23}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
N_t	50000
V_{Rx}	4 μm^3
ϵ	$0.2, 0.4, 0.6\ (m^2/s^3)$
$\tau = \tau_s$	1 ms
τ_n	0.1 ms
T_L	250 ms
g	1
Δt	$0.07\tau_n$

where, the maximization is with respect to the set of a priori probabilities $\{p_i\}$.

We use the method of Lagrange multipliers to maximize the mutual information with respect to the constraint $p_1 + p_2 + \cdots + p_M = 1$. The Lagrange function for our case can be written as

$$\Lambda(p_1, \ldots, p_M, \lambda) = I(X;Y) + \lambda(p_1 + p_2 + \cdots + p_M - 1), \tag{33}$$

where λ is the Lagrange multiplier. Taking the partial derivatives of the Lagrange function with respect to p_k, $1 \le k \le M$, and equating them to zero, we get the set of equations

$$\sum_{j=1}^{M}\left[\sum_{i=1}^{M} p(j|k)\log_2\frac{p(j|i)}{\sum_{i=1}^{M} p_i p(j|i)} + \frac{p_i}{\ln 2}p(j|i)\left(\frac{p(j|k)}{\sum_{i=1}^{M} p_i p(j|i)}\right)\right] = -\lambda, \tag{34}$$

where $k = 1, \ldots, M$.

We solve the set of equations in (34) numerically to obtain the optimum $(\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_M, \bar{\lambda})$. At this optimum, both the Lagrange multiplier as well as the mutual information given by (29) attain their respective maximum values. Using (29), the capacity per channel use can be obtained as

$$C = \sum_{k=1}^{M}\bar{p}_k\left[\sum_{j=1}^{M}\sum_{i=1}^{M} -\frac{\bar{p}_i}{\ln 2}p(j|i)\left(\frac{p(j|k)}{\sum_{i=1}^{M}\bar{p}_k p(j|i)}\right) - \bar{\lambda}\right]. \tag{35}$$

V. NUMERICAL RESULTS

In this section, we present the numerical and simulation results for the statistics and the performance analysis derived in the previous sections for the considered turbulent diffusive MC channel. We have used typical parameter values that have been reported in the literature concerning the Langrangian statistics of a pair of particles in homogeneous isotropic turbulence [30]. The inertial range of time scales is taken as $\tau_n < t < T_L$. The values of the parameters used in the numerical results and simulations are as given in Table I.

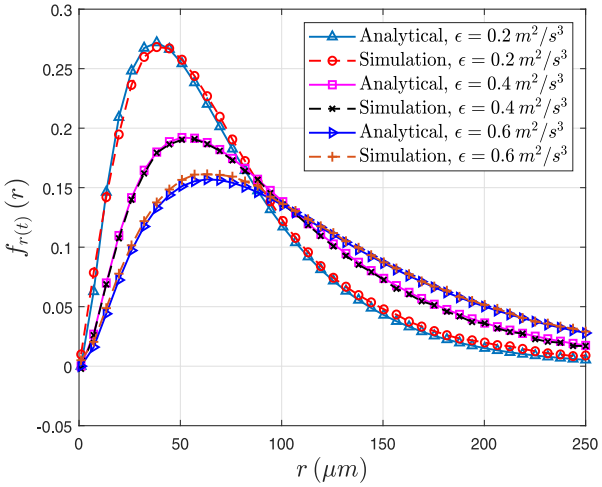


Fig. 2. PDF $f_{r(t)}(r)$ for $\epsilon = 0.2, 0.4$, and 0.6.

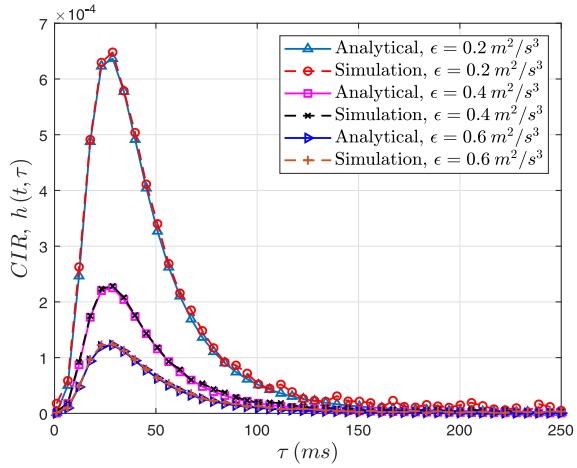


Fig. 3. CIR $h(t,\tau)$ versus τ for $\epsilon = 0.2, 0.4$, and 0.6.

To simulate the turbulent diffusion, we adopt the Monte-Carlo method, where the results are averaged over 10,000 independent realizations of the turbulent system. As mentioned in Remark 2, assuming short range correlations of the fluctuating forces, we adopt a Brownian motion model for the movement of the nano-devices undergoing turbulent diffusion. The algorithm advances in time steps of Δt, and for each iteration step, the turbulent relative diffusivity is calculated using (5). We consider the simulation only over the inertial time range. Different values of ϵ correspond to different values of the Reynolds number [30].

Fig. 2 gives the plot of the PDF of $r(t)$ versus r for values of $\epsilon = 0.2, 0.4$, and 0.6. The plots are obtained from (2). As observed from the plots, the PDF changes more rapidly for smaller values of ϵ. This results from the fact that for smaller values ϵ the mean energy dissipation is low, resulting in faster initial separation. This also indicates that the nano-devices have greater chances of remaining relatively close to each other, however, some of them can move very far away from each other, as indicated by the heavy tails of the PDF.

The effects of the time parameters t and τ on the CIR of the channel are presented in Figs. 3 and 4, respectively

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

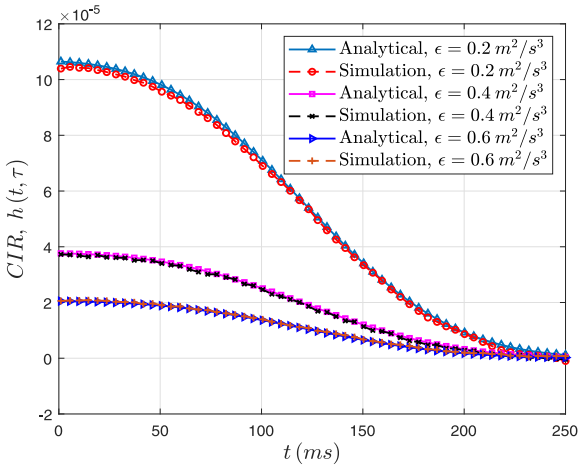


Fig. 4. CIR $h(t,\tau)$ versus t for $\epsilon = 0.2, 0.4$, and 0.6.

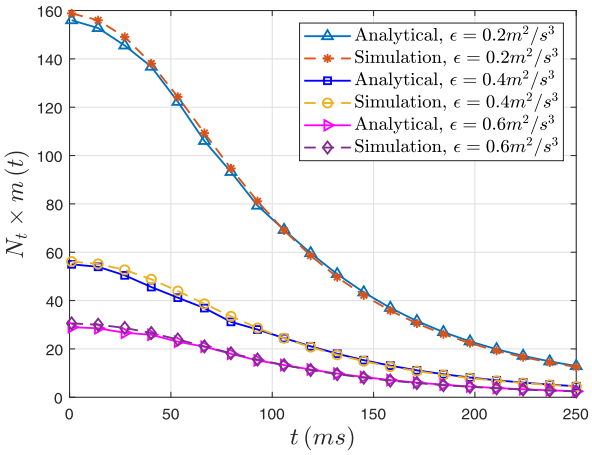


Fig. 5. Mean $m(t)$ multiplied by the number of transmitted molecules N_t versus t for $\epsilon = 0.2, 0.4$, and 0.6.

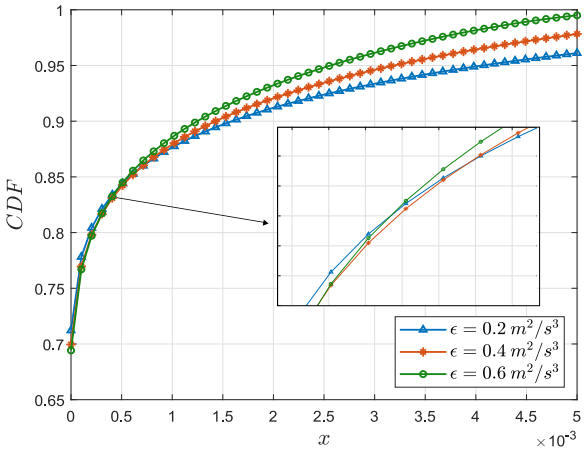


Fig. 6. CDF as a function of the threshold x for $\epsilon = 0.2, 0.4$, and 0.6.

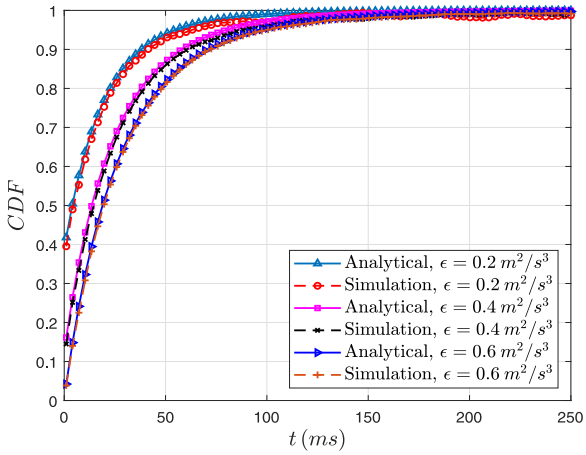


Fig. 7. CDF as a function of the time t at which the transmission occurs for $\epsilon = 0.2, 0.4$, and 0.6 for a fixed threshold $x = 0.0001$.

for three values of the parameter ϵ. The plots are obtained from (7). As seen in Fig. 3, the CIR reaches its maximum quickly for smaller values of τ, then decays exponentially. This decay occurs as with increase in τ, the information molecules diffuse away very quickly, as is obvious from Fig. 3, where the turbulent relative diffusivity increases as τ^2. The plots also show both the analytical and simulation results to be in close agreement with each other. The variation in CIR in Fig. 4 with the time t at which transmission occurs is as expected. As both the nano-devices in the turbulence are moving away from each other, higher value of t means that they are more distant from each other, and hence the number of information molecules reaching the receiving nano-device is less. Here, as seen from the plot, a small value of the energy dissipation parameter ϵ results in more molecules being able to arrive within the volume of the receive nano-device.

Fig. 5 gives the plots of the mean multiplied by the number of transmitted molecules $N_t \times m(t)$ versus t for values of $\epsilon = 0.2, 0.4$, and 0.6. The plots also show the simulation results along with the analytical results. It can be seen from the plots that $N_t \times m(t)$ decreases as t increases. This behavior is due to the fact that with increase in t, the separation between the nano-devices increases. The value of $N_t \times m(t)$ is high for small values of ϵ as more information molecules are able to reach the receiver volume. Also the decrease is more pronounced for smaller value of ϵ, as for small t and ϵ, the devices are relatively close to each other, and then separate rapidly as t increases and cancels out the effect of small ϵ.

We investigate the variation of the CDF of the CIR versus the threshold x and t for a fixed threshold $x = 0.0001$ for different values of $\epsilon = 0.2, 0.4$, and 0.6 in Fig. 6 and Fig. 7, respectively. As can be seen from the plots, for higher values of the mean energy dissipation parameter, the CDF, which gives the outage probability, is also high. One interesting observation in Fig. 6 is that the CDF for larger ϵ is lower than that for smaller ϵ for very small values of the threshold x, and becomes higher as the threshold increases. This behavior can be contributed to the fact that a small ϵ results in small diffusivity, allowing less information molecules to reach the receiving device, however, a large ϵ leads to larger energy dissipation, leading to higher CDF for greater thresholds. A similar pattern

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

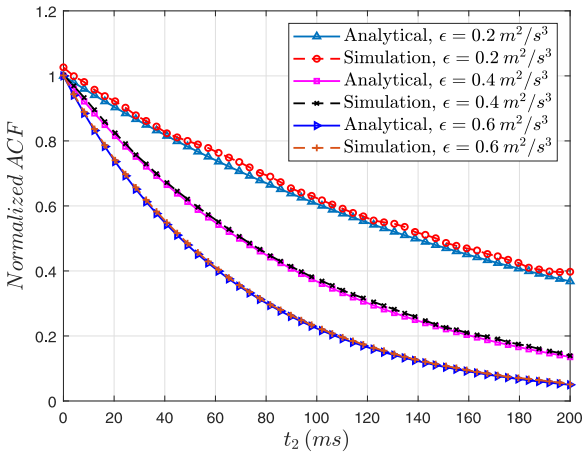


Fig. 8. Normalized ACF versus t_2 for $\epsilon = 0.2, 0.4$, and 0.6.

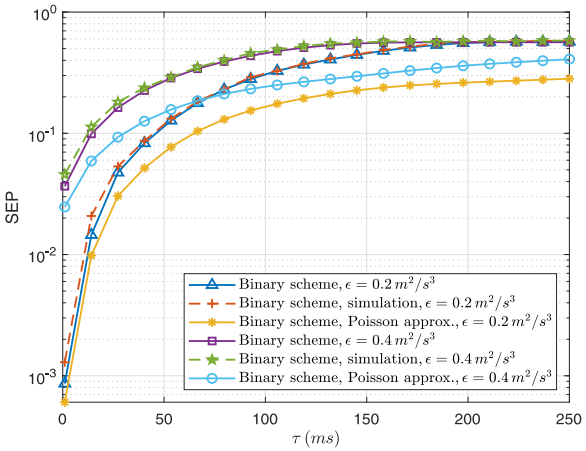


Fig. 9. SEP as a function of the sampling time τ_s for $\epsilon = 0.2$ and 0.4 for binary modulation.

is observed in Fig. 7, where the plots are generated for a very small threshold, and thus the CDF is higher for smaller ϵ. The CDF, and thus, the outage probability increases with increasing time t, as the nano-devices move further apart.

Fig. 8 gives the variation of the normalized ACF versus t_2 for $\epsilon = 0.2, 0.4$, and 0.6 with $t_1 = 0.1$. As expected, the ACF becomes lower as t_2 increases, since with increasing t, the separation $r(t)$ between the nano-devices also increases. This results in a lower correlation between the channel impulses for times t_1 and t_2. This decrease is more rapid for higher values of ϵ, as the relative diffusivity is higher, and the separation between the devices increases rapidly.

Fig. 9 gives the variation of the SEP as a function of the sampling time τ for binary modulation with ϵ values as 0.2 and 0.4. The SEP results are validated using simulation and are found to be in close agreement. The SEP results are shown for the binomial distribution and for the Poisson approximation. It can be observed from the plots that the SEP performance is better for the case of Poisson approximation. Also, with the increase in the value of ϵ the SEP performance deteriorates.

Fig. 10 shows the variation of the capacity as a function of the sampling time τ assuming a binary modulation scheme

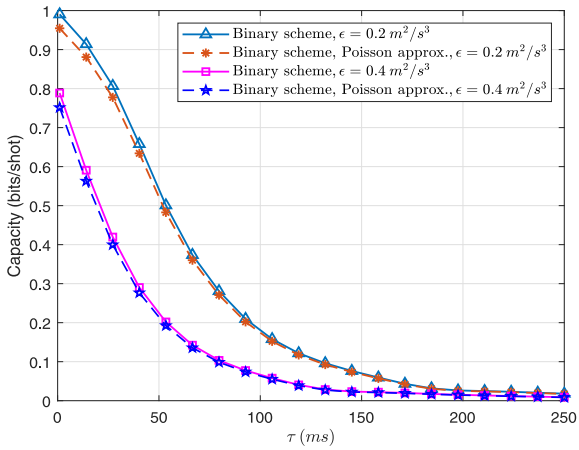


Fig. 10. Capacity as a function of the sampling time τ_s for $\epsilon = 0.2$ and 0.4 for a binary modulation scheme.

for the ϵ values of 0.2 and 0.4. The capacity analysis is shown for the case of binomial distribution as well as Poisson approximation. It is observed that the capacity performance for the binomial and the Poisson approximation case are almost similar. Further, it can be observed that the lower values of ϵ result in higher capacity values which is intuitive.

VI. Conclusion

In this paper, we have considered a mobile MC system in the presence of a fully developed homogeneous turbulence in a fluid medium. Both the transmitting and the receiving nano-devices are assumed to be mobile. We have derived the system statistical properties, namely the mean, the ACF, the CCDF, the CDF, the PDF, and the impulse response, to completely characterize the considered system model. We have also presented the performance analysis of the considered mobile molecular communication system in terms of the SEP and the capacity assuming binary modulation scheme. Furthermore, we have validated the analytical results with the simulation results for the mobile MC system and depicted the effect of the channel parameters on the channel statistics.

References

[1] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A new communication paradigm," *Comput. Netw.*, vol. 52, no. 12, pp. 2260–2279, Aug. 2008.

[2] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1887–1919, 3rd Quart., 2016.

[3] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication—A tutorial review," *Proc. IEEE*, vol. 107, no. 7, pp. 1256–1301, Jul. 2019.

[4] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Interference effects on modulation techniques in diffusion based nanonetworks," *Nano Commun. Netw.*, vol. 3, no. 1, pp. 65–73, Mar. 2012.

[5] N.-R. Kim and C.-B. Chae, "Novel modulation techniques using isomers as messenger molecules for nano communication networks via diffusion," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 847–856, Dec. 2013.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

[6] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, "Modulation techniques for communication via diffusion in nanonetworks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–5.
[7] K. V. Srinivas, A. W. Eckford, and R. S. Adve, "Molecular communication in fluid media: The additive inverse Gaussian noise channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4678–4692, Jul. 2012.
[8] N. Pandey, R. K. Mallik, and B. Lall, "Performance analysis of diffusive molecular timing channels," *IET Commun.*, vol. 13, no. 18, pp. 3059–3067, Nov. 2019.
[9] N. Farsad, Y. Murin, A. W. Eckford, and A. Goldsmith, "Capacity limits of diffusion-based molecular timing channels with finite particle lifetime," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 2, pp. 88–106, Jun. 2018.
[10] N. Pandey, R. K. Mallik, and B. Lall, "Molecular communication: The first arrival position channel," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 508–511, Apr. 2019.
[11] Y. Murin, M. Chowdhury, N. Farsad, and A. Goldsmith, "Diversity gain of one-shot communication over molecular timing channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.
[12] S. Joshi and R. K. Mallik, "Coverage probability analysis in a device-to-device network: Interference functional and laplace transform based approach," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 466–469, Mar. 2019.
[13] C. Rose and I. S. Mian, "A fundamental framework for molecular communication channels: Timing and payload," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1043–1048.
[14] T. Nakano, A. W. Eckford, and T. Haraguchi, *Molecular Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[15] L. Felicetti, M. Femminella, G. Reali, and P. Liò, "Applications of molecular communications to medicine: A survey," *Nano Commun. Netw.*, vol. 7, pp. 27–45, Mar. 2016.
[16] S. Joshi and R. K. Mallik, "Analysis of dedicated and shared device-to-device communication in cellular networks over Nakagami-*m* fading channels," *IET Commun.*, vol. 11, no. 10, pp. 1600–1609, Jul. 2017.
[17] Z. Luo, L. Lin, and M. Ma, "Offset estimation for clock synchronization in mobile molecular communication system," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Doha, Qatar, Apr. 2016, pp. 1–6.
[18] V. Jamali, A. Ahmadzadeh, C. Jardin, H. Sticht, and R. Schober, "Channel estimation for diffusive molecular communications," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4238–4252, Oct. 2016.
[19] S. Qiu, T. Asyhari, and W. Guo, "Mobile molecular communications: Positional-distance codes," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Edinburgh, U.K., Jul. 2016, pp. 1–5.
[20] M. Kuscu and O. B. Akan, "A communication theoretical analysis of fret-based mobile ad hoc molecular nanonetworks," *IEEE Trans. Nanobiosci.*, vol. 13, no. 3, pp. 255–266, Sep. 2014.
[21] A. Ahmadzadeh, V. Jamali, A. Noel, and R. Schober, "Diffusive mobile molecular communications over time-variant channels," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1265–1268, Jun. 2017.
[22] A. Ahmadzadeh, V. Jamali, and R. Schober, "Stochastic channel modeling for diffusive mobile molecular communication systems," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6205–6220, Dec. 2018.
[23] T. N. Cao *et al.*, "Diffusive mobile MC for controlled-release drug delivery with absorbing receiver," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–7.
[24] P. A. Davidson, *Turbulence: An Introduction for Scientists and Engineers*. Oxford, U.K.: Oxford Univ. Press, 2015.
[25] M. Abbaszadeh, H. B. Yilmaz, P. J. Thomas, and W. Guo, "Linearity of sequential molecular signals in turbulent diffusion channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.
[26] M. Abbaszadeh, G. Moutsinas, P. J. Thomas, and W. Guo, "Uncertainty quantification in molecular signals using polynomial chaos expansion," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 1, pp. 39–42, Dec. 2018.
[27] M. Abbaszadeh, P. J. Thomas, and W. Guo, "Toward high capacity molecular communications using sequential vortex rings," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 1, pp. 39–42, Mar. 2018.
[28] B. D. Unluturk and I. F. Akyildiz, "An end-to-end model of plant pheromone channel for long range molecular communication," *IEEE Trans. Nanobiosci.*, vol. 16, no. 1, pp. 11–20, Jan. 2017.
[29] N. Farsad, W. Guo, and A. W. Eckford, "Tabletop molecular communication: Text messages through chemical signals," *PloS One*, vol. 8, no. 12, Dec. 2013, Art. no. e82935.
[30] L. Biferale, G. Boffetta, A. Celani, B. Devenish, A. Lanotte, and F. Toschi, "Lagrangian statistics of particle pairs in homogeneous isotropic turbulence," *Phys. Fluids*, vol. 17, no. 11, Nov. 2005, Art. no. 115101.
[31] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence*. Chelmsford, MA, USA: Courier Corp., 2013.
[32] L. F. Richardson, "Atmospheric diffusion shown on a distance-neighbour graph," *Proc. Roy. Soc. A, Math. Phys. Character*, vol. 110, no. 756 , pp. 709–737, Apr. 1926.
[33] B. Shivamoggi, "A generalized Brownian motion model for turbulent relative particle dispersion," *Phys. Lett. A*, vol. 380, no. 36 , pp. 2809–2814, Aug. 2016.
[34] Y. Meller and A. Liberzon, "Particle–fluid interaction forces as the source of acceleration PDF invariance in particle size," *Int. J. Multiphase Flow*, vol. 76, pp. 22–31, Nov. 2015.
[35] P. Nelson, *Biological Physics*. New York, NY, USA: W. H. Freeman, 2004.
[36] A. M. Obukhov, "Description of turbulence in terms of Lagrangian variables," *Adv. Geophys.*, vol. 6, pp. 113–116, Jan. 1959.
[37] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Cambridge, MA, USA: Academic, 2014.

Nilay Pandey (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Graphic Era University, India, in 2012, and the M.Tech. degree in signal and image processing from the National Institute of Technology, Rourkela, India, in 2014. He is currently pursuing the Ph.D. degree with the Bharti School of Telecommunication Technology and Management, Indian Institute of Technology Delhi, India. From July 2014 to March 2015, he was a Senior Research Fellow, working on a project with the Department of Atomic Energy, India. His research interests include molecular communication, MIMO radar, signal and image processing, and machine learning.

Sandeep Joshi (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Uttar Pradesh Technical University, India, in 2006, the M.E. degree in communication engineering from the Birla Institute of Technology and Science, Pilani, India, in 2009, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology Delhi in 2019. From July 2008 to December 2008, he was an Interim Intern with Qualcomm India Pvt. Ltd., Hyderabad. From February 2009 to June 2009, he was a Senior Research Associate with Evalueserve Pvt. Ltd., Gurgaon. From July 2009 to June 2011, he was a Faculty Member with the Department of Electronics and Communication Engineering, Jaypee Institute of Information and Technology University, Noida. From July 2011 to December 2014, he was a Founding Faculty Member with the Department of Electronics and Communication Engineering, School of Engineering, Shiv Nadar University, Noida. From August 2019 to November 2019, he was a Early-Doctoral Research Fellow with the Department of Electrical Engineering, Indian Institute of Technology Delhi. Since November 2019, he has been with the Samsung R&D Institute India (Bangalore), Bengaluru, where he is currently a Senior Lead Engineer with the Mobile Communications Research Group. His research interests include communication theory, device-to-device communications, smart grid communications, molecular communications, and performance analysis of wireless communication systems.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

Ranjan K. Mallik (Fellow, IEEE) received the B.Tech. degree from the Indian Institute of Technology, Kanpur, in 1987, and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1988 and 1992, respectively, all in electrical engineering. From August 1992 to November 1994, he was a scientist with the Defence Electronics Research Laboratory, Hyderabad, India, working on missile and EW projects. From November 1994 to January 1996, he was a faculty member of the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur. From January 1996 to December 1998, he was with the faculty of the Department of Electronics and Communication Engineering, Indian Institute of Technology, Guwahati. Since December 1998, he has been with the faculty of the Department of Electrical Engineering, Indian Institute of Technology, Delhi, where he is currently a Professor. His research interests are in diversity combining and channel modeling for wireless communications, space-time systems, cooperative communications, multiple-access systems, power line communications, difference equations, and linear algebra.

Dr. Mallik is a member of Eta Kappa Nu. He is also a member of the IEEE Communications, Information Theory, and Vehicular Technology Societies, the American Mathematical Society, and the International Linear Algebra Society; a fellow of the IEEE, the Indian National Academy of Engineering, the Indian National Science Academy, The National Academy of Sciences, India, Prayagraj, the Indian Academy of Sciences, Bengaluru, The World Academy of Sciences-for the advancement of science in developing countries (TWAS), The Institution of Engineering and Technology, U.K., The Institution of Electronics and Telecommunication Engineers, India, and The Institution of Engineers (India); and a life member of the Indian Society for Technical Education. He is a recipient of the Hari Om Ashram Prerit Dr. Vikram Sarabhai Research Award in the field of electronics, telematics, informatics, and automation, the Shanti Swarup Bhatnagar Prize in engineering sciences, the Khosla National Award, and the J. C. Bose Fellowship. He has served as an Area Editor and an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS.

Brejesh Lall (Member, IEEE) received the B.E. degree in electronics and communication engineering and the M.E. degree in signal processing from the Delhi College of Engineering, New Delhi, in 1991 and 1992, respectively, and the Ph.D. degree in signal processing from the Indian Institute of Technology (IIT) Delhi, New Delhi, in 1999.

He is a Professor with the Department of Electrical Engineering, Indian Institute of Technology Delhi. In 1997, he joined Aricent. He joined IIT Delhi as an Assistant Professor in 2005. He worked as an Associate Professor from 2010 to 2018. He is actively working in the area of image processing, signal processing, and communication. The areas include, object representation, tracking and classification, odometry, depth map generation, representation and rendering. He is also exploring vector sensor-based underwater acoustic communications, and performance issues in molecular communications.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:03:19 UTC from IEEE Xplore. Restrictions apply.

Journal of Entomology and Zoology Studies

Available online at www.entomoljournal.com

E-ISSN: 2320-7078
P-ISSN: 2349-6800
www.entomoljournal.com
JEZS 2020; 8(2): 1215-1220
© 2020 JEZS
Received: 08-01-2020
Accepted: 12-02-2020

Anurag Sanghmitra
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India

Anurag Aswal
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India

Ankita Maurya
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India

Apurv Arora
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India

Anuradha Saha
Department of Applied Sciences, Galgotias College of Engineering and Technology, 1, Knowledge Park, Phase 2, Greater Noida, Uttar Pradesh, India

Temesgen Hailu
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India

Rajinder K Gupta
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India

Corresponding Author:
Rajinder K Gupta
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, India

Chemical composition and thermoanalytical analysis of the nest of yellow wasp and mud dauber

Anurag Sanghmitra, Anurag Aswal, Ankita Maurya, Apurv Arora, Anuradha Saha, Temesgen Hailu and Rajinder K Gupta

Abstract
A wasp is a name given to groups of insects in the order *Hymenoptera* and suborder *Apocrita* It is different from a bee as well as an ant. Analysis of nests of yellow colored Wasps and Mud Dauber was carried out with the help of thermogravimetric and spectroscopic techniques. TGA analysis results showed the decomposition profile of the yellow paper wasp nest and mud dauber, indicated the percentage of weight loss with respect to temperature, that is similar to that of cellulose, which further demonstrate that cellulose is the fundamental group in the paper wasp structure which is responsible for adhesive nature and strength of the both type nest sample. Three different stages were observed on the TG curve. First stage starting from 25-250 ºC represents the loss of moisture present in the nest which is not chemically bonded. This stage accounts for 15% weight loss which shows there is a significant amount of moisture content in the outer part of the nest. Second stage is between 250-360 ºC which shows the decomposition of hemicellulose and formation of volatile products and char. Weight loss is approximately 40% of the total weight. Third stage accounts for 33% weight loss due to the thermal degradation of cellulose.
The FT-IR spectral analysis showed the presence of different functional groups. The GC-MS analysis of methanol extracts of paper wasp nest and mud dauber nest showed the presence of different types of bioactive compounds. Elemental analysis of samples showed the most prominent minerals are calcium (31.85ppm) followed by phosphorous (12.8 ppm) and magnesium (9.49 ppm) and different elements. Ash content was low as compared to the mud dauber nest; this indicates that it has higher organic content.

Keywords: Chemical composition, thermo-gravimetric analysis, nests of yellow wasps and mud dauber

Introduction
A wasp is a name given to groups of insects of the order *Hymenoptera* and suborder *Apocrita*. It is different from a bee as well as an ant. They usually form habitat around buildings, houses, trees and bridges. The black and yellow mud dauber, *Sceliphron caementarium*, is a common and distributed widely as solitary wasp that hunts spiders and builds characteristic mud nests for their offspring. It was found in muddy places where building materials were easily available [1]. In their nest construction, a mud dauber wasp uses clay as a material. This makes the mixture of nest adhesive and strong [2]. Clay soils are hydrous aluminium phyllosilicates, with different amounts of magnesium, iron, alkali, alkaline earth metals and other electro positive ions [3]. They collect mud, moisten it with their saliva, which may be functioning as binding material. The saliva derived material stick to the building and hence, vertically on the outside of the button-like structure which are normally rich in phosphorous. The nest of paper wasp was found among bunches of leaves in the tree branches. Structurally, it was oval shaped and attached to minute and string stalk [1]. Bioengineering applies engineering principles to the living systems and relates to biotechnology and medical science [5]. Biomathematics or computational biomedical is an interdisciplinary field which aims at modelling natural and biological processes using mathematical tools [6-8]. The aim of our work was to investigate the nature and chemical composition of yellow wasp & mud dauber nests.

Experimental method
Analysis of the nest of paper wasp and mud dauber sample was done by finding the chemical composition of the nest, which includes elemental analysis (ICP-OES), GC-MS and other

spectroscopic techniques such as FT-IR. Thermo gravimetric analysis of the nest of yellow wasp and mud dauber was analyzed using instrumental techniques like TGA and DSC.

FTIR

FTIR (Fourier Transform Infrared Spectroscopy) analysis was performed by using PerkinElmer spectrum version 10.5.3 to find out the functional groups present in a paper wasp nest and the mud dauber sample. The FT-IR instrument guides, radiation of wavelength of about 4000cm^{-1} to 400 cm^{-1} through our sample with some radiation absorbed and few passes from it.

GC-MS

GC-MS analysis of methanol extract of Paper Wasp and Mud Dauber Nest was done. The methanol extract of paper wasp nest and mud dauber nest were analysed by GC-MS (Agilent 5975B MSD) was used in the scan mode of (m/z 35-1050) to find out the amount of bio active compound. Gas chromatograph induced with mass spectroscopy is one of the widely used instruments used to identify as well quantify the blend of the complex mixtures with variable molecular weight can be measured with the highest accuracy and precision.

Elemental analysis (ICP-OES)

ICP-OES (PerkinElmer) was used to analyse the elemental composition of the sample. The procedure is based on the warm plasma driven by argon gas causing ionization of the sample.

TGA & DSC

TGA determines the quantity and the frequency of the weight variation of the samples against temperature and time in a controlled atmosphere (the purging of nitrogen gas) [4]. TGA of the sample is carried out by using TGA 4000 (PerkinElmer) in the atmosphere of nitrogen. It can be used to primarily investigate thermal strength and viscoelastic property of different nest sample. DSC is a thermo-analytical technique measuring the difference between heat taken from the sample and the reference.

The parameters used for mud dauber

Temperature: 30-250 °C
Heating Rate: 20 ˚ per minute

The parameters used for a paper wasp nest

Temperature: 30-350 ˚C
Heating rate: 20 ˚ per minute
Different nests, building material samples were ground into fine powder, kept in sterilized vials and labelled accordingly.

Ash Content

Determination of ash content in a polymer helps to measure the total amount of inorganic content that is minerals present in the polymer. The sample was heated at 550 ˚C for 4 hours in a Muffle furnace during which all the organic content decomposes and vaporize as carbon dioxide, water vapour and other gases. This process is used to estimate the number of residual solids (ash) in the polymer.
First, the sample was weighed then it is kept in the muffle furnace at 550 ˚C for 4 hours. Sample from the furnace was again weighed till it becomes a constant and final weight gave us the ash content.

Results and Discussion

Paper wasp nest sample results

TGA (Thermogravimetric analysis)

The thermogravimetric analysis of the yellow paper wasp nest was carried out and results were compared with the TGA curve of cellulose fibre obtained from the library. We also noticed different stages in which there was a weight loss with respect to temperature. Figure 1 & 2 shows the curve of thermal decomposition of the outer portion of a paper wasp nest. The experiment was performed in the range of 30-600 ˚C at a rate of 20 °C / minutes.

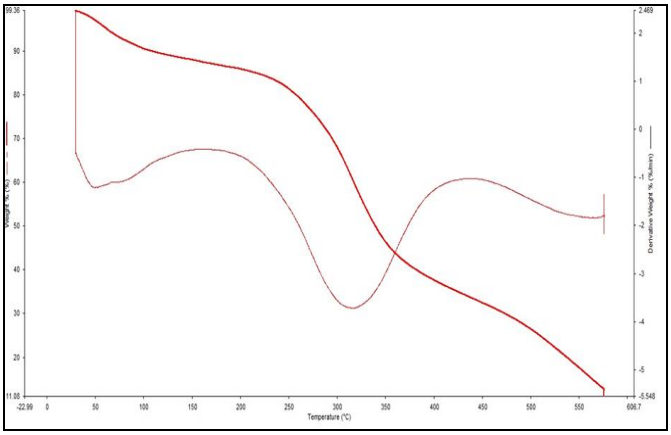


Fig 1: The weight loss with respect to temperature in the paper wasp nest sample

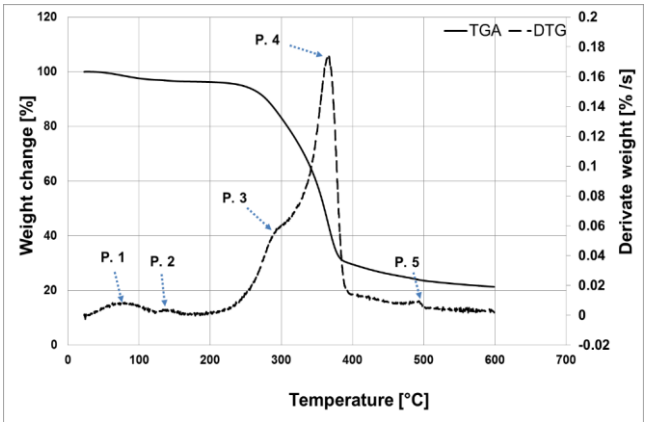


Fig 2: The weight change with respect to temperature in cellulose fibre

Table 1: Comparison between two curves

Stages	Yellow paper nest sample (degree Celsius)	Cellulose fiber (degree Celsius)
First stage	25-250	25-255
Second stage	250-360	255-380
Third stage	360-550	380-600

Three different stages were observed on the TG curve. First stage starting from 25-250 represents the loss of moisture present in the nest, which is not chemically bonded. This stage accounts for 15% weight loss which shows there is a significant amount of moisture content in the outer part of the nest. Second stage is between 250-360 which shows the decomposition of hemicellulose and formation of volatile products and char. Weight loss is approximately 40% of the total weight. Third stage accounts for 33% weight loss due to the thermal degradation of cellulose (Table 1, Table 2).

DSC (Differential Scanning Calorimetry)

DSC is the most widely accepted method to investigate the thermal properties of polymers (both natural and synthetic) such as melting point (Tm), glass transition temperature (Tg), crystallization, oxidation, decomposition stages (Figure 3).

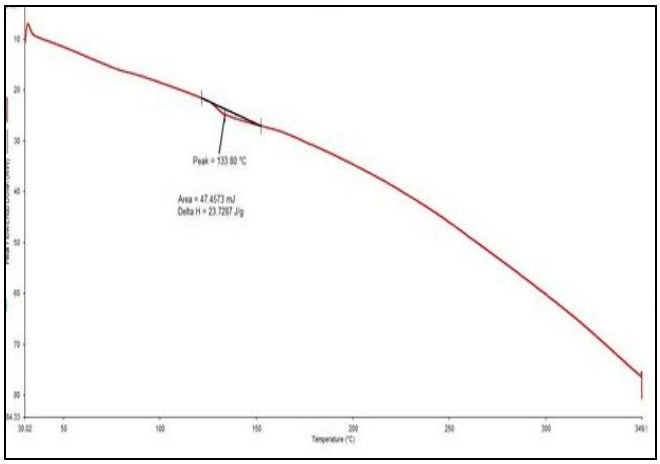


Fig 3: DSC Curve of Paper Wasps Nest Sample

Mud dauber wasp nest results

TGA (Thermogravimetric analysis)

We have carried out the thermo gravimetric analysis of this nest and compared its result with the TGA curve of soil obtained from the library. We have mentioned different stages in which weight loss is occurring with respect to temperature (Figure 4, 5).

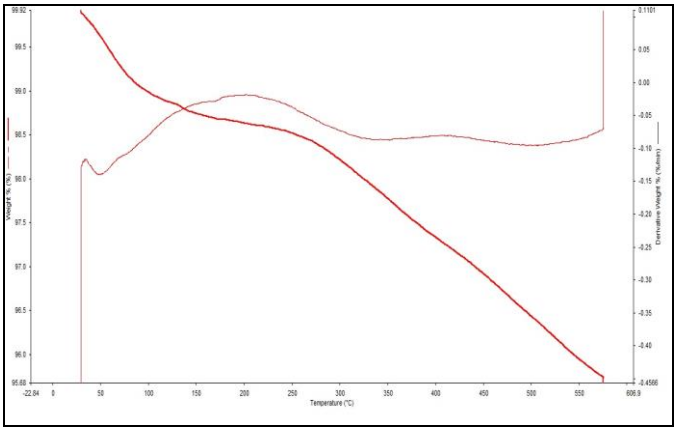


Fig 4: TGA curve of Mud Dauber Nest sample

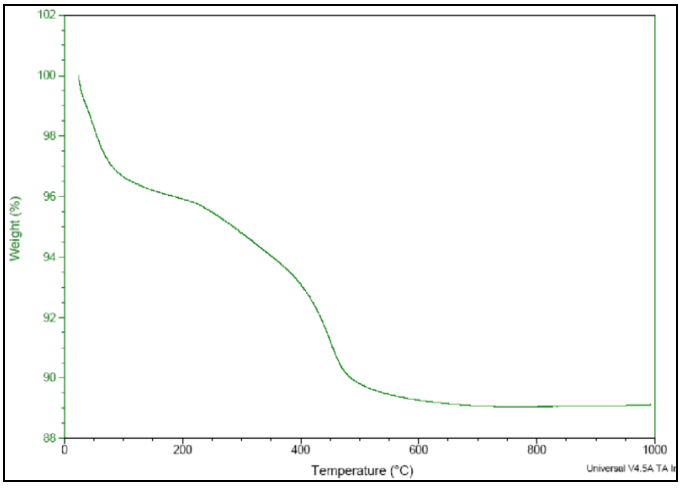


Fig 5: TGA curve of Soil Sample

Comparison between two Curves

Table 2: Comparison between Mud Dauber Nest and soil

Stages	Mud Dauber Nest	Soil
First Stage	25-80	25-90
Second Stage	80-290	90-250
Third Stage	290-550	250-450
Fourth Stage	-	>450

DSC (Differential Scanning Calorimetry)

DSC analysis was carried out to find out the thermal properties of the sample (Figure 6).

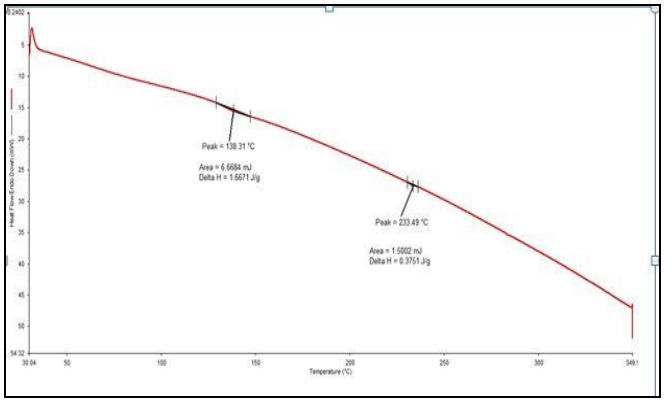


Fig 6: DSC curve of Mud Dauber Nest Sample

Fourier Transform Infrared (FTIR) analysis of Paper Wasp Nest

Transmission FTIR (Fourier Transform Infrared Spectroscopy) analysis was performed to find out the functional groups present in a paper wasp nest sample. Some of the prominent functional groups are listed in the table given below (Figure 7) (Table 3).

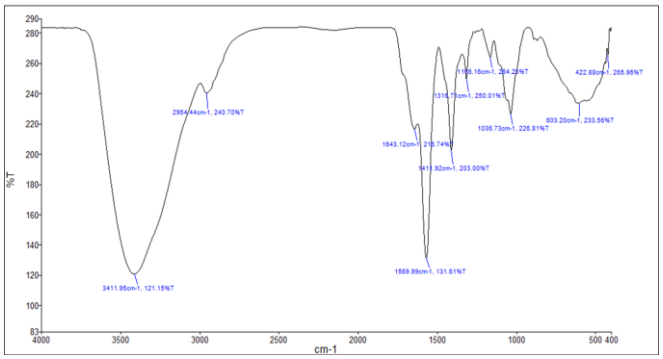


Fig 7: FT-IR spectrum of paper wasp nest sample

Table 3: The functional groups of different wave numbers

Peaks (Wave Number)	Functional Groups
3411.95	Hydroxyl
1643.12	Carbonyl groups
1559.99	Aromatic rings
1411.92	CH_2 and CH_3 bending vibration
1035.73	C-O stretching
603.20	CH_2-, CH_3- rocking

FTIR-ATR

The FTIR-ATR of the yellow paper wasp nest was performed and tissue paper and then compared their curves. This shows that our sample is cellulose based and could have some properties similar to that of tissue paper (Figure 8, 9)

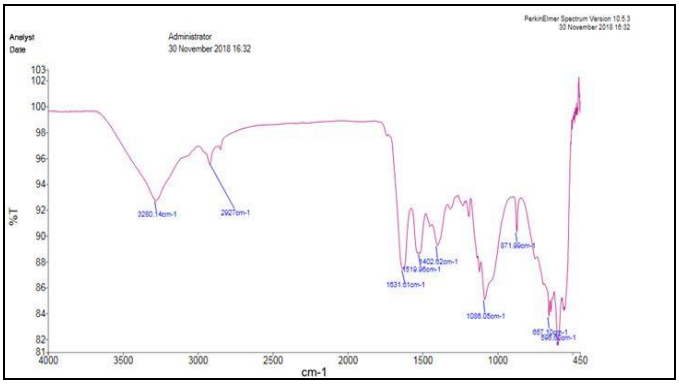


Fig 8: FT-IR ATR Curve Paper Wasp Nest

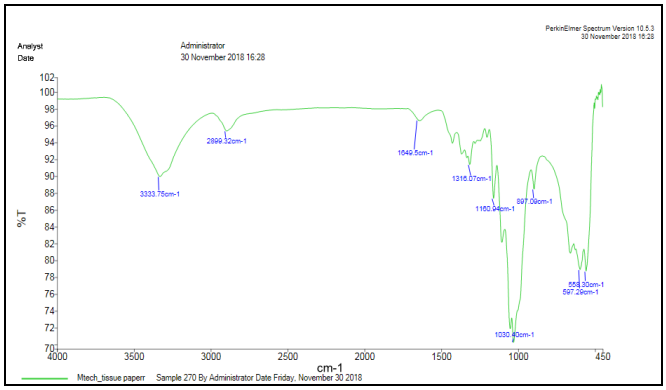


Fig 9: FT-IR ATR curve of Tissue Paper

FTIR Results of mud dauber nest sample

Transmission FTIR (Fourier Transform Infrared Spectroscopy) was performed to find out the functional groups present in the sample (Figure 10). Some of the prominent functional groups are listed in the Table 4.

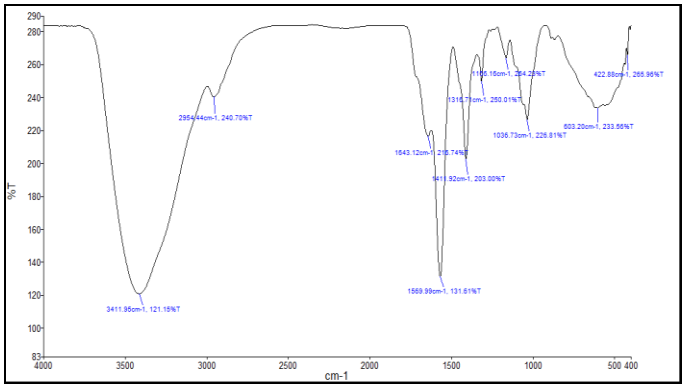


Fig 10: FT-IR curve of Mud Dauber Nest sample

Table 4: Peaks at Different Wave Number corresponding to Different Groups

Peaks (Wave Number)	Functional Groups
3405.93	A hydroxyl group (O-H stretching)
2925.72	Aliphatic SP_3 C-H stretching
2854.96	CH_2 stretching of the alkyl group
1622.55	Aromatic Carbonyl group
1432	CH_2 bending vibration of methylene
1319.30	CH_3 bending vibration of methyl
1028.37	C-O stretching of alcohol

GC-MS Analysis of methanol extract of paper wasp nest and mud dauber nest

The methanol extract of paper wasp nest and mud dauber nest were analysed by GC-MS; to give the gas chromatogram. The details of all the identified bioactive compounds present in nest samples are shown in Table 5 and Table 6.

Table 5: Chemical compounds identified in the methanol extract of a paper wasp nest

Peak No.	Name of compound	R. Time	Area %	Molecular formula	Molecular Weight
1	1H-imidazole, 2-ethyl-	12.709	0.32	$C_5H_8N_2$	96
2	2-Propyldecan-1-ol	12.966	0.62	$C_{13}H_{28}O$	200
3	Octadecane	14.251	1.42	$C_{18}H_{38}$	254
4	Delta. -terpineol. Beta. epoxide (. Alpha., alpha. -dimethyl-3-cyclohexyl	14.978	0.39	$C_{10}H_{18}O_2$	170
5	Cyclohexane ethanol, 4-ethenyl-4-methyl-. Beta. -methylene-3-(1-methyl	15.925	0.39	$C_{15}H_{24}O$	220
6	5-Azulenemethanol, 1,2,3,4,5,6,7,8-octahydro-. Alpha.,. Alpha., 3,8-tetramethyl-	16.271	2.41	$C_{15}H_{26}O$	222
7	Patchouli alcohol	16.542	4.76	$C_{15}H_{26}O$	222
8	Acetamide, n-(1-adamantan-1-ylpropyl)-2-(benzooxazol-2-ylsulfan	16.890	0.83	$C_{22}H_{28}N_2O_2S$	384
9	Acetyl cedrene	17.521	4.85	$C_{17}H_{26}O$	246
10	trans-. alpha. -Bergamotene	17.694	0.94		
11	1,6-Methanonaphthalen-1(2h)-ol, octahydro-4,8a,9,9-tetramethyl-	17.774	1.90	$C_{15}H_{26}O$	222
12	2 (3H) -Naphthalenone, 4,4a,5,6,7,8-hexahydro-4,4a-dimethyl-6-(1-methylethylidene)-	18.364	1.57	$C_{15}H_{22}O$	218
13	Zizanyl acetate	18.454	1.13	$C_{17}H_{26}O_2$	262
14	Hexadecanoic acid, methyl ester	19.016	5.22	$C_{17}H_{34}O_2$	270
15	1,2-Benzenedicarboxylic acid, dibutyl ester	19.369	18.08	$C_{14}H_{18}O_6$	282
16	1,9-Cyclohexadecadiene	20.659	1.33	$C_{16}H_{28}$	220
17	8,11,14-Docosatrienoic acid, methyl ester	20.715	4.50	$C_{23}H_{40}O_2$	348
18	Phytol	20.818	1.01	$C_{20}H_{40}O$	296
19	Heneicosanoic acid, methyl ester	20.948	1.71	$C_{22}H_{44}O_2$	340
20	Hexane, 2,3,4-trimethyl-	22.508	0.35	C_9H_{20}	128
21	Undecane, 2,4-dimethyl-	24.787	0.59	$C_{13}H_{28}$	184
22	1,2-Benzenedicarboxylic acid	25.291	17.08	$C_{24}H_{38}O_4$	390
23	Oxalic acid, 6-ethyloct-3-yl heptyl ester	28.467	2.41	$C_{19}H_{36}O_4$	328
24	Squalene	29.372	0.79	$C_{30}H_{50}$	410
25	Octadecane, 1-chloro-	30.423	1.09	$C_{18}H_{37}Cl$	288
26	Nonadecane	30.825	1.10	$C_{19}H_{40}$	268

The main constituents of the methanol extract of paper wasp nest were Phytol (1.01%), Nonadecane (1.1%), Zizanyl acetate (1.13%), 1,9-Cyclohexadecadiene (1.33%), Octadecane (1.42%), 2(3H)-Naphthalenone, 4,4a,5,6,7,8-hexahydro-4,4a-dimethyl-6-(1-methylethylidene (1.57%), 1,6-Methanonaphthalen-1(2h)-ol, octahydro-4,8a,9,9-tetramethyl-, (1.9%), 5-Azulenemethanol, 1,2,3,4,5,6,7,8-octahydro- alpha, alpha,3,8-tetramethyl-(2.41%), Oxalic acid, 6-ethyloct-3-yl heptyl ester (2.41%), Heneicosanoic acid, methyl ester (1.71%), 1,2-Benzenedicarboxylic acid (17.08%), 8,11,14-Docosatrienoic acid, methyl ester (4.5%), Patchouli alcohol (4.76%), Acetyl cedrene (4.85%), Hexadecanoic acid, methyl ester (5.22%), 1, 2-Benzenedicarboxylic acid, dibutyl ester (18.08%) (Table 6). Hexadecanoic acid methyl ester, are reported to have anti-inflammatory, cancer preventive, hepatoprotective, antiarthritic, and anticoronary properties [9].

Table 6: The major bioactive components identified in methanol extract of mud dauber nest

Peak No.	Name of compound	R. Time	Area %	Molecular formula	Molecular Weight
1	Undecane, 4,7-dimethyl-	14.250	1.34	$C_{13}H_{28}$	184
2	3-Methoxy-5-propylphenol	14.475	3.55	$C_{10}H_{14}O_2$	166
3	Flopropione	14.899	1.60	$C_9H_{10}O_4$	182
4	1,6-Methanonaphthalen-1(2h)-ol, octahydro-4,8a,9,9-tetramethyl-	16.546	1.51	$C_{15}H_{26}O$	222
5	Benzoic acid, 2-hydroxy-4-methoxy-3,5,6-trimethyl-, methyl ester	16.872	1.10	$C_{12}H_{16}O_4$	224
6	1,6-Methanonaphthalen-1(2h)-ol, octahydro-4,8a,9,9-tetramethyl-	17.775	0.69	$C_{15}H_{26}O$	222
7	Neophytadiene	18.092	1.11	$C_{20}H_{38}$	278
8	Pentadecanoic acid, 14-methyl-, methyl ester	19.008	3.91	$C_{17}H_{34}O_2$	270
9	Dibutyl phthalate	19.362	17.99	$C_{16}H_{22}O_4$	278
10	9,12-Octadecadienoic acid, methyl ester	20.649	1.68	$C_{19}H_{34}O_2$	294
11	7-Hexadecenoic acid, methyl ester, (Z)-	20.707	4.26	$C_{17}H_{32}O_2$	268
12	Phytol	20.810	1.17	$C_{20}H_{40}O$	296
13	Octadecanoic acid, methyl ester	20.940	1.68	$C_{19}H_{38}O_2$	298
14	1H-imidazole, 4,5-dihydro-2-(phenylmethyl)-	22.780	4.58	$C_{10}H_{12}N_2$	160
15	Hexane, 3, 3-dimethyl-	24.781	0.43	C_8H_{18}	114
16	1, 2-Benzenedicarboxylic acid	25.283	24.61	$C_{24}H_{38}O_4$	390
17	Hexane, 3, 3, 4-trimethyl-	28.017	0.38	C_9H_{20}	128
18	1-Dodecanol, ethoxy-	28.433	3.43	$C_{14}H_{30}O_2$	230
19	Squalene	29.364	0.90	$C_{30}H_{50}$	410
20	2-Bromotetradecane	30.418	1.79	$C_{14}H_{29}Br$	276
21	Dotriacontane, 1-iodo-	30.810	1.28	$C32H_{65}I$	576
22	Vitamin E	33.983	0.92	$C_{29}H50O_2$	430

The main chemical constituents of the methanol extract of mud dauber nest were:- Vitamin E (0.92%), Neophytadiene (1.11%), Phytol (1.17%), Dotriacontane, 1-iodo-(1.28%), Undecane, 4,7-dimethyl (1.34%), 1, 6-Methanonaphthalen-1(2H)-ol, octahydro-4,8A,9,9-tetramethyl-, (1.51%), Flopropione(1.60%), Octadecanoic acid, methyl ester (1.68%),9,12-Octadecadienoic acid, methyl ester (1.68%), 2-Bromotetradecane (1.79%), 1-Dodecanol, ethoxy-(3.43%), 3-Methoxy-5-propylphenol (3.55%), Pentadecanoic acid, 14-methyl-, methyl ester (3.91%), 1H-imidazole, 4,5-dihydro-2-(phenylmethyl)- (4.58%), 7-Hexadecenoic acid, methyl ester, (Z)- (4.26%), Dibutyl phthalate (17.99%), 1,2-Benzenedicarboxylic acid (24.61%).

Vitamin E is a fat-soluble chiral compound which has high antioxidant potential. It has eight stereoisomers: α, β, γ, δ tocopherol and α, β, γ, δ tocotrienol (with double bonds in side chain) in which α-Tocopherol is the most bioactive form, found in humans [10]. Since, it is fat-soluble, α-tocopherol protects cell membranes from damage by free radicals. Its antioxidant capacity mainly resides in the protection against lipid peroxidation.

Phytol is an acyclic diterpene alcohol that can be used as a precursor for the manufacture of synthetic forms of vitamin E [11] and vitamin K1 [12]. In ruminants, the gut fermentation of ingested plant materials liberates phytol, a constituent of chlorophyll, which is then converted to phytanic acid and stored in fats [13] in shark liver it yields pristane.

Inductively Coupled Plasma (ICP)

Various mineral were find out from the nest sample described in Table 7.

Table 7: Various Metal Ions Present in Mud Dauber Nest Sample

Metal Found	Amount (PPM)
Magnesium	9.49
Calcium	31.85
Titanium	0.70
Copper	0.12
Nickel	0.02
Boron	0.08
Aluminium	8.81
Silicon	0.73
Barium	0.13
Fluoride	8.94
Phosphorous	12.8
Zinc	0.35
Lead	0.06
Vanadium	0.02
Manganese	0.3
Cobalt	0.02
Chromium	0.07

Ash Content

Determination of ash content in a polymer helps to measure total amount of inorganic content that is minerals present in the polymer. The sample was heated at 550 °C for 4 hours in

a Muffle furnace during which all the organic content decomposes and vaporizes as carbon dioxide, water vapour and other gases.

Final weight (ash content) = 0.9061
Initial weight of the sample = 1.0064 g
The difference in initial and final weights = 0.1003g

The difference in initial and final weight is low as compared to the yellow paper wasp nest, this implies that it has more inorganic content.

In case of mud dauber
Initial weight of the sample = 1.0040g Final weight (ash content) = 0.3135g
Difference in initial and final weights = 0.6905g
Difference in initial and final weights is more as compared to mud dauber; this implies it has more organic content.

Conclusion
TGA and DSC analysis of nest samples revealed that there was weight loss with respect to temperature, heat transfer and showed thermal stability. The GC-MS, FT-IR and ICP-OES showed the presence of different types of the chemical compound, functional groups and elements such as hydroxyl, phenolic, carbonyl, calcium, magnesium, titanium, copper, manganese, silicon, phosphorus, zinc, cobalt and aluminium in varying quantities that are responsible for the adhesive nature and strength of the Mud Dauber wasp and yellow wasp nest.

Acknowledgement
Author would like to express gratitude towards Delhi Technological University for the financial support.

Conflict of interest statement: None

References
1. Dress BM, Jackman J, Mud Daubers. A field to Texas insect. Texas A and M University, 1999.
2. Carpenter JM. The phylogenetic relationship and natural classification of the Vespodea (*Hymenoptera*). Systemic Entomology. 1982; 7:11-38.
3. Velde B. Origin and Mineralogy of clays: clays and the environment springer verlag. Berlin Iherdelberg, N.Y, 1995.
4. Perveen F, Muzafar S. Nest Architectural patterns by three wasp species with reference to their behaviour. International Journal of Insect Science. 2020; 5(1).
5. Carpenter JM. A synonymic generic checklist of the *Eumeninae* (*Hymenoptera*; *Vespidae)*. Psyche. 1986; 93:61-90.
6. Bailing SW. Summary of recommendation of AIPEA nomenclature committee on clay minerals, American Mineralogist. 1980; 65:1-7.
7. Iwata K. Habits of four species of *Odynerus (Ancistrocerus*) in Japan, Terithredo. 1938; 2:19-32
8. Smith A. Life strategy and mortality factors of *Sceliphron laetum*(Smith) (*Hymenoptera*) in Australia. Australian Journal of Ecology. 1979; 4:181-186.
9. Surender S, Vinod N, Sweety J, Gupta YK. Evaluation of anti-inflammatory activity of plant lipids containing α-linolenic acid. Indian Journal of Experimental Biology. 2008; 46(6):453-456.
10. (a) Nguyen LA, He H, Pham-Huy C. International Journal of Biomedical Sciences. 2006; 2:85-100. (b) Mayo Clinic Medical Information. Drugs and Supplements. Vitamin E. 2005.
11. Thomas N. Synthesis of Vitamin E. In Litwack, Gerald. Vitamin E. Vitamins and Hormones. Vitamins & Hormones. 2007; 76:155-202.
12. Alison D, Richard P, Humphries M. Abell A. The Synthesis of Naturally Occurring Vitamin K and Vitamin K Analogues. Current Organic Chemistry. 2003; **7**(16):1625–34.
13. Brink DM, Wanders RJA. "Phytanic acid: Production from phytol, its breakdown and role in human disease". Cellular and Molecular Life Sciences. 2006; 63(15):1752-65.

Classification of Cuisines from Sequentially Structured Recipes

Tript Sharma
Department of Mechanical Engineering
Delhi Technological University
Delhi, India
triptsharma22@gmail.com

Utkarsh Upadhyay
Department of Electronics Engineering
Jamia Millia Islamia University
Delhi, India
utkarshdhy@gmail.com

Ganesh Bagler
Department of Computational Biology
Indraprastha Institute of Information Technology Delhi (IIIT-Delhi)
bagler@iiitd.ac.in

arXiv:2004.14165v1 [cs.CL] 26 Apr 2020

***Abstract*—Cultures across the world are distinguished by the idiosyncratic patterns in their cuisines. These cuisines are characterized in terms of their substructures such as ingredients, cooking processes and utensils. A complex fusion of these substructures intrinsic to a region defines the identity of a cuisine. Accurate classification of cuisines based on their culinary features is an outstanding problem and has hitherto been attempted to solve by accounting for ingredients of a recipe as features. Previous studies have attempted cuisine classification by using unstructured recipes without accounting for details of cooking techniques. In reality, the cooking processes/techniques and their order are highly significant for the recipe's structure and hence for its classification. In this article, we have implemented a range of classification techniques by accounting for this information on the RecipeDB dataset containing sequential data on recipes. The state-of-the-art RoBERTa model presented the highest accuracy of 73.30% among a range of classification models from Logistic Regression and Naive Bayes to LSTMs and Transformers.**

***Index Terms*—Recurrent Neural Networks, Transformers, Classification, Sequential Recipes**

I. Introduction

Cuisines represent the culinary imprint of cultures. The structure of cuisines is shaped by composition of their recipes. Increasing availability of data of cuisines has led to data-driven explorations of cuisines such as food pairing, culinary fingerprinting and cuisine classification. Classification of cuisines is an interesting problem with applications for recipe recommendation and generation of novel recipes. Hitherto, cuisine classification has been attempted using ingredients of recipes as a feature. These approaches have overlooked key factors such as cooking techniques and their order which clearly form a key aspect of recipes.

We account for the loss of these information by including these details for the cuisine classification problem. As opposed to treating recipes as an itemset, we propose a methodology that views the problem as a Text Classification (TC) problem. TC refers to annotating text with one category or other based on content words and their collocations. In this article, we propose different architectures for cuisine classification on sequential datasets.

II. Related Work

Availability of detailed recipe data has evoked the interest in recipe recommendation and generation. In the past, classification of cuisines has been attempted on the basis of various factors such as time, ethnicity and place by creating six different feature stylistic sets from the data document [1]. Han Su et. al. [2] have worked on cuisine identification by using ingredients used in recipes as a basis. By identifying ingredients as features, they could provide insights on cuisine similarity. A personalised cuisine recommendation system based on user's preferences has also been proposed [3] where user's preferences are derived from their browsing activities.

Support Vector Machines [4] and several other machine learning techniques have also been implemented towards generation of a cuisine. Recently, a study on classification of cuisine on the basis of the recipe's ingredients [5] suggested a detailed relation between a recipe and its ingredients.

In this article, we propose that, beyond ingredients, even the processes and utensils involved in cooking a recipe and their order of occurrence can provide significant insights into the cuisine. We have used RecipeDB (site) [6] dataset. And to test our hypothesis, we perform classification on the dataset using several machine learning techniques, neural networks and transformers.

III. Dataset

RecipeDB was used as the source of structured data on recipes for the analysis. The dataset contains 118,071 recipes obtained from sources like AllRecipes, Epicurious Food Network, and TarlaDalal. The dataset consists of 26 cuisines as shown in Table II. Moreover, it contains an aggregation of 20280 unique ingredients, 256 unique processes and 69 unique utensils. Sample dataset of RecipeDB can be seen in Table I. Our analysis involves the following substructures of cooking recipes pertaining to traditional recipes, namely, recipes, ingredients, processes and utensils.

RecipeDB consists of ingredients, processes and utensils mined from unstructured recipe scraped from the above mentioned resources. The substructures for the recipes are mined in a sequential fashion depending upon the order in which they are used in preparing the dish. The dataset is highly sparse with a sparsity ratio of 99.50%. Out of the 20,400 distinct ingredients obtained, 11738 occur at most in one recipe such as 'lasagna noodle wheat', while 'add' appeared 1,88,004 number of times. The corresponding cumulative frequency table for the

TABLE I
SAMPLE DATASET FROM RECIPDB

Recipe ID	Continent	Cuisine	Recipe
2610	African	Middle Eastern	['water', 'red lentil', 'rom tomato', ,'smooth', 'stir', 'heat']
3957	Asian	Southeast Asian	['olive oil', 'onion', 'garlic', 'ginger', , 'stir', 'add', 'cook ', 'season', 'garnish', 'pot']
4153	Asian	Indian Subcontinent	['coconut milk', 'milk', 'white sugar', 'basmati rice', , 'stir', 'cook', 'saucepan', 'bowl']
79897	Latin American	Mexican	['beef', 'chunky salsa', 'mushroom', 'garlic', , 'heat', 'simmer', 'serve', 'skillet']
138976	European	Deutschland	['oven buttermilk biscuit', 'onion', 'cream', ..., 'spread', 'sprinkle', 'bake', 'pan']
149191	North American	Canadian	['raisin', 'fig', 'water', 'date', 'butter', , 'chill', 'cut', 'bowl', 'processor', 'pan']

TABLE II
DATASET INFORMATION

Cuisine	Number of Recipes	Cuisine	Number of Recipes
Australian	5823	Japanese	2041
Belgian	1060	Korean	668
Canadian	6700	Mexican	14463
Caribbean	3026	Middle Eastern	3905
Central American	460	Northern Africa	1611
Chinese and Mongolian	5896	Rest Africa	2740
Deutschland	4323	Scandinavian	2811
Eastern European	2503	South American	7176
French	6381	Southeast Asian	1940
Greek	4185	Spanish and Portuguese	2844
Indian Subcontinent	6464	Thai	2605
Irish	2532	UK	4401
Italian	16582	US	5031

TABLE III
FREQUENCY DISTRIBUTION OF FEATURES

Number of Features	Frequency	Number of Features	Frequency
304	>1000	<2	11738
106	>5000	<3	14015
57	>10000	<4	15002
43	>15000	<5	15620
34	>20000	<6	16073
24	>25000	<7	16394
19	>30000	<8	16627
17	>35000	<10	17016
13	>40000	<15	17314
12	>45000	<20	17519

number of items shown in Table III, represents the nature of the dataset.

IV. PREPROCESSING

Before performing cuisine classification on RecipeDB data, preprocessing was implemented on structured and sequential lists of ingredients, processes and utensils. Furthermore, the digits or symbols were omitted from the items to only keep words, thereby reducing the noise in this highly sparse dataset. The preprocessing further involved tokenization followed by lemmatization of the dataset, resulting in 20,400 distinct entities.

The data is further processed to conform to the classification model requirements. Since an individual word itself doesn't impart any semantic or syntactic significance to the classification models that require quantified features as inputs, each item was translated to vectors using two techniques, namely, TF-IDF vectorization and word embedding. Depending upon the preprocessing method used, the models employed can be broadly classified into two categories: sequential models and statistical models. If the dataset is sequential, it is evident that sequential models like RNNs work better while for non sequential datasets, models like Logistic Regression, SVM, etc. perform better.

Word embeddings are essentially word representation as vectors such that semantically similar words have similar vectors whereas TF-IDF vectorization method observes the sequence of items as distinct words. Thus, TF-IDF vectors don't preserve the sequential nature of the data. Yet, we used TF-IDF technique because of its weighted function which reduces the effect of high frequency yet less meaningful words and provides a good analytical cause.

V. CLASSIFICATION

Classifying the recipes region/cuisine-wise, based on the elements involved in cooking a recipe is a major and the most important part of our analysis. The analysis treats recipes either as a sequential or as an unordered set of items. Many state-of-the-art machine learning models with TF-IDF vector inputs, such as Naive Bayes, Logistic Regression, Ensemble models along with boosting and Support Vector Machines, were tested for text classifications. Also, sequence models such as Recurrent Neural Networks and state-of-the-art NLP transformers such as BERT and RoBERTa were tested on our dataset to analyse the 'sequential nature' of the dataset. We will further discuss the implementations of these classifiers in the 'Experiments' section.

A. Naive Bayes

Naive Bayes (NB) classifier is probabilistic in nature. It is based on the supposition that all the features are independent and autonomous. NB selects the label which maximizes the posterior probability:

$$P(C_k|x) = P(C_k) * P(x|C_k)/P(x) \quad (1)$$

while the naive supposition is:

$$P(x_i|x_{i+1},,x_n,C_k) = P(X_i|C_k) \quad (2)$$

In spite of the fact that the naive supposition is false most of the time, NB gave extremely competitive results with respect to other classifiers.

B. Logistic Regression

Logistic Regression (LG) is the most used and a fundamental classifier. LG is also probabilistic in nature. It is based on the following Sigmoidal equation:

$$S(x) = 1/(1+e^{-x}) \quad (3)$$

LG treats the problem as a generalized linear regression model, which can be expressed as:

$$f(k,i) = \beta_{0,k} + \beta_{1,k}x_{1,i} + \beta_{2,k}x_{2,i} + \cdots + \beta_{M,k}x_{M,i}, \quad (4)$$

Here, for our multi-class classification problem LG is trained on a one-vs-rest scheme. Similar steps were followed in a previous research on a different dataset [5] where LG presented with the best results. This hold true in our cse as well in comparison with other baseline models.

C. Support Vector Machine

Support Vector Machines (SVMs) have been demonstrated to be having the best performance when working with textual data [4]. For classification, SVMs require translation of a multi-class classification problem to binary classification. For this, the One-vs-All approach was used. Single classifier per class was trained with the training set belonging to that class annotated as positive while the rest of the samples as negative. A strong real-valued confidence score along with a class label, by the base classifiers is required for the decision. SVM then searches for the two best-fit parallel hyperplanes which separates the two classes of data, so that they are farthest from each other.

D. Random Forest with Boosting

Random forest (RF) is a bagging decision tree approach [7]. When used as a classifier, it might not perform that well when working with a small number of features. But given that our problem is characterized with a large number of features, techniques such as RF with AdaBoost can turn out to be a good text classifier.

E. RNN (LSTM)

Recurrent Neural Networks follow a temporal or sequential connection between nodes of a layer [8]. Therefore, they are an upgrade to the conventional neural networks which consider mutual independence among the sequential inputs. Furthermore, RNNs contain an internal 'memory' and hence making them suitable for remembering previous inputs. Therefore, the characteristics of RNNs align with our problem.

We employed a state-of-the-art RNN, the Long-Short Term Memory based neural network (LSTM) [9]. LSTMs are more complex than simple RNNs as they involve a cell- and gate-like input, output and forget gate. Using these gates it controls the flow of information through the temporal dimension. It decides whether any piece of information is significant for the broader or immediate goal, or should it be removed from the memory. On account of this significant characteristic, we employed a simple 2-layer LSTM.

F. Transformer

RNNs go through words in a temporal fashion and if the sequence is long as in case of RecipeDB, the model tends to forget the crucial features of sequentially distant features. In order to overcome this limitation, attention based transformers were developed. Transformers [10] are the NLP models which are used to boost the speed of attention based models by enabling parallelization. They completely eliminate the recurrence with self attention to establish relationship between input and output, thus making them suitable for multiple language processing applications.

We employed BERT-base [11] and RoBERTa [12] models on RecipeDB. Both models perform bidirectional encoding implementing transformers after pre-training with the exception that RoBERTa is trained differently. RoBERTa was trained on longer sequences for more training steps than BERT.

VI. Experiments

For the purpose of cuisine classification we implemented different machine learning models. We tested the accuracy of the models on the RecipeDB dataset to validate the results obtained. The data was divided into 7:1:2 ratio to obtain training, validation and testing datasets respectively. Therefore out of 1,18,071 recipes training, validation and testing datasets consist of 82,650, 12,021 and 23,380 recipes respectively.

Since recipes were represented as sequences of ingredients, processes and utensils all concatenated together, long sequence were generated. The sequences were pre-processed differently for statistical models and sequential models as described in Section IV. By feeding the features obtained as input to the classifiers mentioned in Section V yielded results shown in Table IV. The corresponding code and relevant files are present in the GitHub repository: https://github.com/cosylabiiit/cuisine-classification.

Among the various statistical models that were implemented for cuisine classification, Logistic Regression performed the best, but with an accuracy of only 57.70%. These models learn the frequency of occurrence of an ingredient or process or utensil to obtain the features unique to a cuisine instead of treating the recipes as an interrelationship among these items. Furthermore, since the dataset is sparse the models couldn't fit better, leading to high bias.

Owing to the temporal relationship among the items in recipes, we observed that sequential models perform better than the statistical models. However, LSTM model gave a lower accuracy than Logistic Regression and Linear kernel SVM used in [5]. The lower accuracy is justified as the model is among the most simplistic models in the recurrent neural

TABLE IV
PERFORMANCE METRICS OF APPLIED MODELS

Dataset	Performance Metric	LogReg	Naive Bayes	SVM (linear)	Random Forest	LSTM	Transformer	
							BERT	RoBERTa
RecipeDB	Accuracy	57.70	51.64	56.60	50.37	53.61	68.71	73.30
	Loss	1.51	7.14	2.97	2.32	1.65	0.21	0.10
	Precision	0.56	0.50	0.54	0.48	0.53	0.58	0.67
	Recall	0.57	0.51	0.56	0.50	0.54	0.60	0.71
	F1 Score	0.56	0.50	0.54	0.49	0.53	0.57	0.69

network class. Furthermore while comparing the LSTM model with Transformers, the sequences are treated differently i.e. LSTMs consider left to right sequence order unlike the bi-directional check in Transformers. Moreover, despite having better memory logic than vanilla RNNs, LSTMs are limited by the number of words in the sequence which further reduces the accuracy.

The limitations of LSTMs have been overcome in Transformers as explained earlier which resulted in the optimal accuracy of 73.3% and a loss of 0.10 on the RecipeDB dataset. Hence, the model is able to predict the class with least errors on minimum number of datasets among the models tested. The model presents a high average precision, recall and F1 score values representing its ability for cuisine classification.

VII. CONCLUSIONS

This article investigates different approaches for cuisine classification as a synthesis of ingredients, processes and utensils inherent to a cuisine. It also examines the effect of temporal relationships among the features to fingerprint the worldwide cuisines with the state-of-the-art RoBERTa model giving optimal results for the problem. Thus, we present a strategy to treat recipes as chains of events that are similar for a region and simultaneously contrasting from others to some extent to enable classification. Further, this articles has raised issues that can help optimise the results in different computational contexts such as recipe generation and recipe recommendation.

Apart from this, the article also raises some new research questions relating to cuisine classification. While our analysis considered for the sequential nature of recipes, the relationship among the three substructures remains unaccounted. Moreover, what features aid or hinder the classification of a recipe which could help one to uniquely distinguish between the cuisines? While maintaining the sequential nature of the recipes, redundant features were not removed. Hence, future analysis needs to identify the effect induced by these features on the classification accuracy of the models. Furthermore the imbalance among the classes affects the cuisine prediction accuracy of the classifiers. This can be reduced by ignoring the low frequency classes but would lead to a limited exploration of the world cuisines. This trade-off presents as a dilemma in this analysis.

We believe this article adds another dimension to the existing body of research on cuisine classification. This is of value for cuisine classification of unknown recipes and also aids in identifying salient features intrinsic to a cuisine.

VIII. ACKNOWLEDGEMENT

G.B. thanks the Indraprastha Institute of Information Technology (IIIT-Delhi) for providing computational facilities and support. T.S and U.U. are Research Interns in Dr. Bagler's lab (Complex Systems Laboratory) at the Center for Computational Biology and thank IIIT-Delhi for the support.

REFERENCES

[1] Y. HaCohen-Kerner, H. Beck, E. Yehudai, M. Rosenstein, and D. Mughaz, "Cuisine: Classification using stylistic feature sets and/or name-based feature sets," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 8, pp. 1644–1657, 2010.

[2] H. Su, T.-W. Lin, C.-T. Li, M.-K. Shan, and J. Chang, "Automatic recipe cuisine classification by ingredients," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication*. ACM, 2014, pp. 565–570.

[3] M. Ueda, M. Takahata, and S. Nakajima, "Recipe recommendation method based on users food preferences," in *Proceedings of the IADIS International Conference on e-Society*, 2011, pp. 591–594.

[4] J. Weston and C. Watkins, "Multi-class support vector machines," Citeseer, Tech. Rep., 1998.

[5] J. Naik and V. Polamreddi, "Cuisine classification and recipe generation," *Online: http://cs229. stanford. edu/proj2015/233 report. pdf*, 2015.

[6] D. Batra, N. Diwan, U. Upadhyay, J. S. Kalra, T. Sharma, A. K. Sharma, D. Khanna, J. S. Marwah, S. Kalathil, N. Singh *et al.*, "Recipedb: A resource for exploring recipes," *Available at SSRN 3482237*, 2019.

[7] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[8] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh annual conference of the international speech communication association*, 2010.

[9] M. Sundermeyer, R. Schlüter, and H. Ney, "Lstm neural networks for language modeling," in *Thirteenth annual conference of the international speech communication association*, 2012.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

This figure "Normalized_Model_Accuracy.png" is available in "png" format from:

http://arxiv.org/ps/2004.14165v1

This figure "feat.png" is available in "png" format from:

http://arxiv.org/ps/2004.14165v1

This figure "feature.png" is available in "png" format from:

http://arxiv.org/ps/2004.14165v1

This figure "fig1.png" is available in "png" format from:

http://arxiv.org/ps/2004.14165v1

This figure "final_edit.png" is available in "png" format from:

http://arxiv.org/ps/2004.14165v1

This figure "flow.png" is available in "png" format from:

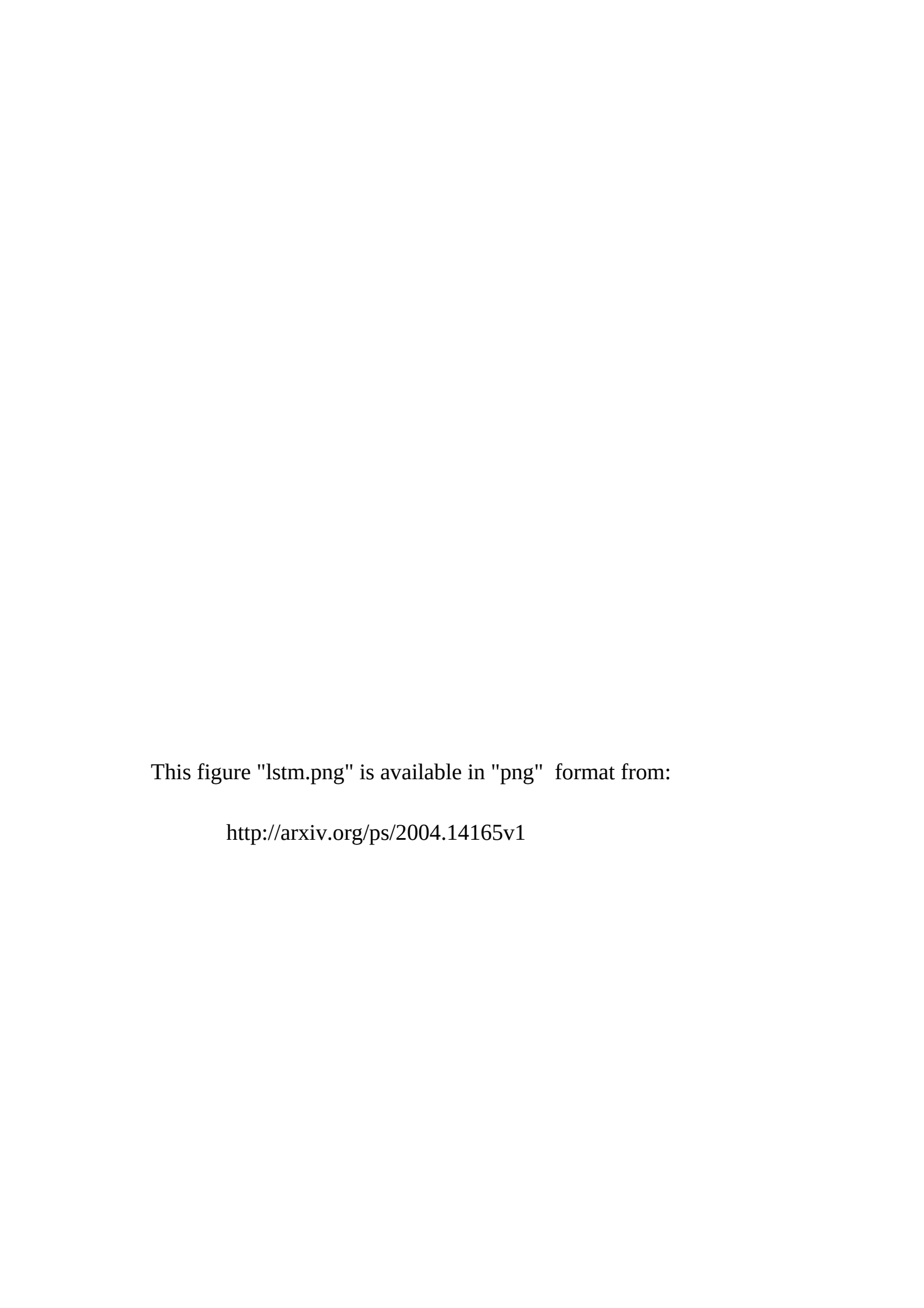
http://arxiv.org/ps/2004.14165v1

This figure "loss_training.png" is available in "png" format from:

http://arxiv.org/ps/2004.14165v1

This figure "loss_val.png" is available in "png" format from:

http://arxiv.org/ps/2004.14165v1


This figure "lstm.png" is available in "png" format from:
http://arxiv.org/ps/2004.14165v1

iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5088

www.ijraset.com

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Design of Tall Chimney for Thermal Power Plant

Karan Dhall
Delhi College of engineering, India

Abstract: ***This research paper presents the Study of along and across wind effects on150m high single flue RCC chimney for Nardana, Maharastra, India where the Basic Wind speed is taken from Figure 1 of IS 875 Part 3 which is 39m/s.***
Index Terms: Reinforced concrete chimney, structural designing of tall chimney, wind analysis of tall chimney, earthquake analysis and design of tall structures.

I. INTRODUCTION

This article was primarily conceptualized to understand behavior of tall structures to wind, in our case it is a tall chimney, 150 meter in height which makes it susceptible to wind forces, along with wind, the research paper also calculates earthquake forces on the chimney.

Wind is essentially the large-scale movement of free air due to thermal currents. It plays an important role in design of tall structures because it exerts static and dynamic loads whose effects on a slender structure, such as a chimney are significant. The wind load exerted at any point on a chimney can be considered as the sum of quasi-static and a dynamic load component.

In case of calculating the forces Simplified method as per as per IS code 875 is used, in order to study the structure completely with random and small variations Random Response method is applied. Random response method also provides us detailed analysis which is much effective than Simplified Method for result evaluations.

II. THE BEHAVIOR OF TALL CHIMNEY IN WIND

A. Along-wind Effects and Across-wind Effects

In a tall freestanding structure like chimneys, wind is generally a governing force. The effect of wind on these tall structures can be separated into two components, known as

along-wind effect

across-wind effect

Along-wind loads are caused by ‘drag’ component of wind force on the chimney, whereas the across-wind loads are caused by the resultant ‘lift’ component. The former is accompanied by the ‘gust buffeting’ causing a dynamic response in the direction of the mean flow, whereas the second is associated with the phenomenon of ‘vortex shedding’ which causes the chimney to oscillate in the direction perpendicular to the direction of wind flow. Evaluation of wind effects therefore includes the estimation of these two types of loads.

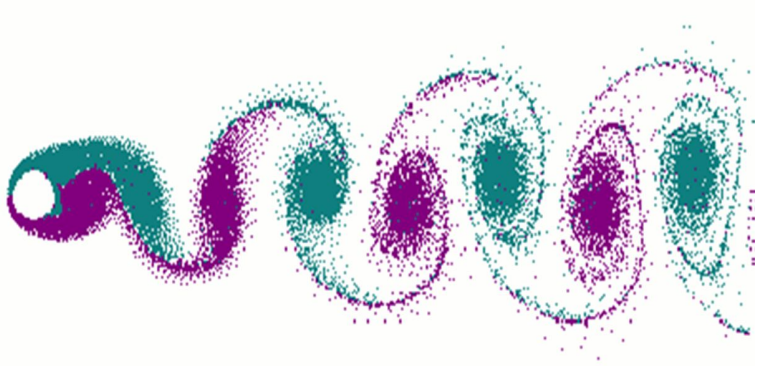
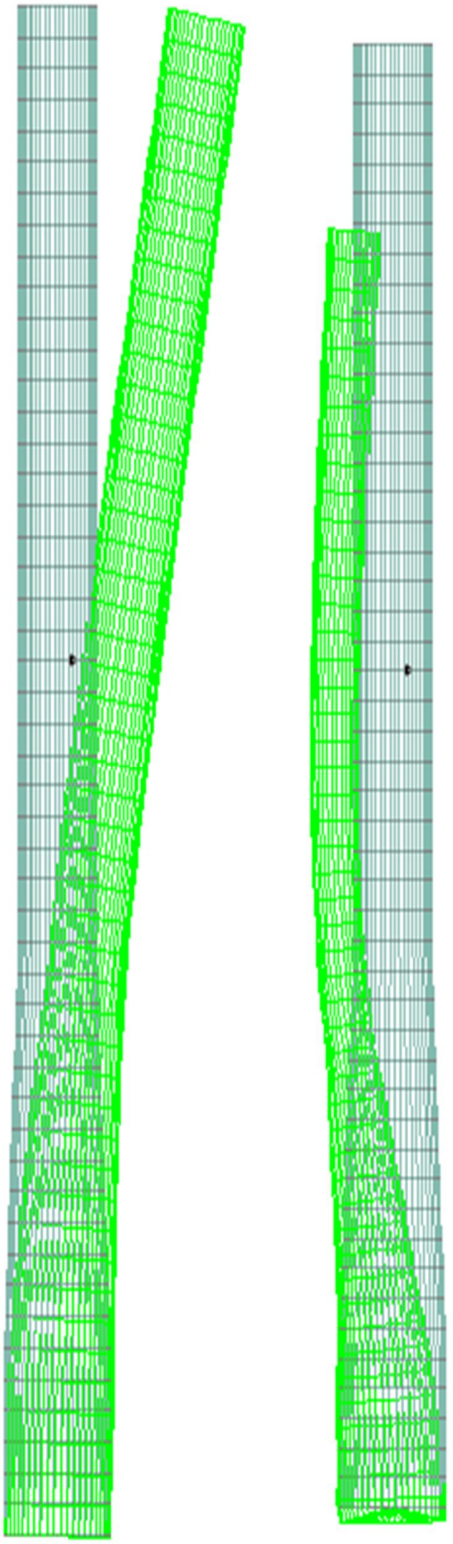


Figure 1 Vortex Shedding

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com



Along wind bending of chimney (N.T.S.) Across wind bending of chimney (N.T.S.)

Figure 2

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

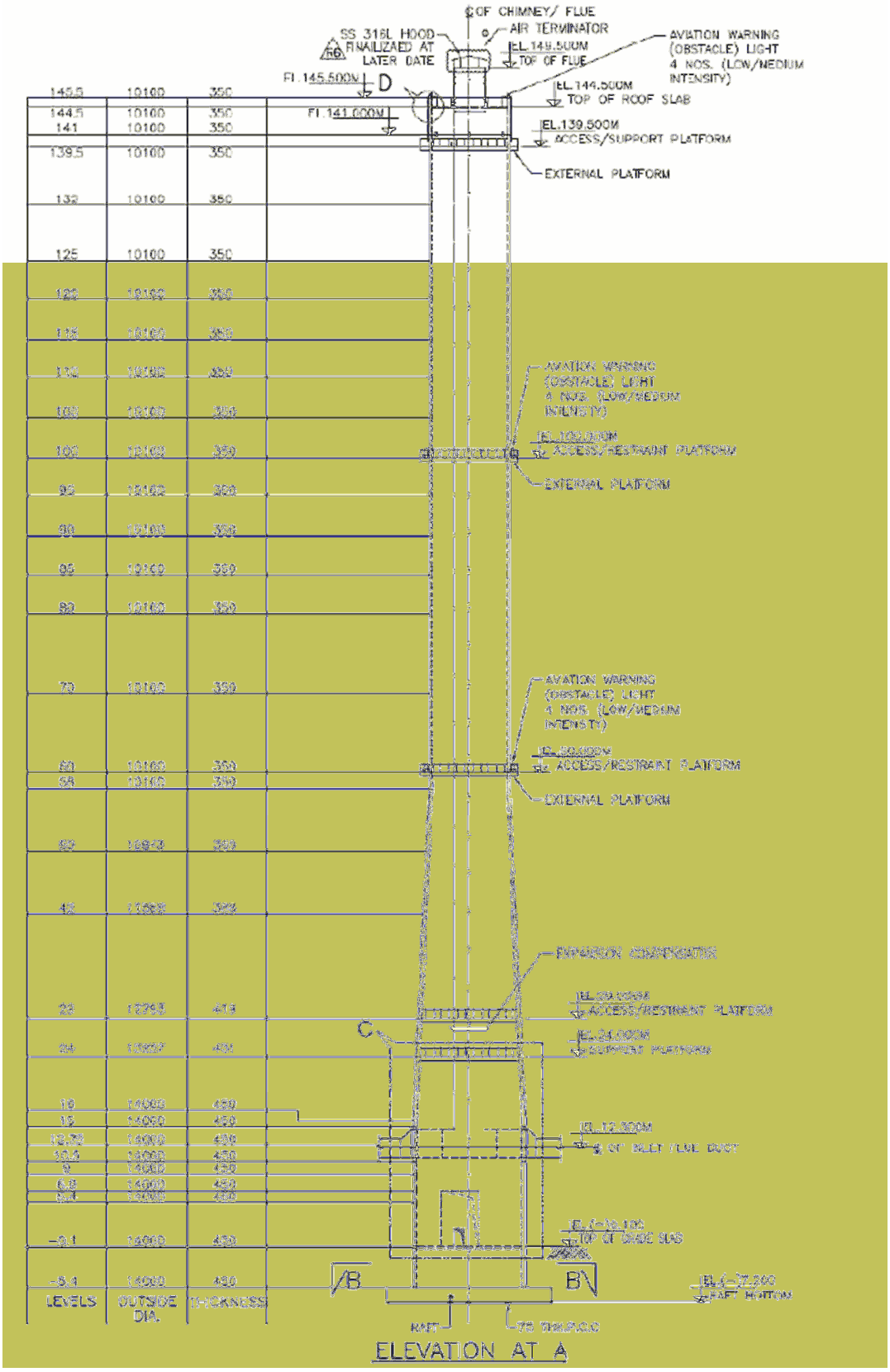


Figure 3 Chimney elevation

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

III. NUMERICAL PROBLEM

A. Design Wind Speed

The basic wind speed(Vb), from figure 1 of IS: 875 (Part 3) – 1987, is 39m/s for Shirpur TPS at Nardana. Basic wind speed is based on peak gust velocity which is averaged over a short time interval of 3 seconds and which corresponds to mean heights 10 m over ground level in an open ground for a 50 year return period. The basic wind speed is modified in order to include the subsequent effects to get design wind velocity at a height (Vz) for the chimney structure:

1) Risk level;
2) topography roughness, elevation and size of structure; and
3) Local geography

It can be mathematically expressed as follows:

$Vz = Vb\ k1k2\ k3$

Where,

Vb= basic wind speed
= 39 m/s
k1 = probability factor (risk coefficient)
= 1.06
k2 = terrain, height and structure size factor, it is calculated for different heights, as per IS 4998 Part 1 we need to consider
K3 = topography factor
= 1.0

Thus, the design wind velocity and related base shear and bending moment for along wind with simplified method.
Class C, Category 2

Height(M)	K1	K2	K3	Vb	Vz	pz(N/Sqm)
145.5	1.06	1.28	1	39	52.9152	1680.011
144.5	1.06	1.28	1	39	52.9152	1680.011
140	1.06	1.28	1	39	52.9152	1680.011
135	1.06	1.28	1	39	52.9152	1680.011
130	1.06	1.28	1	39	52.9152	1680.011
125	1.06	1.28	1	39	52.9152	1680.011
120	1.06	1.28	1	39	52.9152	1680.011
115	1.06	1.28	1	39	52.9152	1680.011
110	1.06	1.28	1	39	52.9152	1680.011
105	1.06	1.28	1	39	52.9152	1680.011
100	1.06	1.24	1	39	51.2616	1576.651
95	1.06	1.24	1	39	51.2616	1576.651
90	1.06	1.24	1	39	51.2616	1576.651
85	1.06	1.24	1	39	51.2616	1576.651
80	1.06	1.24	1	39	51.2616	1576.651
75	1.06	1.24	1	39	51.2616	1576.651
70	1.06	1.24	1	39	51.2616	1576.651
65	1.06	1.24	1	39	51.2616	1576.651
60	1.06	1.24	1	39	51.2616	1576.651
58	1.06	1.24	1	39	51.2616	1576.651
50	1.06	1.17	1	39	48.3678	1403.6664
42	1.06	1.17	1	39	48.3678	1403.6664
29	1.06	1.12	1	39	46.3008	1286.2584
24	1.06	1.12	1	39	46.3008	1286.2584
16	1.06	1.07	1	39	44.2338	1173.9774
15	1.06	1.05	1	39	43.407	1130.5006
12.75	1.06	1.05	1	39	43.407	1130.5006
10.5	1.06	1.05	1	39	43.407	1130.5006
9	1.06	1	1	39	41.34	1025.3974
6.9	1.06	1	1	39	41.34	1025.3974
5.4	1.06	1	1	39	41.34	1025.3974
0	1.06	1	1	39	41.34	1025.3974

©IJRASET: All Rights are Reserved

B. Circumferential Wind Moments = 0.33 x pz x r^2 N-m/m(Ref. Clause 5.4 IS 4998 Part 1)

Where

Moe and Moi = internal and external ring moment

Pz = Design Wind Pressure at Z in N/sqm

rm = Mean Radius of the shell at the cross-section under consideration in M

Height	pz	r radius in m	Moe	Moi
145.5	1680	5.05	14139	14139
144.5	1680	5.05	14139	14139
140.0	1680	5.05	14139	14139
135.0	1680	5.05	14139	14139
130.0	1680	5.05	14139	14139
125.0	1680	5.05	14139	14139
120.0	1680	5.05	14139	14139
115.0	1680	5.05	14139	14139
110.0	1680	5.05	14139	14139
105.0	1680	5.05	14139	14139
100.0	1577	5.05	13269	13269
95.0	1577	5.05	13269	13269
90.0	1577	5.05	13269	13269
85.0	1577	5.05	13269	13269
80.0	1577	5.05	13269	13269
75.0	1577	5.05	13269	13269
70.0	1577	5.05	13269	13269
65.0	1577	5.05	13269	13269
60.0	1577	5.05	13269	13269
58.0	1577	5.05	13269	13269
50.0	1404	5.4215	13615	13615
42.0	1404	5.793	15545	15545
29.0	1286	6.3965	17367	17367
24.0	1286	6.6285	18650	18650
16.0	1174	7	18983	18983
15.0	1131	7	18280	18280
12.8	1131	7	18280	18280
10.5	1131	7	18280	18280
9.0	1025	7	16581	16581
6.9	1025	7	16581	16581
5.4	1025	7	16581	16581
0.0	1025	7	16581	16581

©IJRASET: All Rights are Reserved

C. Calculation For Static Wind Load

Along wind load –Simplified method (Peak factor method)**:**

$$Fz, = PzCD\ d\ z$$

Where

Pz= design wind pressure obtained in accordance IS 875 (part3):1987.

Z = height of section of the chimney in m measured from the top of foundation.

CD = drag coefficient of the chimney taken as 0.8&dz = external diameter of the chimney.

NOTE — Take the proper factor leading to the class of the structure as defined in the IS 875 (part3):1987.

Height	pz	CD	FZ	DIA	SHEAR	BASE SHEAR	Bending
m			N/sqm/M	dz	M-Ton	M-Ton	Mton-M
146	1680	0.8	13574	10.1	1.36	0.00	
145	1680	0.8	13574	10.1	6.11	1.36	0.678724
140	1680	0.8	13574	10.1	6.79	7.47	22.05854
135	1680	0.8	13574	10.1	6.79	14.25	76.3565
130	1680	0.8	13574	10.1	6.79	21.04	164.5907
125	1680	0.8	13574	10.1	6.79	27.83	286.7611
120	1680	0.8	13574	10.1	6.79	34.61	442.8677
115	1680	0.8	13574	10.1	6.79	41.40	632.9106
110	1680	0.8	13574	10.1	6.79	48.19	856.8896
105	1680	0.8	13574	10.1	6.58	54.98	1114.283
100	1577	0.8	12739	10.1	6.37	61.56	1405.091
95	1577	0.8	12739	10.1	6.37	67.92	1728.79
90	1577	0.8	12739	10.1	6.37	74.29	2084.339
85	1577	0.8	12739	10.1	6.37	80.66	2471.735
80	1577	0.8	12739	10.1	6.37	87.03	2890.98
75	1577	0.8	12739	10.1	6.37	93.40	3342.073
70	1577	0.8	12739	10.1	6.37	99.77	3825.015
65	1577	0.8	12739	10.1	6.37	106.14	4339.805
60	1577	0.8	12739	10.1	2.55	112.51	4876.889
58	1577	0.8	12739	10.1	9.97	115.06	5111.88
50	1404	0.8	12176	10.843	10.07	125.03	6072.661
42	1404	0.8	13010	11.586	17.01	135.10	7140.926
29	1286	0.8	13164	12.793	6.70	152.11	8940.798
24	1286	0.8	13642	13.257	10.72	158.82	9728.16
16	1174	0.8	13149	14	1.29	169.53	11003.85
15	1131	0.8	12662	14	2.85	170.82	11174.8
13	1131	0.8	12662	14	2.85	173.67	11562.36
11	1131	0.8	12662	14	1.81	176.52	11955.16
9	1025	0.8	11484	14	2.41	178.33	12221.75
7	1025	0.8	11484	14	1.72	180.74	12598.05
5	1025	0.8	11484	14	6.20	182.47	12873.81
0	1025	0.8	11484	14		188.67	13859.13

©IJRASET: All Rights are Reserved

D. Calculations Of Dynamic Wind Loads

Across wind load –Simplified method:

$$\eta_{oi} = \left\{ \frac{\int_0^H d_z \, \phi_{zi} \, d_z}{\int_0^H \phi^2_{zi} d_z} \right\} \times \frac{C_L}{4 \pi S^2_n K_{si}}$$

Where

ηoi= peak tip deflection due to the vortex shedding in the ith mode of the vibration in m

dz = external diameter of chimney.

CL= peak oscillatory lift coefficient to be taken as 0.16

H = height of the chimney in m.

Ksi = mass damping parameter for the ith mode of vibration

Sn= Strouhal number to be taken as 0.2

Øzi= mode shape function normalized with respect to the dynamic amplitude at top of the chimney in the ithmode of vibration

1) Calculating Ø (mode shape): This is calculated from STAAD software using Dynamic analysis using plates in space frame for details refer appendix: A

Consolidated data for different modes

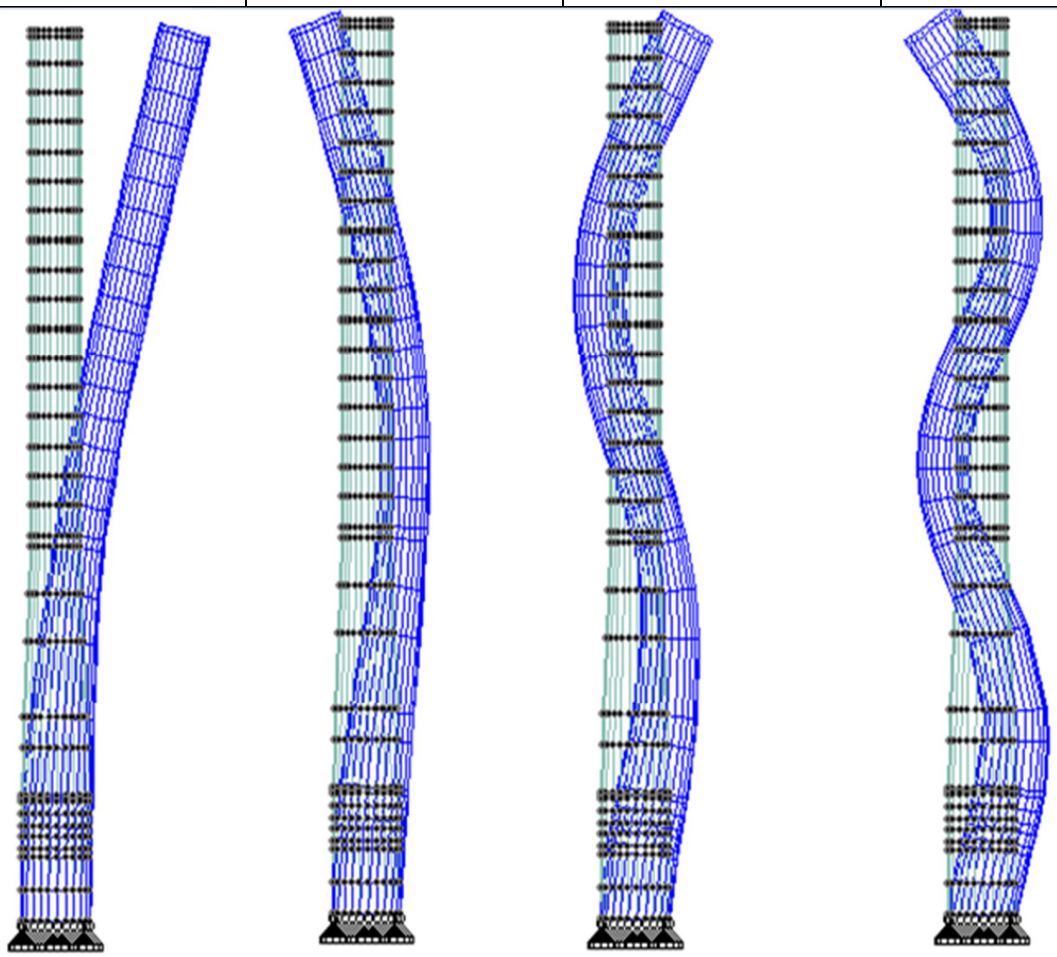
Height	Outer Diameter	Thickness	Mass	Mode 1	Mode 2	Mode 3	Mode 4	
M	OD	M	Kg/m	Ø1	Ø2	Ø3	Ø4	
145.5	10.1	0.35	26788.13	1	-1	1	-1	
144.5	10.1	0.35	26788.13	0.989	-0.966	0.947	-0.927	
140	10.1	0.35	26788.13	0.942	-0.813	0.711	-0.61	
135	10.1	0.35	26788.13	0.89	-0.642	0.448	-0.254	
130	10.1	0.35	26788.13	0.837	-0.472	0.19	0.081	
125	10.1	0.35	26788.13	0.785	-0.305	-0.052	0.364	
120	10.1	0.35	26788.13	0.732	-0.142	-0.268	0.569	
115	10.1	0.35	26788.13	0.68	0.014	-0.447	0.675	
110	10.1	0.35	26788.13	0.629	0.16	-0.579	0.673	
105	10.1	0.35	26788.13	0.578	0.294	-0.659	0.567	
100	10.1	0.35	26788.13	0.528	0.414	-0.683	0.375	
95	10.1	0.35	26788.13	0.48	0.518	-0.651	0.126	
90	10.1	0.35	26788.13	0.432	0.604	-0.568	-0.146	
85	10.1	0.35	26788.13	0.386	0.67	-0.439	-0.401	
80	10.1	0.35	26788.13	0.342	0.716	-0.277	-0.605	
75	10.1	0.35	26788.13	0.3	0.742	-0.094	-0.729	
70	10.1	0.35	26788.13	0.261	0.747	-0.095	-0.756	
65	10.1	0.35	26788.13	0.224	0.734	0.277	-0.684	
60	10.1	0.35	26788.13	0.19	0.702	0.438	-0.527	
58	10.1	0.35	26788.13	0.178	0.686	0.493	-0.449	
50	10.843	0.36	29624.96	0.134	0.595	0.637	-0.052	
42	11.586	0.388	34106.87	0.097	0.487	0.679	0.318	
29	12.793	0.419	40699.94	0.052	0.305	0.568	0.637	
24	13.257	0.431	43394.85	0.039	0.24	0.486	0.641	
16	14	0.45	47865.38	0.022	0.148	0.335	0.536	
15	14	0.45	47865.38	0.02	0.137	0.316	0.518	
12.75	14	0.45	47865.38	0.016	0.114	0.271	0.466	
10.5	14	0.45	47865.38	0.013	0.093	0.228	0.41	
9	14	0.45	47865.38	0.011	0.079	0.2	0.371	
6.9	14	0.45	47865.38	0.008	0.063	0.164	0.315	
5.4	14	0.45	47865.38	0.007	0.052	0.138	0.274	
-0.1	14	0.45	47865.38	0.002	0.021	0.06	0.129	
-5.5	14	0.45	47865.38	0	0	0	0	
				2.68	0.55	0.222	0.124	period sec
				2.34	11.42	28.28	50.82	ώ

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Mode shapes

Mode 1	Mode 2	Mode 3	Mode 4



Shapes for different modes.

Figure 4 view showing different mode shapes

Knowing the values of Ø following equation can be calculated using graphical method as specified by GeoffryM.Pinfort, Nachshen, crofts and Leggatt in book "Reinforced concrete Chimneys and towers"

2) Finding Ksi = mass damping parameter for the ith mode of vibration

Given by

$$K_{si} = \frac{2m_{ei}\delta_s}{\sigma.d^2}$$ where

mei = equivalent mass per unit length in kg/m in the ith mode of vibration as defined,

δs = logarithmic decrement of the structural damping = 2πβ

β = structural damping as a fraction of critical damping to be taken 0.016,

σ = mass density of air to be taken 1.2 kg/m3,

d = effective diameter taken as average diameter above the top 1/3rd height of chimney in m.

= 10.1M

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

3) The equivalent mass per unit length in ith mode of vibration (mei):

Follwing eq is shown the ith mode of vibration

$$m_{ei} = \frac{\int_0^H m_z \phi^2_{zi} . d_z}{\int_0^H \phi^2_{zi} . d_z}$$

4) Consolidated data and corresponding results of K_{s1}, m_{e1}, η_{o1} for first mode

Height	Outer	Thickness	Mass	Mode	mz ø1 dz	mz ø1 (z-z0)	mz	ø1^2 dz	Dz ø1 dz	ø1^2 dz
M	OD	M	Kg/m							
145.5	10.1	0.35	26788.125	1	26640.79	3876234.99	26494	0.98903	10.04445	0.98903025
144.5	10.1	0.35	26788.125	0.989	116387.7	16818023.53	112372	4.194856	43.881975	4.19485613
140	10.1	0.35	26788.125	0.942	122689.6	17176545.75	112384	4.19528	46.258	4.19528
135	10.1	0.35	26788.125	0.89	115657.7	15613793.51	99870	3.728161	43.60675	3.72816125
130	10.1	0.35	26788.125	0.837	108625.8	14121360.09	88096	3.288605	40.9555	3.288605
125	10.1	0.35	26788.125	0.785	101594	12699245.51	77059	2.876611	38.30425	2.87661125
120	10.1	0.35	26788.125	0.732	94562.08	11347449.75	66761	2.49218	35.653	2.49218
115	10.1	0.35	26788.125	0.68	87664.14	10081375.99	57376	2.141851	33.05225	2.14185125
110	10.1	0.35	26788.125	0.629	80833.17	8891648.391	48783	1.821061	30.47675	1.82106125
105	10.1	0.35	26788.125	0.578	74069.17	7777262.391	40960	1.529045	27.9265	1.529045
100	10.1	0.35	26788.125	0.528	67506.07	6750607.5	34023	1.27008	25.452	1.27008
95	10.1	0.35	26788.125	0.48	61076.92	5802307.875	27851	1.03968	23.028	1.03968
90	10.1	0.35	26788.125	0.432	54781.72	4930354.406	22406	0.836405	20.6545	0.836405
85	10.1	0.35	26788.125	0.386	48754.39	4144122.937	17747	0.66248	18.382	0.66248
80	10.1	0.35	26788.125	0.342	42994.94	3439595.25	13801	0.515205	16.2105	0.515205
75	10.1	0.35	26788.125	0.3	37570.35	2817775.898	10538	0.393401	14.16525	0.39340125
70	10.1	0.35	26788.125	0.261	32480.6	2273642.109	7877	0.294031	12.24625	0.29403125
65	10.1	0.35	26788.125	0.224	27725.71	1802171.109	5739	0.214245	10.4535	0.214245
60	10.1	0.35	26788.125	0.19	9858.03	591481.8	1814	0.067712	3.7168	0.067712
58	10.1	0.35	26788.125	0.178	35201.76	2041702.3	5215	0.194688	12.6048	0.194688
50	10.843	0.36	29624.958	0.134	29444.1	1472205.19	3162	0.106722	10.018932	0.106722
42	11.586	0.388	34106.868	0.097	36225.2	1521458.315	2461	0.072153	11.221041	0.07215325
29	12.793	0.419	40699.942	0.052	9565.782	277407.6859	421	0.010351	2.9104075	0.01035125
24	13.257	0.431	43394.847	0.039	11133.75	267209.9303	323	0.007442	3.234708	0.007442
16	14	0.45	47865.375	0.022	1005.173	16082.766	21	0.000441	0.294	0.000441
15	14	0.45	47865.375	0.02	1938.548	29078.21531	35	0.000729	0.567	0.000729
12.75	14	0.45	47865.375	0.016	1561.608	19910.50021	23	0.000473	0.45675	0.00047306
10.5	14	0.45	47865.375	0.013	861.5768	9046.555875	10	0.000216	0.252	0.000216
9	14	0.45	47865.375	0.011	954.9142	8594.228081	9	0.00019	0.2793	0.00018953
6.9	14	0.45	47865.375	0.008	538.4855	3715.549734	4	8.44E-05	0.1575	8.4375E-05
5.4	14	0.45	47865.375	0.007	1163.129	6280.894508	5	0.000109	0.3402	0.00010935
0	14	0.45	47865.375	0.002	263.2596	0	0	5.5E-06	0.077	0.0000055
-5.5	14	0.45	47865.375	0						
				∑	1441330	156627690.9	883641	32.94353	536.88186	32.9435252
							me1	26822.91	ηo1	0.02357315
							Ks1	44.03434		

©IJRASET: All Rights are Reserved

5) Consolidated data and corresponding results of K_{s2}, m_{e2}, η_{o2} for second mode

Height	Outer	Thickness	Mode	mz ø2 dz	mz ø2 (z-z0)	mz ø2^2	ø2^2 dz	Dz ø2 dz	ø2^2 dz
M	OD	M							
145.5	10.1	0.35	-1	-26332.7	-3831411.8	25885.07	0.966289	-9.9283	0.966289
144.5	10.1	0.35	-0.966	-107226	-15494181	95377.68	3.560446	-40.4278	3.560446
140	10.1	0.35	-0.813	-97441.8	-13641853	70888.91	2.646281	-36.7388	2.646281
135	10.1	0.35	-0.642	-74604.9	-10071665	41554.94	1.551245	-28.1285	1.551245
130	10.1	0.35	-0.472	-52035.9	-6764671.3	20215.96	0.754661	-19.6193	0.754661
125	10.1	0.35	-0.305	-29935.7	-3741966.2	6690.636	0.249761	-11.2868	0.249761
120	10.1	0.35	-0.142	-8572.2	-1028664	548.6208	0.02048	-3.232	0.02048
115	10.1	0.35	0.014	11652.83	1340075.95	1013.797	0.037845	4.3935	0.037845
110	10.1	0.35	0.16	30404.52	3344497.41	6901.826	0.257645	11.4635	0.257645
105	10.1	0.35	0.294	47414.98	4978573.03	16784.9	0.62658	17.877	0.62658
100	10.1	0.35	0.414	62416.33	6241633.12	29086.01	1.08578	23.533	1.08578
95	10.1	0.35	0.518	75140.69	7138365.61	42153.93	1.573605	28.3305	1.573605
90	10.1	0.35	0.604	85320.18	7678816.03	54348.95	2.028845	32.1685	2.028845
85	10.1	0.35	0.67	92820.85	7889772.52	64324.85	2.401245	34.9965	2.401245
80	10.1	0.35	0.716	97642.72	7811417.25	71181.54	2.657205	36.8145	2.657205
75	10.1	0.35	0.742	99718.8	7478909.65	74240.64	2.771401	37.59725	2.771401
70	10.1	0.35	0.747	99183.03	6942812.3	73445.04	2.741701	37.39525	2.741701
65	10.1	0.35	0.734	96169.37	6251008.97	69049.61	2.57762	36.259	2.57762
60	10.1	0.35	0.702	37181.92	2230915.05	25804.25	0.963272	14.0188	0.963272
58	10.1	0.35	0.686	144530.3	8382758.48	92571.67	3.281922	51.7524	3.281922
50	10.84	0.36	0.595	137915.7	6895783.62	74612.38	2.341448	46.9285	2.341448
42	11.59	0.388	0.487	192552.7	8087214.67	76250.88	2.038608	59.64473	2.038608
29	12.79	0.419	0.305	57289.58	1661397.68	15611.41	0.371281	17.43046	0.371281
24	13.26	0.431	0.24	70817.93	1699630.38	13738.68	0.301088	20.57486	0.301088
16	14	0.45	0.148	6820.816	109133.055	971.9663	0.020306	1.995	0.020306
15	14	0.45	0.137	13515.99	202739.779	1696.256	0.035438	3.95325	0.035438
12.75	14	0.45	0.114	11146.65	142119.777	1153.678	0.024103	3.26025	0.024103
10.5	14	0.45	0.093	6174.633	64833.6504	531.0185	0.011094	1.806	0.011094
9	14	0.45	0.079	7136.727	64230.5467	506.7076	0.010586	2.0874	0.010586
6.9	14	0.45	0.063	4128.389	28485.8813	237.3823	0.004959	1.2075	0.004959
5.4	14	0.45	0.052	9434.265	50945.0332	344.3507	0.007194	2.7594	0.007194
0	14	0.45	0.021	2764.225	0	29.02437	0.000606	0.8085	0.000606
-5.5	14	0.45	0						
			∑	1103145	42141657.1	1067753	37.92054	379.6942	37.92054
						me2	28157.63	ηo2	0.013797
						Ks2	46.22551		

©IJRASET: All Rights are Reserved

Calculated values for the first 2 modes we get

ηo1	ηo2
0.02357315	0.013797

Calculation of the sectional shear force (F_{zoi})&Sectional Bending Moment (M_{zoi})

$$F_{zoi} = 4\pi^2 f_1^{\,2} \eta_{oi} \int_{z0}^{H} m_z \phi_{zi} d_z$$

$$M_{zoi} = 4\pi^2 f_1^{\,2} \eta_{oi} \int_{z0}^{H} m_z \phi_{zi} (z - z_0) d_z .$$

where
f1 = natural frequency of the chimney in Hz in the ith mode of vibration
mz = mass per unit length of the chimney at section z in kg/m.

Result taken from STAAD Pro.

Mode	Frequency Hz
1	0.373
2	1.818

Critical wind speed

$$V_{cri} = \frac{f_1 d}{S_n}$$

d	avgdia of top 1/3	10.1
Sn	strouhal number	0.2
Vcr1	18.8365	
Vcr1	91.809	does not govern

Thus calculated values of the Shear force Kg&Bending Moment Kg-M at base is as follows:

Fzo1	Fzo2	Mzo1	Mzo2
186431	Doesn't govern	20259236	Doesn't govern

©IJRASET: All Rights are Reserved

E. Along wind R.R.M. Analysis

The along wind load per unit height at any height z on a chimney shall be calculated from the equation: Fz = Fzm + Fzf

							Along	Wind	R.R.M				SHEAR	Bending
Height (M)	V_{10}	f_1	S	E	B	r	Vt	g_r	G	∫Fzm.Z.dz	F_{zf}	Fzf+ Fzm	M-Ton	Mton-M
145.5	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	137102	28	13602	0	
144.5	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	616961	123	13698	1	1
140	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	133	13707	8	22
135	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	128	13703	14	77
130	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	123	13698	21	166
125	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	119	13693	28	289
120	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	114	13688	35	447
115	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	109	13684	42	639
110	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	104	13679	49	864
105	49.92	0.37	0.3	0.09	0.63	0.24	1148.96	3.91	2.42	685512	100	13674	55	1123
100	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	88	12827	62	1415
95	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	83	12823	68	1739
90	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	79	12818	74	2096
85	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	75	12814	81	2484
80	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	70	12810	87	2904
75	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	66	12805	94	3357
70	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	62	12801	100	3841
65	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	643337	57	12796	106	4357
60	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	257335	21	12760	113	4896
58	48.36	0.37	0.3	0.09	0.63	0.24	1143.13	3.91	2.4	1029339	82	12821	115	5132
50	45.63	0.37	0.28	0.09	0.63	0.24	1131.88	3.9	2.37	1056192	70	12246	126	6097
42	45.63	0.37	0.28	0.09	0.63	0.24	1131.88	3.9	2.37	1959586	110	13120	136	7171
29	43.68	0.37	0.27	0.09	0.63	0.24	1122.92	3.9	2.34	842041	32	13196	154	8983
24	43.68	0.37	0.27	0.09	0.63	0.24	1122.92	3.9	2.34	1446767	45	13687	160	9776
16	41.73	0.37	0.27	0.08	0.63	0.24	1113.06	3.9	2.31	184080	4	13152	171	11063
15	40.95	0.37	0.26	0.08	0.63	0.24	1108.85	3.9	2.3	398841	8	12669	172	11235
12.75	40.95	0.37	0.26	0.08	0.63	0.24	1108.85	3.9	2.3	398841	6	12668	175	11625
10.5	40.95	0.37	0.26	0.08	0.63	0.24	1108.85	3.9	2.3	265894	4	12665	177	12019
9	39	0.37	0.25	0.08	0.63	0.24	1097.56	3.9	2.27	337643	4	11488	179	12287
6.9	39	0.37	0.25	0.08	0.63	0.24	1097.56	3.9	2.27	241173	2	11487	182	12665
5.4	39	0.37	0.25	0.08	0.63	0.24	1097.56	3.9	2.27	868224	6	11490	183	12940
0	39	0.37	0.25	0.08	0.63	0.24	1097.56	3.9	2.27	0	0	11484	186	13929
									=	20670804				

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Height	Mode 1	β	CL	L	Sn	σ	∩	me1	ka	ø1^2 dz	deflection
M	ø1					mass density of air				with respect to height	M
145.5	1	0.016	0.12	1	0.2	1.2	14.40594059	26822.90553	0.5	32.94352519	0.073196556
144.5	0.989	0.016	0.12	1	0.2	1.2	14.30693069	26822.90553	0.5	31.95449494	0.073250134
140	0.942	0.016	0.12	1	0.2	1.2	13.86138614	26822.90553	0.5	27.75963881	0.073680438
135	0.89	0.016	0.12	1	0.2	1.2	13.36633663	26822.90553	0.5	23.56435881	0.074194765
130	0.837	0.016	0.12	1	0.2	1.2	12.87128713	26822.90553	0.5	19.83619756	0.074629771
125	0.785	0.016	0.12	1	0.2	1.2	12.37623762	26822.90553	0.5	16.54759256	0.075145234
120	0.732	0.016	0.12	1	0.2	1.2	11.88118812	26822.90553	0.5	13.67098131	0.075534619
115	0.68	0.016	0.12	1	0.2	1.2	11.38613861	26822.90553	0.5	11.17880131	0.075963401
110	0.629	0.016	0.12	1	0.2	1.2	10.89108911	26822.90553	0.5	9.036950063	0.076432881
105	0.578	0.016	0.12	1	0.2	1.2	10.3960396	26822.90553	0.5	7.215888813	0.076793022
100	0.528	0.016	0.12	1	0.2	1.2	9.900990099	26822.90553	0.5	5.686843813	0.077115638
95	0.48	0.016	0.12	1	0.2	1.2	9.405940594	26822.90553	0.5	4.416763813	0.077534526
90	0.432	0.016	0.12	1	0.2	1.2	8.910891089	26822.90553	0.5	3.377083813	0.077674464
85	0.386	0.016	0.12	1	0.2	1.2	8.415841584	26822.90553	0.5	2.540678813	0.077761758
80	0.342	0.016	0.12	1	0.2	1.2	7.920792079	26822.90553	0.5	1.878198813	0.077739994
75	0.3	0.016	0.12	1	0.2	1.2	7.425742574	26822.90553	0.5	1.362993813	0.077508464
70	0.261	0.016	0.12	1	0.2	1.2	6.930693069	26822.90553	0.5	0.969592563	0.077239433
65	0.224	0.016	0.12	1	0.2	1.2	6.435643564	26822.90553	0.5	0.675561313	0.076527319
60	0.19	0.016	0.12	1	0.2	1.2	5.940594059	26822.90553	0.5	0.461316313	0.07546994
58	0.178	0.016	0.12	1	0.2	1.2	5.742574257	26822.90553	0.5	0.393604313	0.075257235
50	0.134	0.016	0.12	1	0.2	1.2	4.611269944	26822.90553	0.5	0.198916313	0.083858767
42	0.097	0.016	0.12	1	0.2	1.2	3.625064733	26822.90553	0.5	0.092194313	0.087830751
29	0.052	0.016	0.12	1	0.2	1.2	2.266864692	26822.90553	0.5	0.020041063	0.097161465
24	0.039	0.016	0.12	1	0.2	1.2	1.810364336	26822.90553	0.5	0.009689813	0.100612033
16	0.022	0.016	0.12	1	0.2	1.2	1.142857143	26822.90553	0.5	0.002247813	0.10471596
15	0.02	0.016	0.12	1	0.2	1.2	1.071428571	26822.90553	0.5	0.001806813	0.102808572
12.75	0.016	0.016	0.12	1	0.2	1.2	0.910714286	26822.90553	0.5	0.001077813	0.098177894
10.5	0.013	0.016	0.12	1	0.2	1.2	0.75	26822.90553	0.5	0.00060475	0.096640754
9	0.011	0.016	0.12	1	0.2	1.2	0.642857143	26822.90553	0.5	0.00038875	0.094425433
6.9	0.008	0.016	0.12	1	0.2	1.2	0.492857143	26822.90553	0.5	0.000199225	0.083994889
5.4	0.007	0.016	0.12	1	0.2	1.2	0.385714286	26822.90553	0.5	0.00011485	0.085632706
0	0.002	0.016	0.12	1	0.2	1.2	0	26822.90553	0.5	5.5E-06	
-5.5	0	0.016	0.12	1	0.2	1.2					
										∑	32.94352519
										ηo1 (max)	0.10471596

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

F. Across Wind R.R.M. Analysis

Thus values of the Shear force Kg&Bending Moment Kg-M at base is as follows:

Fzo1	Fzo2	Mzo1	Mzo2
53268.48743	DOES NOT GOVERN	43560.9311	DOES NOT GOVERN

G. Seismic Analysis

1) Horizontal Seismic Force The horizontal seismic coefficient Ah, shall be obtained using the period T, described as under.

2) When using site specific spectra, the seismic coefficient shall be calculated from the expression :

$$A_h = \frac{\left[S_a/g\right]}{(R/I)}$$

where / g = spectral acceleration coefficient calculated from the expression:

$$A_h = \frac{\left[Z/2\right]\left[S_a/g\right]}{(R/I)}$$

where

Z = Zone factor

Sa/g = Spectral acceleration coefficient

R = Response reduction factor

Table 1 Importance factor for various Industrial structures II

Sl No.	Categories of Structures (*see* 7.1)	Importance Factor
(1)	(2)	(3)
i)	Structures in Category 1	2.00
ii)	Structures in Category 2	1.75
iii)	Structures in Category 3	1.50
iv)	Structures in Category 4	1.00

NOTE — Higher importance factor may be assigned to different structures at the discretion of the project authorities.

T lamental time period for stake-like structures, ‘T” is given by:

$$T = C_T\sqrt{\frac{W_t.h}{E_s.A.g}}$$

C_T = coefficient depending upon slenderness ratio.

W_T = Total weight of structure.

h = Height of structure above the base.

Table 2 Values of C_T andC_v

Sl No.	$k = h/r_c$	Coefficient, C_T	Coefficient, C_v
(1)	(2)	(3)	(4)
i)	5	14.4	1.02
ii)	10	21.2	1.12
ii)	15	29.6	1.19
iv)	20	38.4	1.25
v)	25	47.2	1.30
vi)	30	56.0	1.35
vii)	35	65.0	1.39
viii)	40	73.8	1.43
ix)	45	82.8	1.47
x)	50 or more	1.8 *k*	1.50

NOTE — k = slenderness ratio, and
r_c = radius of gyration of the structural shell at the base section.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Calculation for Time period

Area	g	Es (M25)	wt	H/r	Ct	T
19.14615	9.81	32000000000	4651909	30.3	56	1.86

3) Seismic Calculations

Height	Outer Diameter	Thickness	Mass	Mass				time	medium soil	Ah1	HORIZO SEISMIC	BENDING
M	OD	M	Kg/m	kg	Z	I	R	period	Sa/g		LOAD	MOMENT
145.5	10.1	0.35	26788.13	13394.06	0.24	1.75	3	1.8613	0.731	0.0511	685.0619904	99676.51961
144.5	10.1	0.35	26788.13	73667.34	0.24	1.75	3	1.8613	0.731	0.0511	3767.840947	544453.0169
140	10.1	0.35	26788.13	127243.6	0.24	1.75	3	1.8613	0.731	0.0511	6508.088909	911132.4473
135	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	924833.6871
130	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	890580.5876
125	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	856327.488
120	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	822074.3885
115	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	787821.289
110	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	753568.1895
105	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	719315.09
100	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	685061.9904
95	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	650808.8909
90	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	616555.7914
85	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	582302.6919
80	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	548049.5923
75	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	513796.4928
70	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	479543.3933
65	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	445290.2938
60	10.1	0.35	26788.13	93758.44	0.24	1.75	3	1.8613	0.731	0.0511	4795.433933	287726.036
58	10.1	0.35	26788.13	133940.6	0.24	1.75	3	1.8613	0.731	0.0511	6850.619904	397335.9545
50	10.84	0.36	29624.96	236999.7	0.24	1.75	3	1.8613	0.731	0.0511	12121.74884	606087.4419
42	11.59	0.388	34106.87	358122.1	0.24	1.75	3	1.8613	0.731	0.0511	18316.76171	769303.9917
29	12.79	0.419	40699.94	366299.5	0.24	1.75	3	1.8613	0.731	0.0511	18735.00666	543315.1931
24	13.26	0.431	43394.85	282066.5	0.24	1.75	3	1.8613	0.731	0.0511	14426.76874	346242.4497
16	14	0.45	47865.38	215394.2	0.24	1.75	3	1.8613	0.731	0.0511	11016.70018	176267.203
15	14	0.45	47865.38	77781.23	0.24	1.75	3	1.8613	0.731	0.0511	3978.252844	59673.79267
12.75	14	0.45	47865.38	107697.1	0.24	1.75	3	1.8613	0.731	0.0511	5508.350092	70231.46368
10.5	14	0.45	47865.38	89747.58	0.24	1.75	3	1.8613	0.731	0.0511	4590.291744	48198.06331
9	14	0.45	47865.38	86157.68	0.24	1.75	3	1.8613	0.731	0.0511	4406.680074	39660.12066
6.9	14	0.45	47865.38	86157.68	0.24	1.75	3	1.8613	0.731	0.0511	4406.680074	30406.09251
5.4	14	0.45	47865.38	165135.5	0.24	1.75	3	1.8613	0.731	0.0511	8446.136808	45609.13876
0	14	0.45	47865.38	129236.5	0.24	1.75	3	1.8613	0.731	0.0511	6610.020111	0
				4651909							231319.722	15251248.78
				wt							KG	KG-M
											DOSENT GOVERN	

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

IV. RESULTS AND DISCUSSIONS

A. After Analysis of chimney with respect to IS 875, IS 4998, IS 1893, following table Shows Resultant Moment

Table 3Results of chimney

S.NO.	EFFECTS OF WIND	UNIT	W.R.T. (IS CODES), SIMPLIFIED METHOD	RESULTANT	W.R.T. (IS CODES), RANDOM RESPONSE	RESULTANT
1	Along wind maximum bending moment	KTON-M	13.86	24.55	13.93	13.93
2	ACROSS wind maximum bending moment	KTON-M	20.26		0.044	
3	Along wind maximum BASE SHEAR	KTON	0.188	0.26	0.186	0.19
4	ACROSS wind maximum BASE SHEAR	KTON	0.186		0.053	

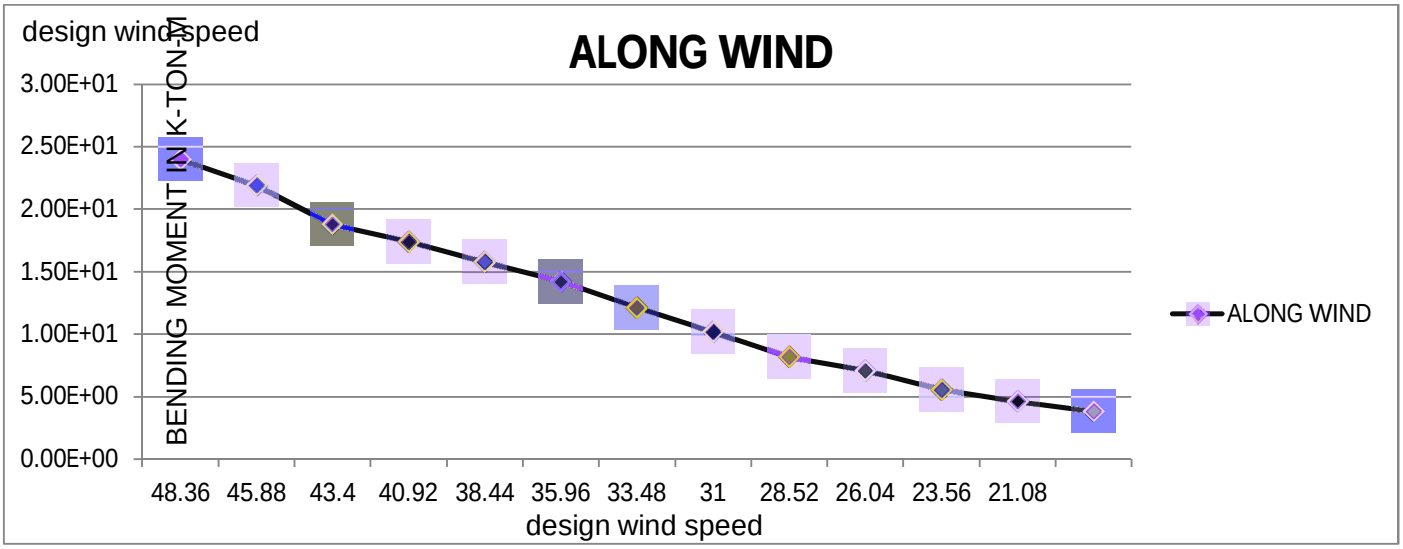


Figure 5 Bending Moment Vs Design wind speed (Along wind)

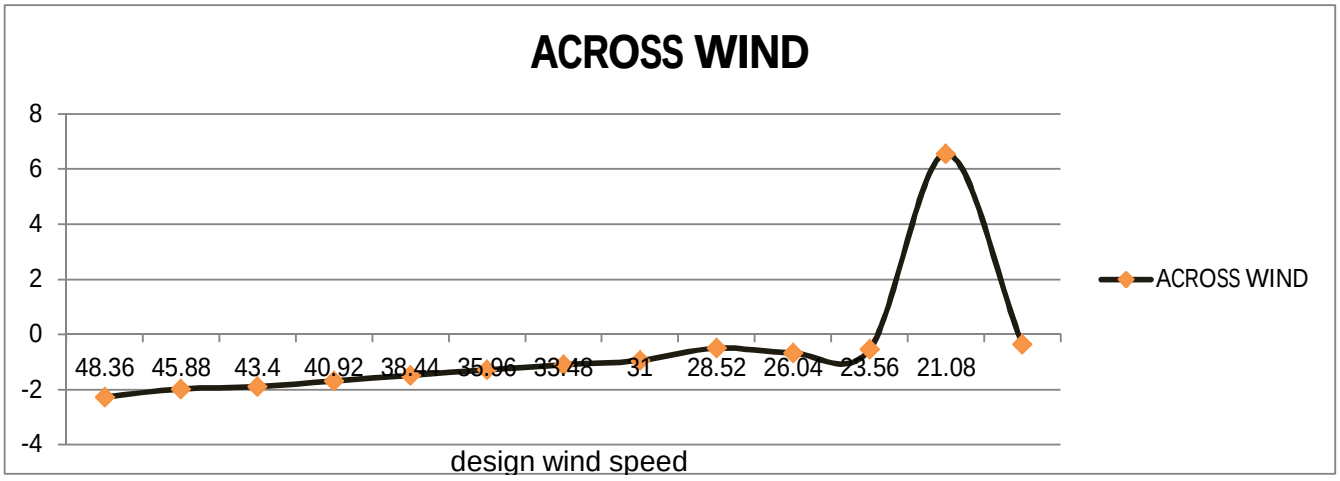


Figure 6 Moment Vs Design wind speed (Across wind)

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Graph Showing Deflection In Mm Vs Height Of The Chimney For Random Response Calculation As Per Is 4998 Equations

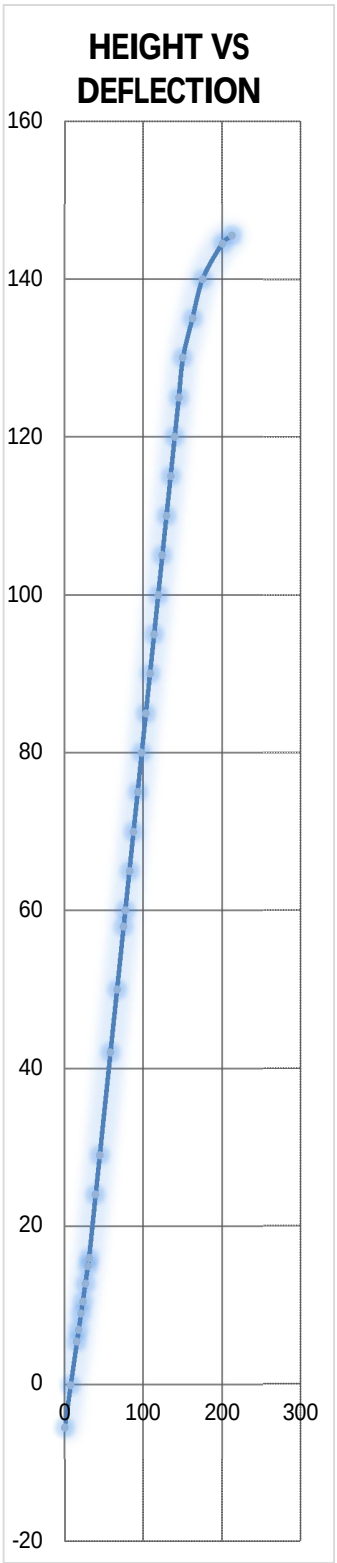


Figure 7 Deflection Vs Height (Along R.R.M.)

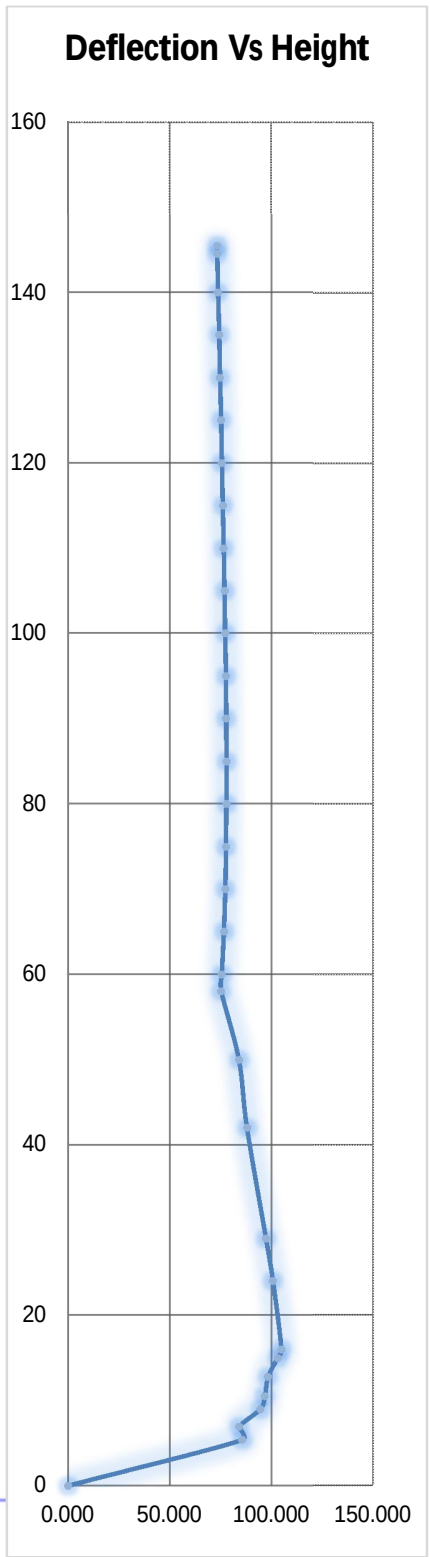


Figure 8 Deflection Vs Height (ACCROS R.R.M)

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Design of CHIMNEY

	SHEAR	BASE SHEAR	Bending	DESIGN CALCULATIONS								
Height(M)	ton	Ton	ton-M	Weight(KN)	Eccentricity (e)mm	Min stress	Max stress	providereinf.	providedreinf.	Compression capacity	Tension capacity	CHECK
						in concrete(N/Sq mm)		1% thus Ast	every 300mm			
145.500	1.37	0.0	0.0									
144.500	6.17	1.4	228.9	273	8	0	0	107153	1050	10	-2	HENCE SAFE
140.000	6.85	7.5	324.9	1502	2122	0	0	107153	1050	10	-2	HENCE SAFE
135.000	6.85	14.3	408.4	2867	1397	0	0	107153	1050	10	-2	HENCE SAFE
130.000	6.85	21.1	571.4	4233	1324	0	1	107153	1050	10	-2	HENCE SAFE
125.000	6.84	27.8	772.1	5598	1353	0	1	107153	1050	10	-2	HENCE SAFE
120.000	6.84	34.6	908.6	6963	1280	0	1	107153	1050	10	-2	HENCE SAFE
115.000	6.84	41.5	1784.2	8329	2102	0	1	107153	1050	10	-2	HENCE SAFE
110.000	6.63	48.3	2326.3	9694	2354	0	2	107153	1050	10	-2	HENCE SAFE
105.000	6.41	55.6	3842.9	11059	3409	0	2	107153	1050	10	-2	HENCE SAFE
100.000	6.41	62.8	4541.5	12425	3586	0	3	107153	1050	10	-2	HENCE SAFE
95.000	6.41	70.0	5150.4	13790	3664	-1	3	107153	1050	10	-2	HENCE SAFE
90.000	6.41	76.5	5390.9	15155	3489	0	3	107153	1050	10	-2	HENCE SAFE
85.000	6.40	83.5	5904.6	16521	3506	-1	4	107153	1050	10	-2	HENCE SAFE
80.000	6.40	90.8	6503.7	17886	3567	-1	4	107153	1050	10	-2	HENCE SAFE
75.000	6.40	98.3	7183.7	19251	3661	-1	4	107153	1050	10	-2	HENCE SAFE
70.000	6.39	104.4	7420.7	20617	3531	-1	5	107153	1050	10	-2	HENCE SAFE
65.000	6.40	112.2	8200.4	21982	3660	-1	5	107153	1050	10	-2	HENCE SAFE
60.000	2.51	120.3	9046.2	23347	3801	-1	5	107153	1050	10	-2	HENCE SAFE
58.000	10.15	125.3	9782.6	23894	4016	-1	6	107153	1050	10	-2	HENCE SAFE
50.000	10.53	137.6	10957.8	26078	4122	-1	5	118500	1080	10	-2	HENCE SAFE
42.000	17.47	150.5	12256.4	28494	4220	-1	5	136427	1164	10	-2	HENCE SAFE
29.000	6.71	170.0	14067.5	33014	4180	-1	5	162800	1257	10	-2	HENCE SAFE
24.000	10.33	180.8	15306.7	35088	4279	-1	5	173579	1293	10	-2	HENCE SAFE
16.000	1.27	195.7	16904.6	38627	4293	0	4	191462	1350	10	-2	HENCE SAFE
15.000	2.85	203.1	17839.5	39115	4474	0	5	191462	1350	10	-2	HENCE SAFE
12.750	2.72	212.8	18967.2	40213	4627	-1	5	191462	1350	10	-2	HENCE SAFE
10.500	1.72	223.5	20151.9	41311	4785	-1	5	191462	1350	10	-2	HENCE SAFE
9.000	2.41	234.4	21318.6	42043	4974	-1	5	191462	1350	10	-2	HENCE SAFE
6.900	1.72	246.7	22605.9	43067	5149	-1	5	191462	1350	10	-2	HENCE SAFE
5.400	114.95	258.4	23775.3	43799	5325	-1	6	191462	1350	10	-2	HENCE SAFE
0.000	0.00	265.2	24546.1	46434	5186	-1.1	6	191462	1350	10	-2	HENCE SAFE

©IJRASET: All Rights are Reserved

V. CONCLUSIONS, DISCUSSION, FUTURE SCOPE.

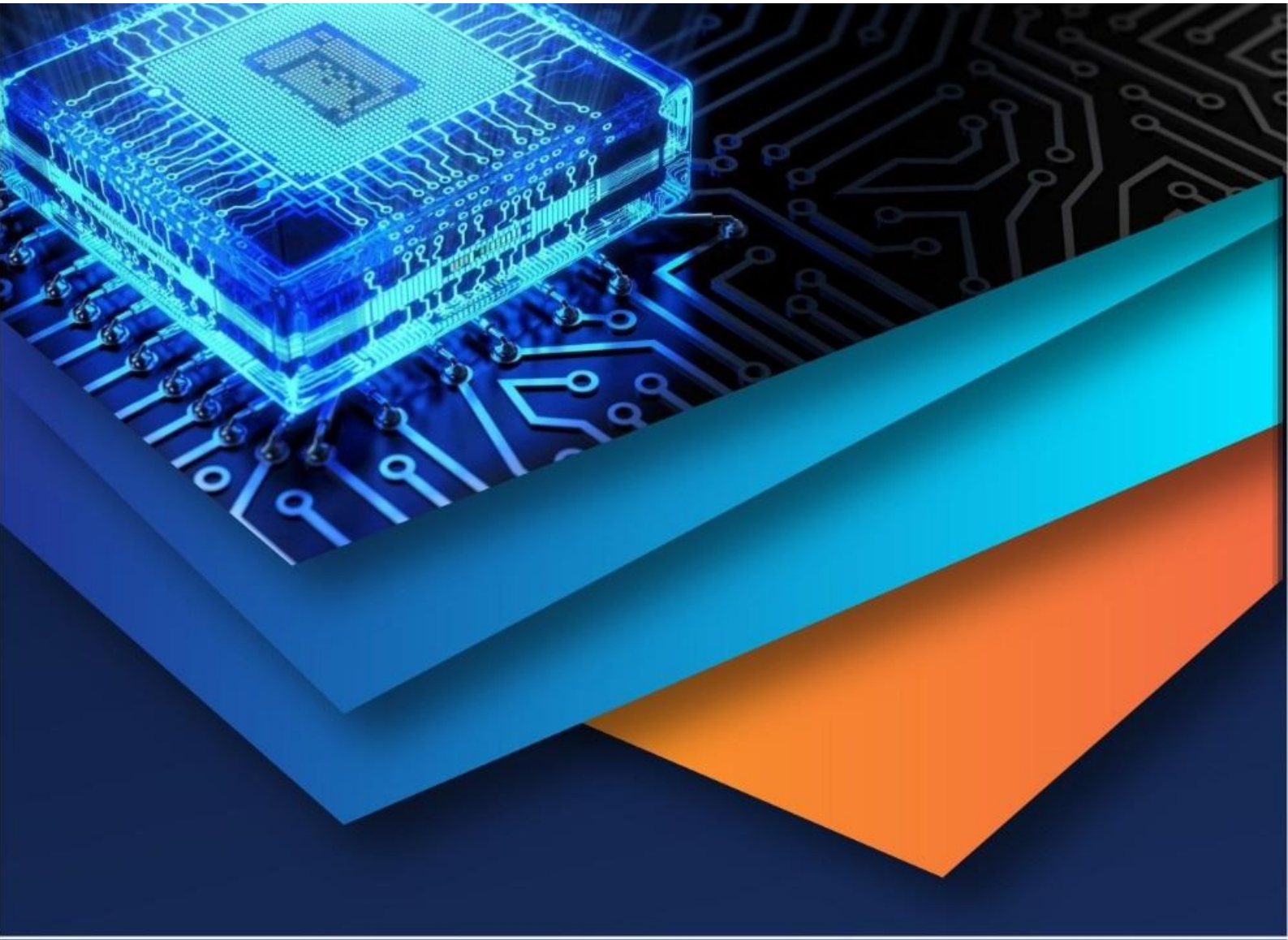
Our objective of study included Analysis of structure as per IS Codes and comparing the results with each other, after we obtained results we found that Bending moment and shear forces obtained from Simplified method and R.R.M method of analysis as per IS 4998 are higher than bending moment and shear force obtained due to earthquake analysis, also every structure is unique and with changes in geometry the governing forces and their values differs, after in depth study of codes in this paper I have worked out a case study of a chimney and using this solution chimney any other chimney problem can be solved quickly and with accuracy.

Future scope includes analysis of chimney of different height and material properties in similar fashion as explained in this case study to understand behavior of structures in a better manner, wind velocity may also be increased to study behavior of structure when excited in higher modes, one may also go for wind tunnel testing and CFD analysis to testify results obtained with Indian standard codes.

REFERENCES

[1] Batham, J.P., "Parameters required for the wind-tunnel simulation of the wind loads on large power station chimneys", Wind Engineering and Industrial dynamics, Elsevier science publishers -Amsterdam, Vol.18, 1985.
[2] Milford R.V., "Structural reliability and cross wind response of tall chimneys". Engineering structures, Butterworth & Co. (Publishers) Ltd, Vol.4, 1982,
[3] Reddy K.R.C, Jaiswal O.R and Godbole P.N, "Wind response control of tall RC chimneys", Wind and Engineering, Vol. 8, 2011.
[4] Reddy K. R. C, Jaiswal O. R. and P. N. Godbole., "Wind and Earthquake Analysis of Tall RC Chimneys", Earth sciences and Engineering, October 2011,
[5] Vickery B.J. and Basu.R. "Simplified approaches to the evaluation of the across wind Response of chimneys", Wind Engineering and Industrial dynamics, Elsevier science Publishers -Amsterdam, 1983, pp. 153-166.
[6] Manohar, S.N., "Tall Chimneys", Tata McGraw-Hill Publishing Company Limited, New Delhi, 1981.
[7] Code of practice for design loads for buildings and structures, IS: 875(Part-III):1987, published by Bureau of Indian standards.
[8] GeoffryM.Pinfort, Nachshen, crofts and Leggatt in their book "Reinforced concrete Chimneys and towers"
[9] Criteria for design of Reinforced concrete Chimneys, IS: 4998(Part-I):1992, published by Bureau of Indian standards.
[10] IS 1893(Part1) – 2002, Indian Standard Code of Practice for Criteria for Earthquake Resistant Design of Structures. Bureau of Indian Standard, (New Delhi)
[11] K. Suresh Kumar and G. N. V. Rao, Wind Loading over the Top Portions of Tall Stacks with and Without External Landing Platforms, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 51, Page 319-338, July 1993
[12] Lawrence C Maugh and Wadi S. Rumman, Dynamic Design of Reinforced Concrete Chimneys,
[13] ACI Journal, Vol. 64, No. 47, Page 560-567, September 1967
[14] S. R. Joshi, N. S. Pendse, V. T. Patilkakad, Some Special Aspects of the Design and Analysis of Tall Chimneys, Irrigation and Power Journal, Vol. 42, No. 1, January 1985
[15] K. S. Babu Narayan, Subhash C. Yarogal, and Yukio Tamura, "Interaction Envelops For Limit State Design of Chimneys", Fourth International Symposium on Computational Wind Engineering, Yokohama, 2006 [10] J. L. Wilson, Code Recommendation for the aseismic Design of Tall Reinforced Concrete Chimneys, CICIND's Report Australia, Bibao, Vol. 16, No. 2, September 2000
[16] N. S. Pour, Indrajit Chowdhary, Dynamic Soil – Structure Interaction Analysis of Tall Multi flue Chimneys Under Aerodynamic and Seismic Force, Twelfth International Conference of IACMAG India, Goa, Page 1-6, October 2008
[17] Jaiswal, O.R., Srinivas, V., 2005, 'Effect of tuned mass damper on across-wind response of tall RC Chimneys', Journal of Wind & Engineering, India, vol.2, No.1, pp.9-21
[18] Shivaji, M., Raju, V.S.N., 'Dynamic analysis of R.C.C. chimneys'
[19] H.Taibi Zinai, A. Plumier and D. Kerdal, "Computation of Buckling Strength of Reinforced Concrete Columns by the Transfer-Matrix Method", International Journal of Civil Engineering & Technology (IJCIET), Volume 3, Issue 1, 2012, pp. 111 - 127, ISSN Print: 0976 – 6308, ISSN Online: 0976 – 6316.
[20] B.C.Punmia, Ashok K Jain and Arun K Jain, "Reinforced concrete structures- Vol.II", Laxmi Publication (P) Ltd. New Delhi-110002.
[21] Mohammed S. Al-Ansari, "Flexural Safety Cost of Optimized Reinforced Concrete Beams", International Journal of Civil Engineering & Technology (IJCIET), Volume 4, Issue 2, 2013, pp. 15 - 35, ISSN Print: 0976 – 6308, ISSN Online: 0976 – 6316.
[22] A Model code for concrete chimneys, Part-A-The shell (1984)-CICIND, 136 North street, Brighton, England.
[23] Shravya Donkonda and Dr.Devdas Menon, "Optimal design of reinforced concrete retaining walls", The Indian Concrete Journal, Vol.86, No.04, pp. 9-18.
[24] A Flaga and T Lipecki (2010), "Code approaches to vortex shedding and own model", Engineering Structures. 32, pp.1530-1536.
[25] A Kareem and J Hseih (1986), "Reliability analysis of concrete chimneys under wind loading", Journal of Wind Engineering and Industrial Aerodynamics. 25, pp. 93-112.
[26] A Hlaga (1983), "A analysis of along-across and torsional wind effect on slender engineering structures in stochastic formulation", Wydawnictwa politechniki, Monografia No 22, Krakow (in Polish).
[27] A. Castelani (1983), "Construzioni in zona sismica.Milano", Masson Italia Editori.
[28] CICIND, "Model code for steel chimneys (Revision 1-December 1999)", Amendment A, March 2002.
[29] D Menon and PS Rao (1997), "Uncertainties in codal recommendations for across-wind load analysis of R/C chimneys", Journal of Wind Engineering and Industrial Aerodynamics.72, pp. 455-468.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

arXiv:2005.02796v1 [math.CO] 6 May 2020

Domineering games with minimal number of moves

Rohan Karthikeyan Siddharth Sinha
Department of Applied Mathematics
Delhi Technological University
Delhi, India
fromrohank07@gmail.com

May 7, 2020

Abstract

Domineering is a two-player game played on a checkerboard in which one player places dominoes vertically, while the other places them horizontally. In this paper, we find out the minimum number of moves for a game of Domineering to end on several rectangular $m \times n$ boards. We also formulate two conjectures pertaining to $2 \times n$ (similarly $m \times 2$) and on $3 \times n$ (similarly, $m \times 3$) boards.

Mathematics Subject Classification. 91A46; 05C30, 05C57

1 Introduction

Combinatorial games are two-player games with perfect information and no chance moves such as rolling dice or shuffling cards. The rules are defined in such a way that play will always come to an end; in theory we could determine by computer which player would win if both players play perfectly. In practice, however, the complexity of the required search complicates our analysis.

In this paper, we consider the game **Domineering**, invented by Göran Andersson around 1973 and popularized by Martin Gardner [2]. The version introduced by Andersson and Gardner was the 8×8 board. Play consists of the two players alternately placing a 1×2 tile (domino) on adjacent empty squares; Left places vertically and Right places horizontally. The game ends when the player whose turn it is cannot place a piece; the player who cannot place loses - this is the *normal play* condition. Since the board is gradually filled, Domineering is a converging game, the game always ends, and ties are impossible. For more information we would like to direct the attention of the interested reader to the books of Berlekamp et al. [3] and Albert et al. [4].

After stating Huntemann et al.'s [1] results of counting Domineering positions satisfying certain properties, we obtain values for the minimum number of moves for a $m \times n$ Domineering game to end in Section 3, before finishing with further research directions.

2 Counting Domineering positions

This section is adapted from [1]. We are interested in enumerating positions at the end of the game. Therefore, we define the following terms:

Definition 2.1. *A **Right end** position is a position in which Left potentially has moves available but Right has no moves; a **Left end** position is defined similarly.*

Definition 2.2. *A **maximal end** position is a position which is both a left end and a right end - a position in which no player can place a domino.*

2.1 Maximal End Positions

For counting the maximal Domineering positions on a $m \times n$ board, we find the generating function

$$F_{m,n}(x,y) = \sum f(a,b)x^a y^b$$

where $f(a,b)$ is the number of maximal Domineering positions with a vertical (placed by Left) and b horizontal (placed by Right) dominoes.

Theorem 2.3. *The generating function for the maximal position of an $m \times n$ Domineering board is*

$$F_{m,n}(x,y) = \sum_{u \in \{0,1\}^n} \left(M_{0,n} + M'_{0,n} \right)^m (1 + \sum_{i=1}^{n} u_i 3^i, 1)$$

where $M_{0,0} = [1], M_{1,0} = [0], M_{2,0} = [0], M'_{0,0} = [0], M'_{1,0} = [1], M'_{2,0} = [0]$,

$$M_{0,\,(q+1)} = \begin{bmatrix} M_{2,q} & M_{2,q} & xM_{0,q} \\ M_{1,q} & \mathbf{0} & \mathbf{0} \\ M_{0,q} & M_{0,q} & \mathbf{0} \end{bmatrix}, \; M'_{0,\,(q+1)} = \begin{bmatrix} M'_{2,q} & M'_{2,q} & xM'_{0,q} \\ M'_{1,q} & \mathbf{0} & \mathbf{0} \\ M'_{0,q} & M'_{0,q} & \mathbf{0} \end{bmatrix}$$

$$M_{1,\,(q+1)} = \begin{bmatrix} M_{2,q} & M_{2,q} & xM_{0,q} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ M_{0,q} & M_{0,q} & \mathbf{0} \end{bmatrix}, \; M'_{1,\,(q+1)} = \begin{bmatrix} M'_{2,q} & M'_{2,q} & xM'_{0,q} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ M'_{0,q} & M'_{0,q} & \mathbf{0} \end{bmatrix}$$

$$M_{2,\,(q+1)} = \begin{bmatrix} yM_{0,q} & yM_{0,q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \textit{ and } M'_{2,\,(q+1)} = \begin{bmatrix} yM'_{0,q} & yM'_{0,q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

2.2 Left and Right End Positions

For counting the Right end Domineering positions on a $m \times n$ board, we find the generating function

$$R_{m,n}(x,y) = \sum r(a,b)x^a y^b$$

where $r(a,b)$ is the number of Right end positions with a vertical and b horizontal dominoes. The generating polynomial for Left end positions on an $m \times n$ board, which we denote by $L_{m,n}(x,y)$, can be found by obtaining $R_{n,m}(x,y)$ and then switching x and y.

Theorem 2.4. *The generating polynomial of Domineering Right ends on an $m \times n$ board is the $(1,1)$ entry of $(R_{0,n} + R'_{0,n})^m$, denoted by $R_{m,n}(x,y)$, where* $R_{0,0} = [1], R_{1,0} = [0], R_{2,0} = [0], R'_{0,0} = [0], R'_{1,0} = [1], R'_{2,0} = [0]$,

$$R_{0,\,(q+1)} = \begin{bmatrix} R_{1,q} + R_{2,q} & xR_{0,q} \\ R_{0,q} & \mathbf{0} \end{bmatrix}, \; R'_{0,\,(q+1)} = \begin{bmatrix} R'_{1,q} + R'_{2,q} & xR'_{0,q} \\ R'_{0,q} & \mathbf{0} \end{bmatrix}$$

$$R_{1,\,(q+1)} = \begin{bmatrix} R_{2,q} & xR_{0,q} \\ R_{0,q} & \mathbf{0} \end{bmatrix}, \; R'_{1,\,(q+1)} = \begin{bmatrix} R'_{2,q} & xR'_{0,q} \\ R'_{0,q} & \mathbf{0} \end{bmatrix}$$

$$R_{2,\,(q+1)} = \begin{bmatrix} yR_{0,q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \textit{ and } R'_{2,\,(q+1)} = \begin{bmatrix} yR'_{0,q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

3 Results

In this section, we state and discuss the results we've obtained and pose some problems not considered in the literature thus far.

One observes that the presence of a monomial of the form $x^a y^b$ in either the Left end ($L_{m,n}(x,y)$) or Right end ($R_{m,n}(x,y)$) or the Maximal end ($F_{m,n}(x,y)$) polynomials indicates the tiling of the $m \times n$ Domineering board with a vertical (placed by Left) tiles and b horizontal (placed by Right) tiles satisfying the required properties of the position.

It naturally follows that the minimum number of moves for a $m \times n$ Domineering game to end, which we denote by $\alpha_{m,n}$, equals the lowest sum of the degrees of x and y in the Left end, Right end and Maximal end polynomials. That is:

$$\alpha_{m,n} = \min\{a+b : x^a y^b \text{occurs in } L_{m,n}(x,y) \text{ or } R_{m,n}(x,y) \text{ or } F_{m,n}(x,y)\}$$

We have incorporated the matrix recurrence relations for the Left, Right and Maximal end positions defined previously into our program written in Maple. The program works by first determining the terms in $L_{m,n}(x,y), R_{m,n}(x,y)$ and $F_{m,n}(x,y)$ where the difference between the powers of x and y is at most 1. That is because these polynomials actually count all legal Domineering positions satisfying the required properties. Then, we write another procedure to determine the polynomial(s) corresponding to $\alpha_{m,n}$. Additionally, the polynomials were analysed to check whether the term(s) occur in the Left and/or Right and/or Maximal end positions.

We have generated results for all rectangular boards up to size 8×8 and some other rectangular boards. Beyond that point, the computations quickly grows too large to fit in the average home computer's main memory. An overview of the results is given in Table 1.

	n									
m	1	2	3	4	5	6	7	8	9	10
1	1^{LR}	1^{LR}	1^{LR}	1^{LR}	1^{LR}	1^{LR}	1^{LR}	1^{LR}	1^{LR}	1^{LR}
2	1^{LR}	1^{LR}	1^{R}	3^{LR}	3^{LR}	3^{R}	4^{R}	4^{R}	5^{R}	5^{R}
3	1^{LR}	1^{L}	2^{LR}	3^{LR}	3^{L}	4^{L}	5^{LR}	5^{R}	6^{L}	
4	1^{LR}	3^{LR}	3^{LR}	4^{LR}	5^{LR}	7^{LR}	7^{LR}	8^{LR}	9^{LR}	
5	1^{LR}	3^{LR}	3^{R}	5^{LR}	7^{LR}	8^{LR}	9^{LR}	11^{LR}		
6	1^{LR}	3^{L}	4^{R}	7^{LR}	8^{LR}	9^{LR}	11^{LR}	12^{LR}		
7	1^{LR}	4^{L}	5^{LR}	7^{LR}	9^{LR}	11^{LR}	12^{LR}	15^{LR}		
8	1^{LR}	4^{L}	5^{R}	8^{LR}	11^{LR}	12^{LR}	15^{LR}	16^{LR}		
9	1^{LR}	5^{L}	6^{R}	9^{LR}						
10	1^{LR}	5^{L}								

Table 1: The minimum number of moves for $m \times n$ Domineering games to end. The superscript over the numbers denote the polynomials (Left end and/or Right end and/or Maximal end) in which these numbers occur. For instance, the 2×4 game ends in 3 moves and this position occurs in both the Left end and Right end polynomials. Hence, the entry is denoted 3^{LR}.

Some observations are in order:

1. The term with the least degree for all $m \times 1$ and $1 \times n$ boards with $m, n \geq 1$ is actually $x^0 y^0 = 1$. However, by the definition of a combinatorial game, someone

must make a move. Hence, we take the next smallest term which is x for all $m \times 1$ boards with $m > 1$ and y for all $1 \times n$ boards with $n > 1$.

2. The minimum number of moves for a $m \times n$ Domineering game to end is the same as that for a $n \times m$ Domineering game to end, that is, $\alpha_{m,n} = \alpha_{n,m}$.

3. For large m, n, it seems that $\alpha_{m,n}$ occurs in both the Left end and the Right end polynomials.

4. We observe that $\alpha_{m,n}$ coincides with the minimum number of moves obtained in perfect play only for small m, n.

5. We were unable to analyse the maximal end polynomials mainly because the computations quickly grow large even for small m, n.

On close observation, one can notice the following patterns which, to the best of our knowledge, haven't been widely studied:

Problem 3.1. *The minimum number of moves for a $2 \times n$ Domineering game to end is given by*

$$\alpha_{2,\,n} = \left\lceil \frac{n}{2} \right\rceil \ \forall\, n \geq 5.$$

A similar result also holds for $m \times 2$ boards. One can easily construct Domineering positions that use the minimum number of moves shown in Table 1 on $2 \times n$ boards up to $n = 5$. However, we're unable to explicitly construct Domineering positions on $2 \times n$ boards with $n > 5$ using the minimum number of moves for the game to end.

Problem 3.2. *The minimum number of moves for a $3 \times n$ Domineering game to end is given by*

$$\alpha_{3,\,n} = \left\lfloor \frac{2n+1}{3} \right\rfloor \ \forall\, n \geq 1.$$

A similar result also holds for $m \times 3$ boards. We have been able to construct Domineering positions that use the least number of moves shown in Table 1 on $3 \times n$ boards up to $n = 6$.

4 Conclusions and Future research

We have analysed the minimum number of moves for a $m \times n$ Domineering game to end and formulated two new problems. We find two avenues to expand upon the research presented in this paper.

First, one could continue to find values for unsolved Domineering boards using better programs and/or more computer memory. We have already started working in SageMath to analyze this problem. Secondly, we intend to find the minimum number of moves for a $m \times n$ Domineering game to end if both players play perfectly.

5 Acknowledgements

We would like to acknowledge Dr. Chandra Prakash Singh, our thesis advisor, for his constant encouragement during our preparatory studies and in the preparation of this manuscript. We would also like to thank Dr. Samrith Ram of IIIT Delhi for introducing us to this wonderful topic of combinatorial game theory. Last but not the least, we would like to thank our parents for providing moral support all through our lives.

References

[1] S. Huntemann, N. A. McKay. *Counting Domineering positions*. arXiv preprint, arXiv:1909.12419 (2019).

[2] M. Gardner. *Mathematical games: cram, crosscram and quadraphage: new games having elusive winning strategies*. Sci. Am., **230** (1974), 106-108.

[3] E. Berlekamp, J.H. Conway, R.K. Guy. *Winning Ways for your Mathematical Plays*, vol. 1 (2001), vols. 2, 3 (2003), vol. 4 (2004). A K Peters.

[4] M.K. Albert, R.J. Nowakowski, D. Wolfe. *Lessons in Play: An Introduction to Combinatorial Game Theory*, A K Peters, 2007.

Metallography, Microstructure, and Analysis (2020) 9:403–418
https://doi.org/10.1007/s13632-020-00640-7

TECHNICAL ARTICLE

Effect of Friction Stir Processing on Microstructure and Mechanical Properties of TIG Welded Joint of AA6061 and AA7075

Husain Mehdi[1] · **R. S. Mishra**[1]

Received: 21 November 2019 / Revised: 24 February 2020 / Accepted: 7 April 2020 / Published online: 28 April 2020
© ASM International 2020

Abstract
In tungsten inert gas welding (TIG), micro-cracks, porosity, coarse grain structure and high residual stress distribution were found due to persisting thermal conditions. The TIG welded joint is processed using friction stir processing with input process parameters to avoid these defects. In present work, the experimental investigation was conducted to examine the influence of friction stir processing (FSP) on microstructure and mechanical properties of TIG welded joint of AA6061 and AA7075. Tensile test, Vickers hardness test, x-ray diffraction, microscopy and energy-dispersive x-ray test were performed for concluding the optimum set of parameters. The tensile test results shows that the hybrid TIG + FSP welded joint had higher tensile strength than TIG welded joint with filler ER4043, whereas the increment in the micro-hardness of TIG + FSP welded joint was observed. The grain size also decreases when tool pin rotates on TIG welding with different processing parameters. It was found that the maximum tensile stress, % elongation and micro-hardness at nugget zone for TIG + FSP welded joint are 255 MPa, 29.2, and 105 HV respectively. The present investigation demonstrates that the tool rotational speed and traverse speed are the dominating parameters to improve the mechanical properties of TIG welded joint.

Keywords Tensile strength · Micro-hardness · Residual stress · Microstructure

Introduction

Modern development in the automobile sector and defense sector has been transformed from conventional materials to light material such as aluminum alloy. Due to excellent mechanical and physical properties of aluminum alloy such as high strength, excellent corrosion resistance, low density and high thermal conductivity, it is mostly used for making various components such as military aircraft, rocket, and rocket launcher, axle shafts, rims, bumpers and car bodies [1–2]. The friction stir welding (FSW) is used to fabricate the automotive and shipping components, aircraft structural components and reducing structural weight as shown in Fig. 1. This solid-state welding is used in defense and aerospace [3–5]. Many aluminum alloys like AA2219 and AA7075 have large range of applications in defense; it includes fabrication of fuel tank of cryogenic rocket and ammunition hardware [6–9]. The fusion welding (FW) of Al-alloy has a prodigious challenges for researcher. The problems associated with aluminum alloys welded joints are aluminum oxide, coarse grain structure, micro-crack, solidification shrinkage and high residual stress [10]. Due to eutectic melting and re-solidification of the fusion zone (FZ), the brittle inter dendritic structure was found, which influence the low tensile strength, ductility and low hardness [11–13]. Therefore, fusion welding has a major defect during welding process. Friction stir welding has many benefits such as low residual stress and distortion, low welding defects and can also weld thick and thin plates. During welding, the thermal stress induced in the material is a big concern, because these stresses affect microstructure, plastic flow and residual stresses. The residual stresses present in weldment have a strong cause on failures and crack on welded joint [5]. The effects of tool feed rate on residual stresses of FSW of Al-alloy joints were studied with synchrotron x-ray measurement and analyze the residual stress in longitudinal and transverse direction [14]. FSW can affect the performance of the residual stresses in the weldment. Due to low heat input, the residual stress in FSW joint is very low. In FSW, the processing parameters may affect the development of residual stresses [15–16].

✉ Husain Mehdi
husainmehdi4u@gmail.com

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

Many past researchers were studied and investigated about processing parameters of friction stir welding for enhancement of mechanical properties of welded joint [17–23]. Many researchers have been developed the models for predicting the optimized tensile strength FSW welded joint of various aluminum alloy by response surface method [24–26]. Various optimization methods have been used to optimize welding parameters and established empirical relationship between welding parameters and their responses. Most popular methods to solve the optimization problems are RSM using design expert and MINITAB software [27]. The hardness of TIG welded joints was measured to be higher as compared to gas tungsten arc welding (GTAW) due to higher weight percentage of carbon in weld fusion zone [28–29]. Mechanical properties and residual stresses of friction stir welded joint and TIG welded joint of different steel have been studied with various process parameters in different welded zone. The residual stress has been carried out in heat-affected zone and fusion zone by using blind hole drilling method. The residual stress in TIG welded joint was observed higher than friction stir welding. Compressive residual stress was observed in base material, whereas tensile residual stress was observed in heat-affected zone [30].

The effects of rotation speed on microstructure and mechanical properties of welded joints by dual rotation friction stir welding were analyzed and found fine equiaxed recrystallization grains and defect-free joints under various process parameters [31]. A new welding approach of TIG + FSP was successfully applied to the AA2024 to enhance the mechanical properties of welded joint. They conclude that the defects and porosities in the TIG welded joint are completely reduced by FSP process and modified the microstructure and mechanical properties of TIG + FSP welded joint [32]. The application of FSP on TIG welded joint improved the ductility and tensile strength of welded joint of AA5083-H111, and results revealed best reduced dimple size [33].

This work is focused on to enhance the microstructure and mechanical properties of TIG welded joint of AA6061 and AA7075 by FSP process under various processing parameters. This topic is carefully selected after noticing that there is little literature review available on TIG + FSP welding approach. The outcomes of this paper will give the new approach for enhancing the mechanical properties of TIG welded joint which play a huge role in welded strcuturs.

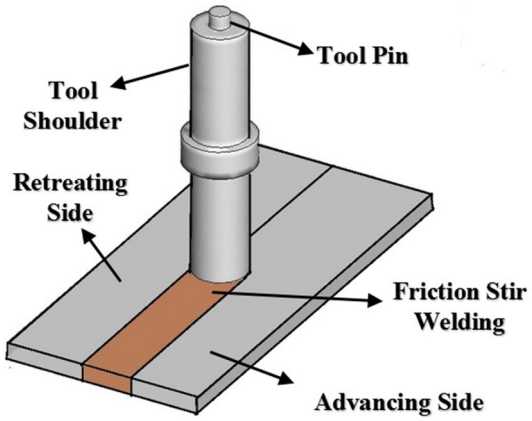


Fig. 1 Friction stir welding

Material and Experiment Work

The base plates of A6061 and AA7075-T6 of 6.2 mm thickness were used for present experimental work. AA7075 is an Al-Zn-Mg alloy which gains higher strength from the precipitation hardening. The chemical composition of AA6061 and AA7075 obtained by energy-dispersive x-ray spectroscopy (EDX) is presented in Table 1.

The filler wire of ER4043 of diameter 2.4 mm was used to join single V groove plate at voltage 23 V and current 110 amp. Argon (22 lit/min) is used as a shielding gas and travel speed 3.2 mm/sec to fabricate the joint.

After TIG welding, FSP technique was used on TIG weldment to improve the microstructure and mechanical properties of the welded joint. The non-consumable H13 tool steel with pin diameter (octagonal cross section), shoulder diameter and pin length of 7 mm, 19.5 mm and 5.5 mm respectively was used. The input parameters range for TIG + FSP welded joint have been taken as rotational speed (1000 rpm to 1300 rpm), traverse speed (30–60 mm/min) and tilt angles (0°–2°). Only defect-free TIG + FSP welded joints were further investigated.

Figure 2 demonstrates the dimension of tensile sub test specimen as per ASTM E8. The tensile test specimen was sliced and machined from the welded joint using milling cutter and shaper machine. Single-pass FSP used to fabricate TIG + FSP welded joint is shown in Fig. 3. Three sub tensile specimens were tested on UTM machine at room

Table 1 Chemical composition of filler wire and parent material

Material	Si	Cu	Fe	Zn	Mg	Mn	Cr	Ti	Al
ER 4043	5.3	0.025	0.8	0.1	0.05	…	…	0.02	Bal.
AA7075	0.58	1.2	0.35	5.1	2.1	0.12	0.2	0.05	Bal.
AA6061	0.4–0.8	0.2	0.7	0.25	0.8	0.1	0.2	0.15	Bal.

temperature, and average of these three results is taken. Vickers micro-hardness machine was used for measuring the hardness across the welded joint with a load of 100 g and dwell time 30 s. Microstructure observation was carried out using scanning electron microscope (SEM) for TIG and TIG + FSP welded joint.

Twenty experiments have been conducted with three independent input variables in which some fabricated joints are shown in Fig. 4. The processing parameters of TIG + FSP and its level are shown in Table 2.

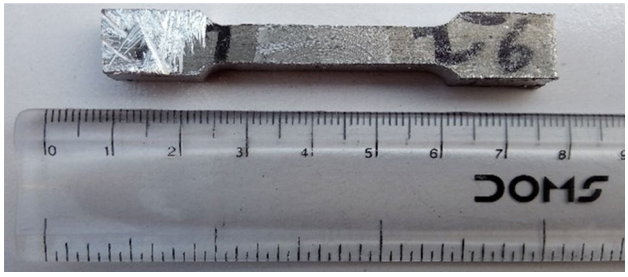


Fig. 2 Dimension of tensile test specimen as per ASTM E8

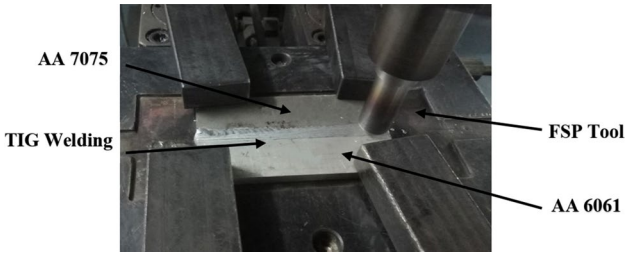


Fig. 3 FSP process after TIG welding

Results and Discussion

Microstructure Analysis Of TIG And TIG + FSP Welded Joint

The microstructure of TIG welded joint of Al-6061 and 7075 with filler ER4043 was observed at different locations using various magnifications. Figure 5 shows the different zones of TIG welded joint such as welded zone, fusion zone and HAZ. Fusion zone was formed due to re-solidification and melting during the welding. The dispersed precipitates of Mg_2Si intermetallic compound were observed in welded region, and coarse equiaxed grains were also observed in fusion zone. The equiaxed grains absent due to the material close to the fusion line provide plenty of sources for the crystal nucleation of the liquid metal during the TIG welding [34]. Fine grains with small precipitates were reported in heat-affected zone (HAZ) as compared to other zone.

Fig. 4 Various TIG + FSP fabricated joint

Table 2 Processing parameter of friction stir processing and its level

Parameters	Units	Notation	Range	Levels		
				− 1	0	1
TRS	rpm	A	1000–1300	1000	1150	1300
TS	mm/min	B	30–60	30	45	60
TA	degree	C	0–2	0	1	2

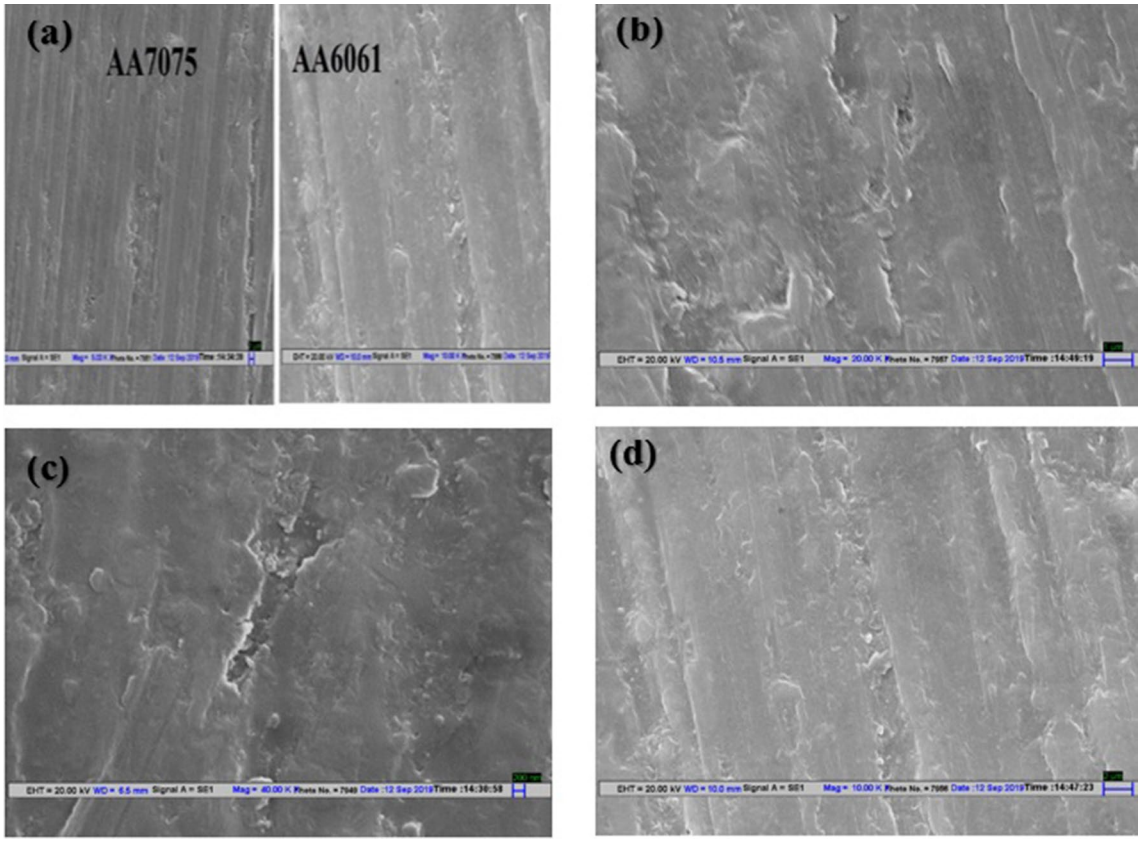


Fig. 5 SEM images of TIG welded joint (**a**) base material, (**b**) welded zone, (**c**) fusion zone, (**d**) HAZ

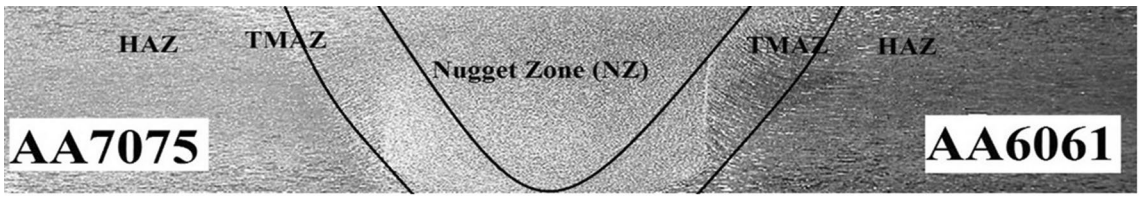


Fig. 6 Optical micrograph of various zone for TIG + FSP welded joint of AA7075 and AA6061

The relationship between heat input and the grain size is measured by electron backscatter diffraction (EBSD) in the stir zone [35]. The mean grain size generally decreased with the heat input under the studies conditions. The revealed relationship was characterized by a significant experimental scattering. This deteriorates their predictive ability and therefore requires a deeper understanding for the heat input and grain growth. The heat input is directly related to the variation of the welding temperature. The final grain size in the stir zone is depended on the peak welding temperature and weakly on the cooling rate [36].

Three different zones have been recognized in TIG + FSP weldment at low magnification due to mechanical and thermal stresses caused by the processing parameters. These zones are nugget zone (NZ), thermo-mechanically affected zone (TMAZ), and heat-affected zone (HAZ) are shown in Fig. 6. The formation of nugget zone shape in TIG + FSP welded joint is recognized to the maximum deformation and plasticization in the material which shows the fine recrystallized equiaxed grains. The formation of nugget shapes depends on thermal gradient, processing parameters and tools geometry in the workpiece [37]. Therefore, coarse grain structure of TIG welded joint is

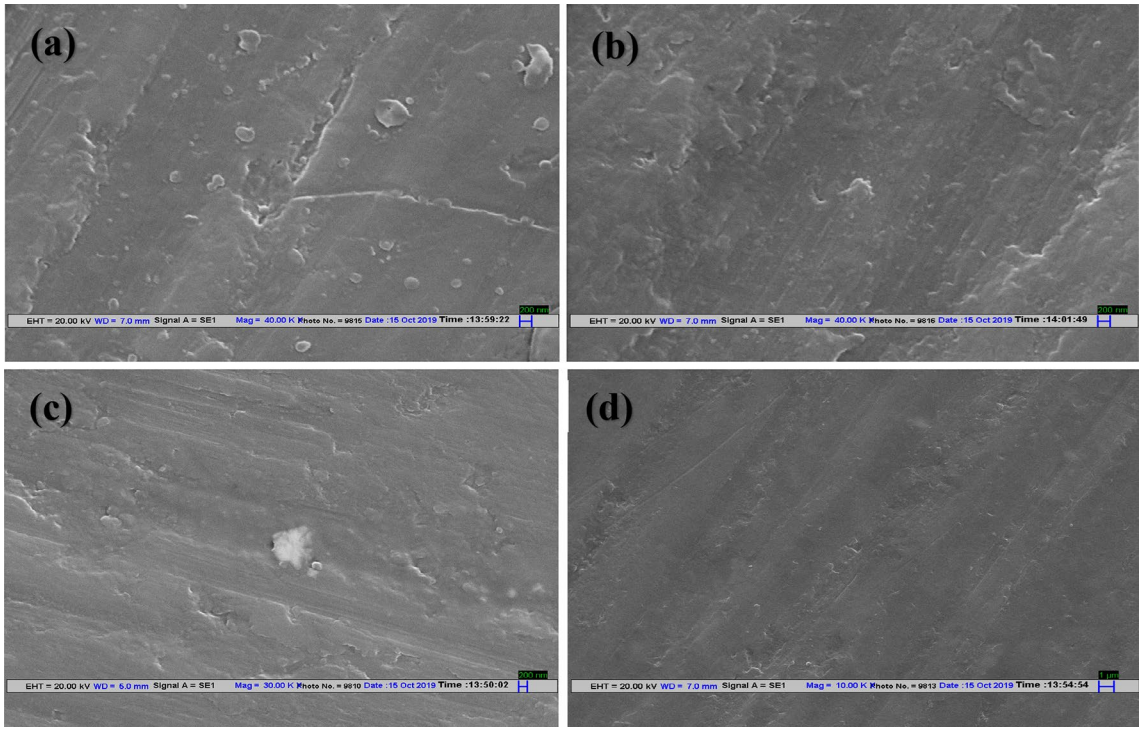


Fig. 7 SEM images of TIG + FSP welded joint (**a**) sample 1, (**b**) sample 5, (**c**) sample 12 and (**d**) sample 18

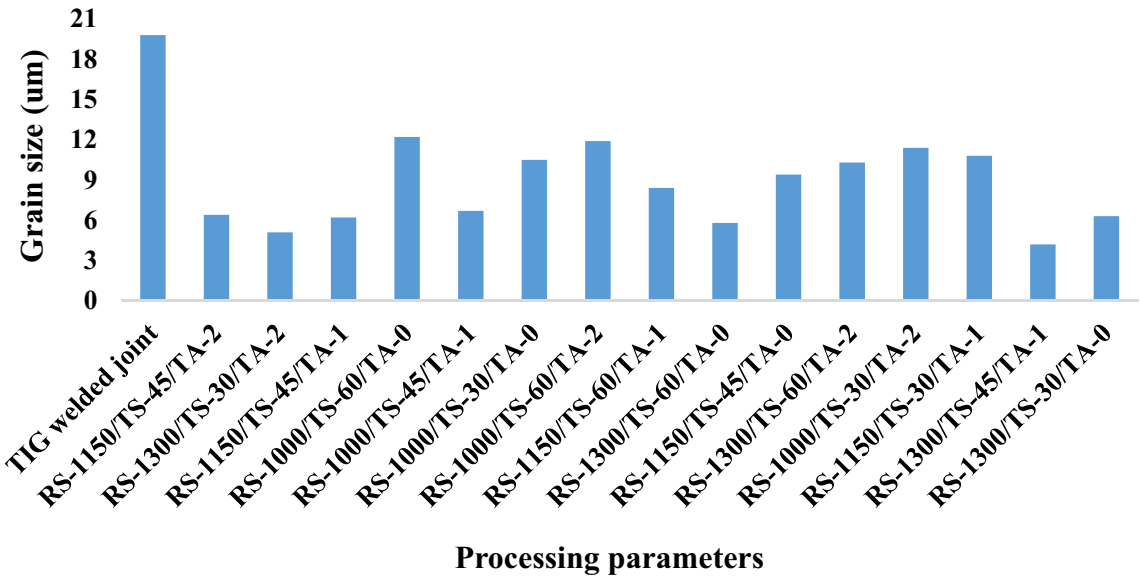


Fig. 8 Variation of grain size at the nugget zone to the processing parameters

transformed into the uniform and fine grains structures in the weld nugget zone due to adequate softening of material revealed the maximum tensile strength and micro-hardness of the TIG + FSP welded joint as shown in Fig. 7. At high welding speed, the weld nugget zone is more homogenous than those produce low welding speed because high heat input gives the effective recrystallization and more homogenous temperature distribution in weld nugget zone. The grains sizes in stir zone change crucially which is depended on the heat input and processing parameters [38–39]. The grain sizes of TIG weldment were analyzed

by the image J software, and observed grain sizes in fusion zone at three different positions are 19.2 µm, 19.8 µm and 18.5 µm, whereas 5.4 µm, 5.9 µm and 4.2 µm were observed in TIG + FSP welded joint at 1300 rpm, 45 mm/min as shown in Fig. 8.

Fracture Surface Analysis

Fractured surfaces of the fragmented tensile specimens were analyzed by SEM machine to understand the effect of microstructure on the failure pattern of TIG and TIG + FSP welded joints. Some important observation can be made. All TIG welded specimens are failed at the welded joint, whereas most of the TIG + FSP welded joint failed on the advancing side, while some specimens were failed at retreating side because fracture was initiated from the interface of TMAZ and stir zone in retreating side. The reasons for such localized fracture behavior are explained by analyzing the texture, grain size and strain localization characteristics in retreating side [40]. The scanning electron microscope (SEM) fractograph has been taken from the fractured tensile test specimen of TIG and TIG + FSP welded joint of AA6061 and AA7075 at room temperature as shown in Fig. 9. The fracture morphology between the TIG and TIG + FSP shows the clear difference that TIG welded portion shows the deep dimples, whereas TIG + FSP welded portion shows shallow dimples with fractured lines; this is the evdence of crack nuleation and growth 4 mm away from the weld line. The small grain particles were found in the TIG + FSP welded zone, while big grains were found in TIG welded region. A number of small and big silicon particles are observed in both fractured specimens. Because of formation and consequent growth and coalescence of cavity or voids, the ductile fracture of welded joint occurs; an improvement in ductility may be achieved when the cavity nucleation could be suppressed [41]. The maximum interfacial normal stress depends upon the grain particle size and the volume fraction of the grain particles [42].

Pandey et al. [43–44] have explained that the coalescence of micro-voids results in equiaxed dimples on tensile fracture surface normal to the loading axis. Hence, the equiaxed and spherical dimples on a flat crater bottom loaded in tension and elongated ellipsoidal dimples on the shear lips oriented at 45°.

SEM fractograph of the fractured tensile test specimen shows the size and distribution of large void and cavity of different TIG + FSP welded joint. The fractured surface is

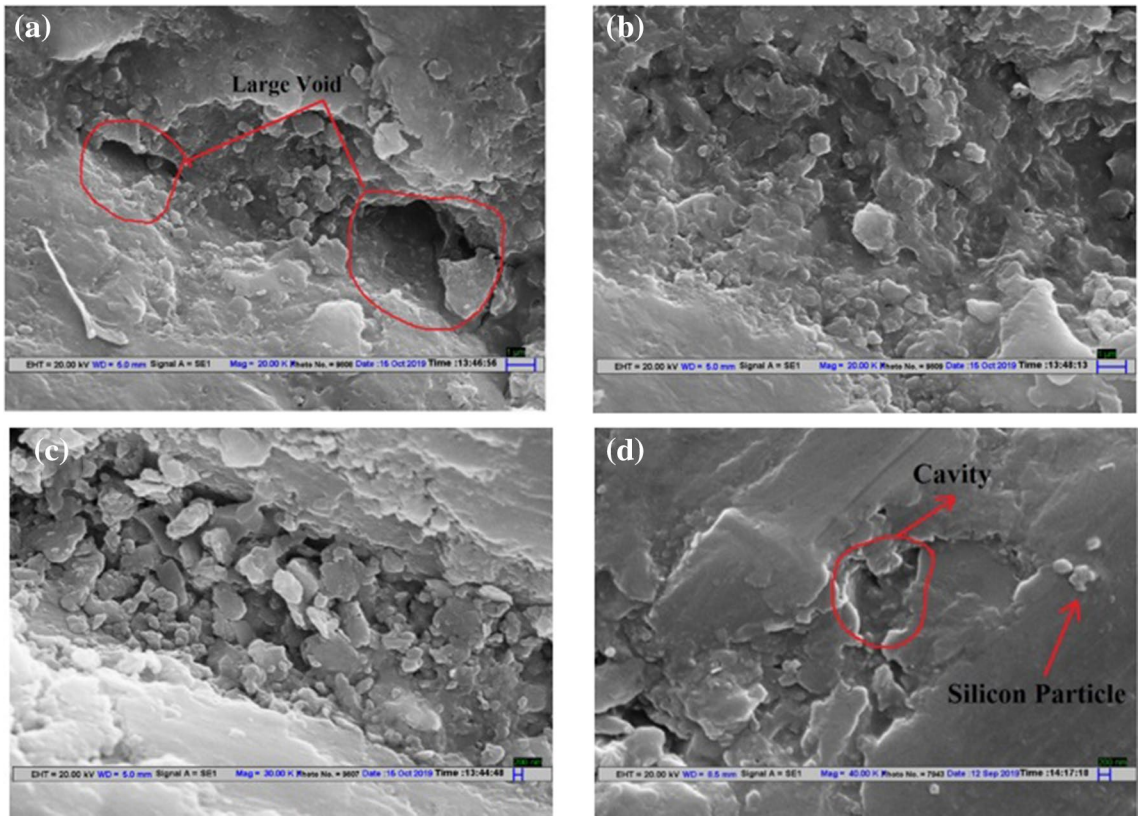


Fig. 9 SEM images of tensile fractured specimen, (**a**) TIG welded joint, (**b**) TIG + FSP welded joint (sample 4), (**c**) TIG + FSP welded joint (sample 9), (**d**) TIG + FSP welded joint (sample 18),

characterized by ductile dimple and some area of large void as shown in Fig. 9. The fine dimple percentage increases by the variation of process parameters of TIG + FSP welded joint. The cavity formation occurred due to precipitate coarsening that layer coalesce led to the shear rupture.

EDX Analysis

The energy-dispersive x-ray spectroscopy (EDX/EDS) of tungsten inert gas welded joint with filler ER4043, and TIG + FSP welded joint has been analyzed. It was found that atomic percentage of Al and Si in the welded joint is higher than the other elements. Figure 10a illustrates the EDX image, while Table 3 shows the percentage of element concentration at fusion zone (FZ) in TIG welded joint with filler ER4043. The percentage of aluminum being highest (88.3%), Si is 4.58% (2nd highest), and Zn is 2.2% (3rd highest) in TIG welded joint, whereas alloying elements are uniformly distributed in TIG + FSP welded joint as shown in Fig. 10b and Table 4. Zinc (Zn), magnesium (Mg) and silicon (Si) elements were found in the weldment besides the aluminum (Al); it is found that Mg, Zn and Si created the phase after the precipitation reaction in the weldment. A very high intensity was found from aluminum, because of fragmentation of precipitates the intensity of Mg_2Si and $MgZn_2$ decreases after friction stir processing on tungsten inert gas welded joint.

XRD Analysis of TIG and TIG + FSP Welded Joint

X-ray diffraction is used for examine the phase detection of dissimilar welded joint AA6061 and AA7075 with filler ER4043 and found three major phases Al, Al_2CuMg and Mg_2Si as shown in Fig. 11. Magnesium (Mg) and silicon (Si) elements were found in the weldment besides the aluminum (Al); it is found that Mg and Si created the phase after the precipitation reaction in the weldment. A very high intensity was found from aluminum, because of fragmentation of precipitates the intensity of Mg_2Si increases after friction stir processing on TIG welded joint with filler ER4043 [45–46]. The alloying elements such as Si and Mg existing in weld center make precipitation reaction and form a strong precipitate of Mg_2Si to give a higher strength. Same phase of Al_2CuMg was detected in both the nugget zone.

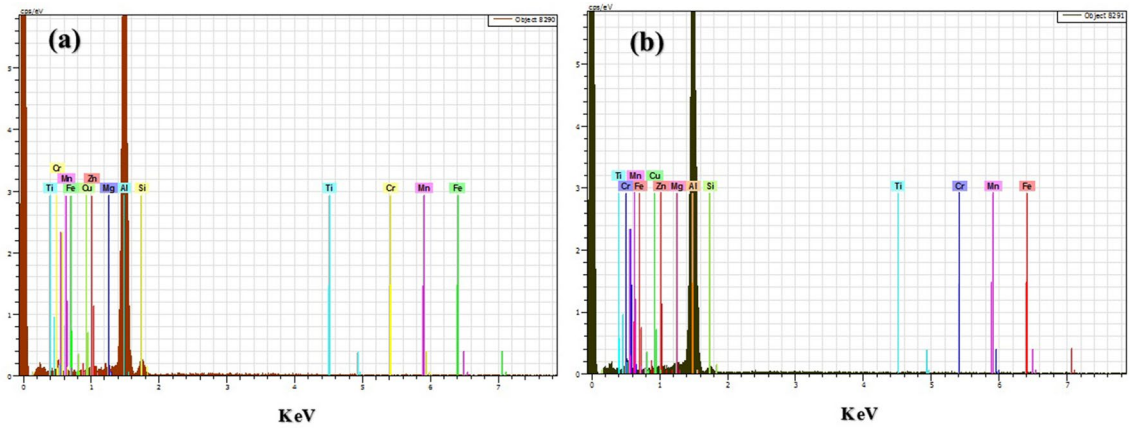


Fig. 10 EDX analysis of (**a**) TIG welded joint, (**b**) TIG + FSP welded joint

Table 3 Concentration of elements in TIG welded joint of AA6061 and AA7075 with filler ER4043 at 3 different points

Elements (%)	Al	Si	Zn	Cu	Fe	Mg	Mn	Cr
Point-1	83.1	3.47	2.81	2.62	2.5	1.77	1.25	0.9
Point-2	81.5	5.94	2.61	2.2	2.04	1.88	1.22	0.96
Point-3	88.3	4.58	2.2	1.84	1.72	1.13	0.86	0.47

Table 4 Concentration of elements in TIG + FSP welded joint for sample 18 at 3 different points

Elements (%)	Al	Si	Zn	Cu	Mg	Fe	Mn	Cr
Point-1	85.88	4.94	2.61	1.76	1.64	0.95	0.69	0.49
Point-2	84.63	4.91	2.53	1.58	1.52	1.12	1.08	0.77
Point-3	85.31	4.52	2.23	2.12	1.82	1.03	0.65	0.51

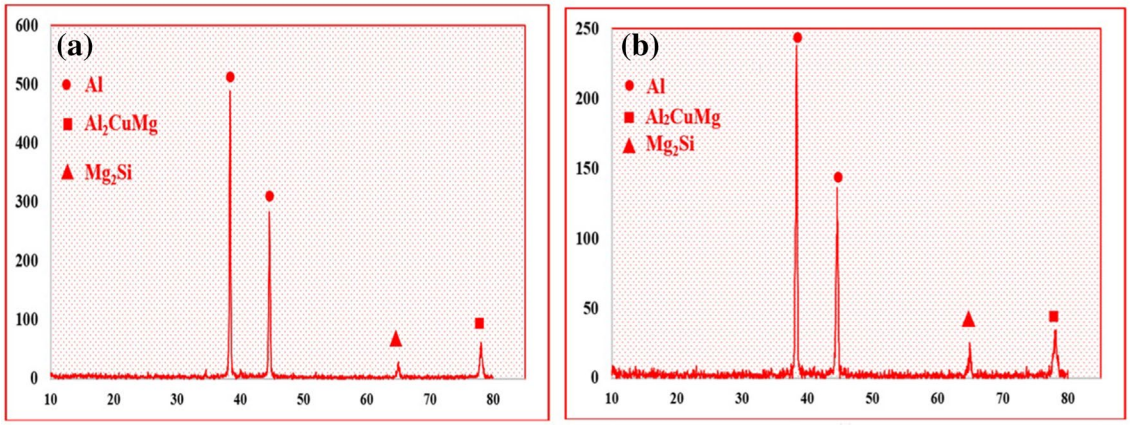


Fig. 11 XRD peaks with filler ER4043, (**a**) TIG welding, (**b**) TIG + FSP welding

Micro-hardness of the TIG + FSP joints is based on boundary energy, brittle intermetallic formation, strain hardening and precipitates formation in the joint.

Fine recrystallized grains and increase in grain boundaries in the (SZ) of AA6061 and AA7075 joints predict higher micro-hardness. Because of fine precipitates and fine grains structure, the stir zone is associated with plastic deformation and high temperature, due to precipitates formation at high temperature along the grain boundaries. When the rotational speed of tool is less than the 1000 rpm, the worm hole at R.S was observed due to insufficient of metal transportation and insufficient heat generation in TIG + FSP welded joint, whereas when the TRS is higher than the 1300 rpm, the high heat was observed in the NZ, results widen the HAZ and weaken the microstructure bonds. Pin holes defect was observed at traverse speed less than 30 mm/min due excessive heat input per unit length in the weldment. When feed rate is greater than 60 mm/min, the bottom tunnel in the advancing side was perceived due to inadequate flow of material and insufficient heat input in the weldment.

Tensile Strength

The influences of FSP on TIG weldment were analyzed with the help of computer controlled UTM machine at room temperature using different processing parameters. According to the hall patch equation $\sigma_1 = \sigma_i + kd^{(-1/2)}$, where d is grain size, σ_1 is yiels stress, σ_i is material constant, and the tensile stress is inversely proportional to the grain size [47]. When the speed of rotating tool increases, the TIG welded metal unswervingly contacted with the tool experienced the most extensive plastic deformation. The number of new grains nucleation increases and forms large number of small grains due to increase in lattice distortion and intragranular dislocation density such that tensile strength of TIG + FSP welded joint increases. When the tool rotational speed exceeded 1300 rpm, the turbulent flow arises at this moment and they could not produce high quality defect-free weld. The elongation of TIG + FSP welded joint is lower than both the base alloys. The decrease in elongation is due to the clustering of the strengthening precipitates, and also effect arising from the localization of strain occurs in the weld region during the friction stir processing [48].

Due to TIG + FSP approach with filler wire ER4043, the porosity, large grain size and presence of extra material are absent in the welded region, which shows the higher tensile strength than the conventional TIG joining process. The tensile strength of TIG welded joint was observed as 176.1 MPa, whereas maximum tensile strength (255 MPa) was found in TIG + FSP welded joint at 1300 rpm, 45 mm/min with tilt angle 1° as shown in Fig. 12 and Table 5.

3D responses surfaces plot and contour of tensile strength are shown in Fig. 13. When the tool rotation speed increases, tensile stress also increases due to strain hardening effect convinced by tool stirring, whereas excess heat input plays the predominant role at high tool rotation speed such as lower tensile strength was observed when the tool rotation decreases. As the feed rate increases, tensile strength first increases and then decreases. The minimum tensile strength, i.e., 185 MPa, was found at lower tool rotation speed (1000 rpm) due to inadequate tool stirring action [9]. The maximum tensile strength (255 MPa) was observed at 1300 rpm, 45 mm/min and 1°, when the feed rate or traverses speed increases, the tensile strength and hardness also increase up to a certain value. Large heat was found in the welded region at lower feed rate. As the feed rate increases, the influences of thermal cycle on the welded joint weakened leading to enhancement in tensile strength of the welded joints as shown in Fig. 13.

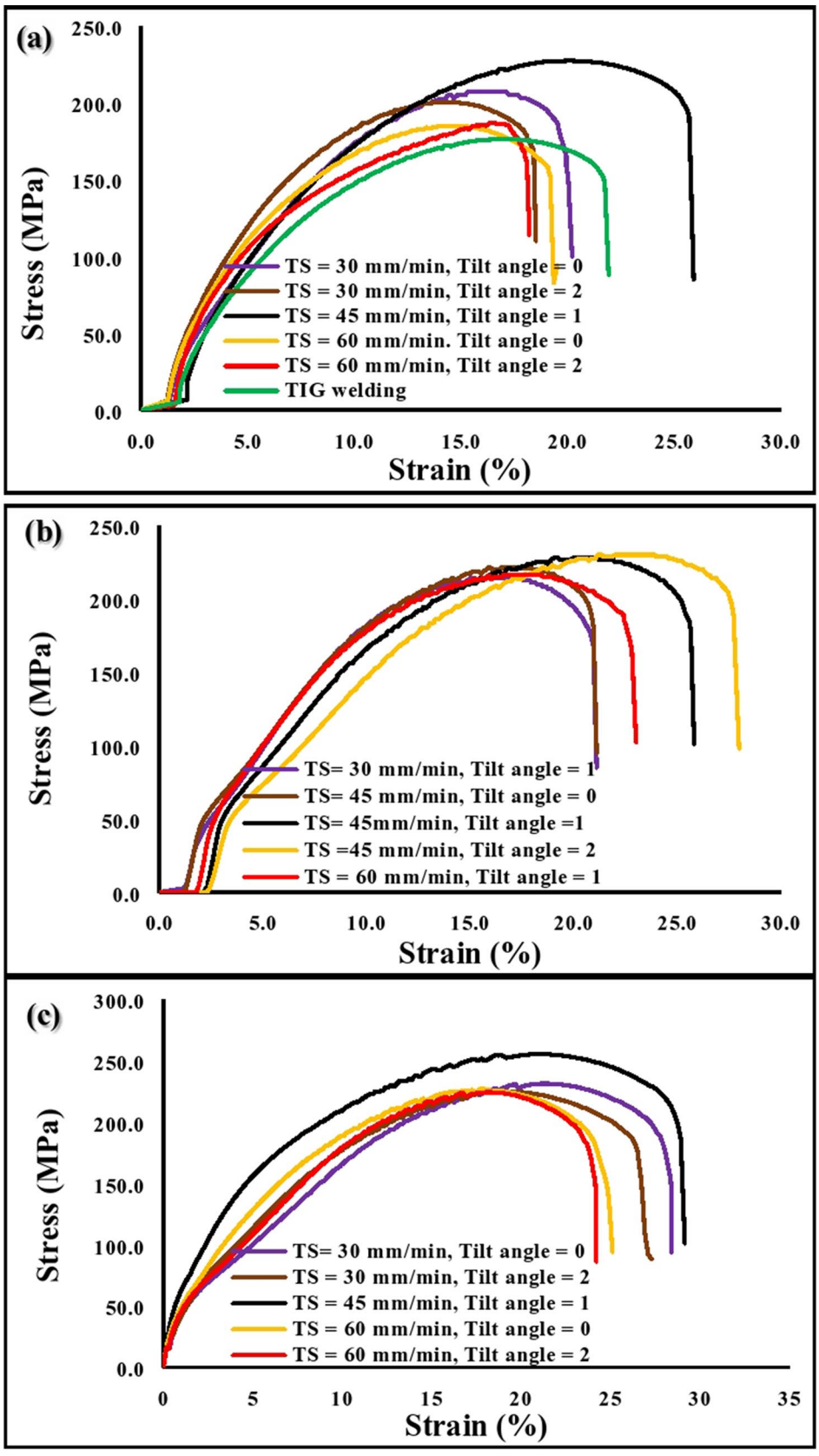


Fig. 12 Stress strain diagram of TIG + FSP welding, (**a**) tool rotation at 1000 rpm, (**b**) tool rotation at 1150 rpm and (c) tool rotation at 1300 rpm

Table 5 Design experiment and their results

Sample	A: Tools rotations speeds (rpm)	B: Traverses speeds (mm/min)	C: Tilts angles (°)	UTS MPa	% *E*	Micro-hardness at nugget (HV)	Grain size (μm)
1	1150	45	2	230	28	100	6.4
2	1300	30	2	225	26.9	97	5.1
3	1150	45	1	225	28.1	95	6.2
4	1000	60	0	185	19.5	77	12.2
5	1000	45	1	228	25.9	94	6.7
6	1150	45	1	232	28.4	99	6.9
7	1000	30	0	207	20.2	81	10.5
8	1000	60	2	187	18.2	78	11.9
9	1150	45	1	233	27.5	101	5.7
10	1150	60	1	219	22.9	89	8.4
11	1300	60	0	226	25.1	93	5.8
12	1150	45	1	228	27.2	98	7.2
13	1150	45	1	232	26.1	99	4.9
14	1150	45	0	221	23.5	91	9.4
15	1300	60	2	224	24.2	91	10.3
16	1000	30	2	201	18.7	74	11.4
17	1150	30	1	214	21.1	87	10.8
18	1300	45	1	255	29.2	105	4.2
19	1300	30	0	231	28.4	101	6.3
20	1150	45	1	228	25.8	99	7.8

Micro-hardness

The micro-hardness distribution of TIG and TIG + FSP welded joint of AA6061 and AA7075 with different processing parameters was analyzed by Vickers hardness testing machine with a load of 100 g and dwell time 30 s are shown in Fig. 14. The micro-hardness values are asymmetrical in the weld center due to unsteady plastic flow and large grain structure in the weld center [49–51]. The alloying elements such as Si and Mg existing in weld center make precipitation reaction and form a strong precipitate of Mg_2Si to give a higher strength as shown in Fig. 11. It showed significant difference, where TIG welded joint using filler ER4043 showed a lower average hardness value compared to TIG + FSP joint. The maximum micro-hardness value, i.e., 105 HV, was obtained at the nugget zone of weldment with processing parameter TRS-1300, TS-45 and TA-1°. During welding, the filler wire ER4043 shows the columnar grains, while fine equiaxed grains are found in TIG + FSP welded joint. It may be analyzed that the fine equiaxed grains improve mechanical properties than the columnar grains [52].

Residual Stress Measurement by cosα Method

For residual stress determination, the cosα method was described [53–54]. The translation from the diffractometer space to the sample is inherently more complex due to the 2D planar geometry of the measurement and can be expressed as

$$q_s = \begin{bmatrix} \cos\eta\sin\psi_0 + \sin\eta\cos\psi_0\cos\alpha \\ \cos\eta\sin\psi_0\sin\phi_0 + \sin\eta\cos\psi_0\sin\phi_0\cos\alpha + \sin\eta\cos\phi_0\sin\alpha \\ \cos\eta\cos\psi_0 + \sin\eta\sin\psi_0\cos\alpha \end{bmatrix} \quad (1)$$

The strain projection along (η, α) coordinates can be written as in terms of scattering vector and strain component as

$$\varepsilon_\alpha = q_i q_j \varepsilon_{ij}$$
$$\varepsilon_{ij} = \frac{1+\nu}{E}\sigma_{ij} - \delta_{ij}\frac{\nu}{E}\sigma_{kk} \quad (2)$$

So, the strain projection may be written as

$$\varepsilon_\alpha = \frac{1+\nu}{E} q_i q_j \sigma_{ij} - \frac{\nu}{E}\sigma_{kk} \quad (3)$$

Now, defining two parameter a_1 and a_2 for linear determination of σ_{11} and σ_{22}

$$\mathrm{a}_1 = \frac{1}{2}\left[\left(\varepsilon_\alpha - \varepsilon_{\pi+\alpha}\right) + \left(\varepsilon_{-\alpha} - \varepsilon_{\pi-\alpha}\right)\right] \quad (4)$$

$$\mathrm{a}_2 = \frac{1}{2}\left[\left(\varepsilon_\alpha - \varepsilon_{\pi+\alpha}\right) - \left(\varepsilon_{-\alpha} - \varepsilon_{\pi-\alpha}\right)\right] \quad (5)$$

After re-expressing of Eqs. (4) and (5) to lead the final relationship for this method.

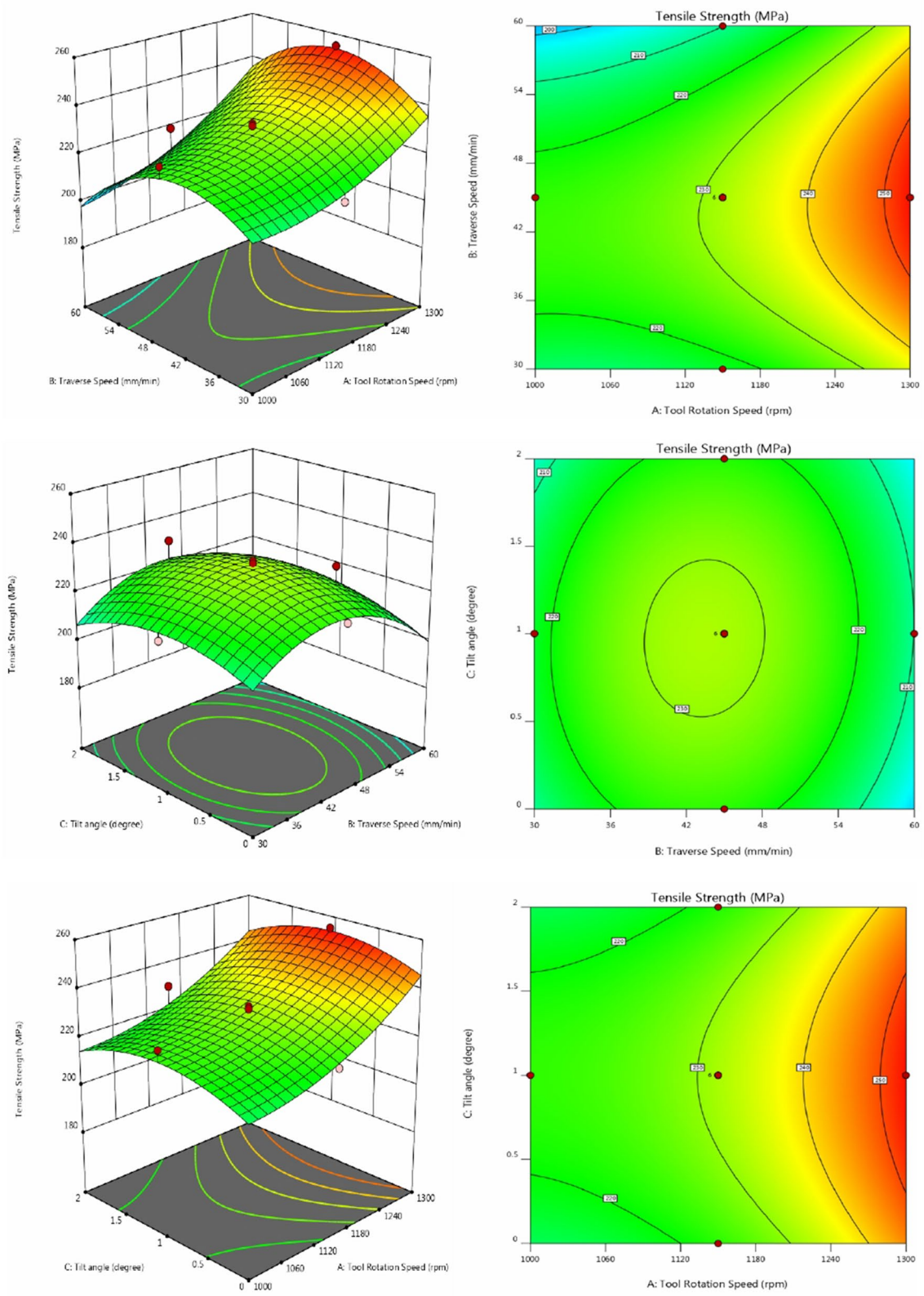


Fig. 13: 3D response surface plot and contour plot for tensile strength of TIG + FSP welded joint

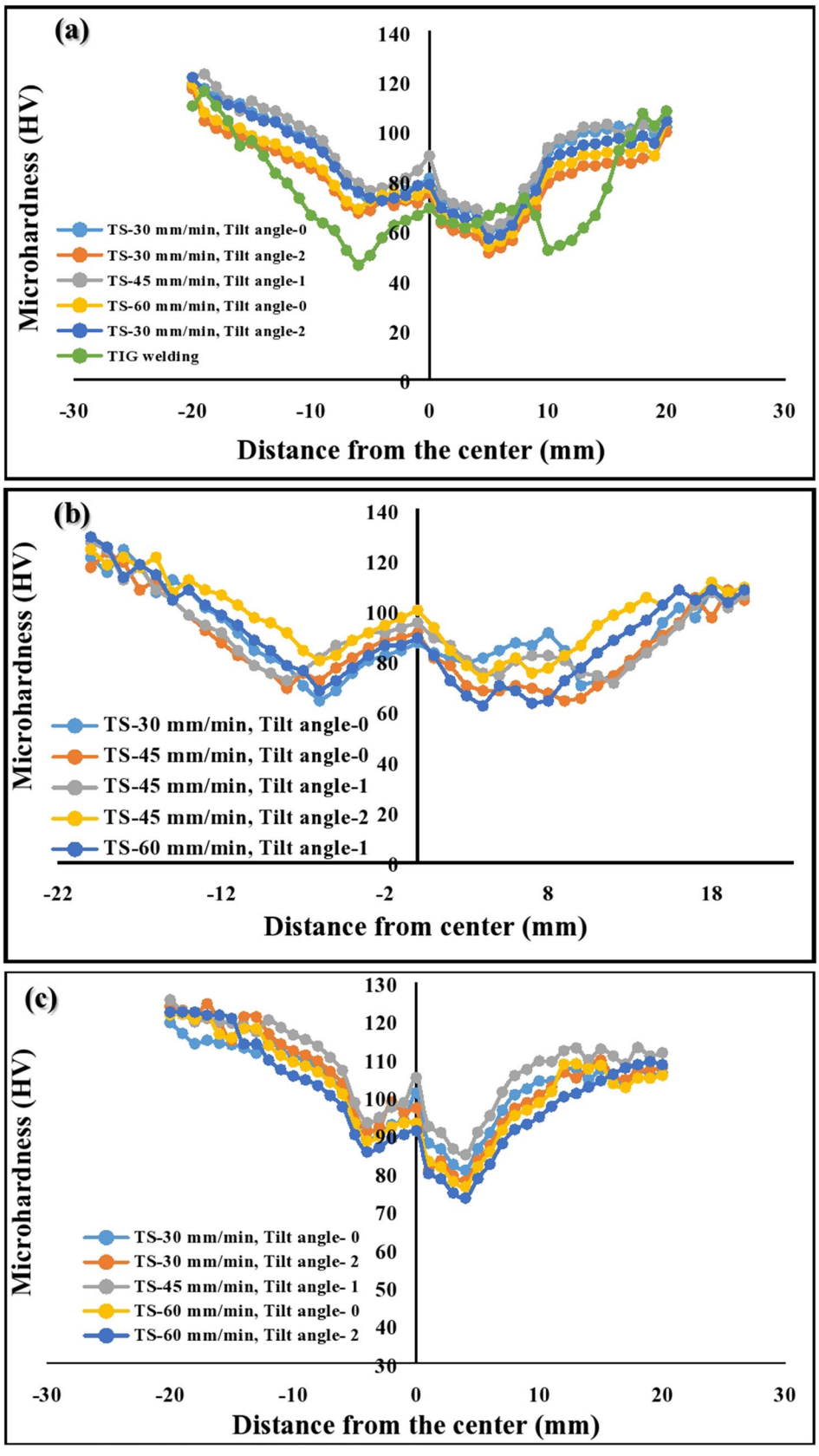


Fig. 14 Distribution of microhardness of welded joint, (**a**) TIG welding, (**b**–**d**) TIG + FSP welded joint, (**b**) tool rotation 1000 rpm, (**c**) tool rotation 1150 rpm and (**d**) tool rotation 1300 rpm

$$a_1 = \frac{1+\nu}{E}\sin 2\psi_0 \sin 2\eta \cos\alpha[\sigma_{11}(1+\cos 2\phi_0) + \sigma_{22}(1-\cos 2\phi_0) + 2\sigma_{12}(\sin 2\phi_0)] \quad (6)$$

$$a_2 = \frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \sin\alpha[\sigma_{22}\sin 2\phi_0 - \sigma_{11}\sin 2\phi_0) + 2\sigma_{12}(\cos 2\phi_0)] \quad (7)$$

at $\phi_0 = 0$, the above equations will be

$$a_1 = \sigma_{11}\frac{1+\nu}{E}\sin 2\psi_0 \sin 2\eta \cos\alpha \quad (8)$$

$$a_2 = 2\sigma_{11}\frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \sin\alpha \quad (9)$$

Thus, the term cosα in Eq. (8) is the origin of the name for this method.

The value of stresses after re-expression may be written as

$$\sigma_{11} = -\frac{E}{1+\nu}\frac{1}{\sin 2\psi_0 \sin 2\eta} \quad (10)$$

$$\sigma_{12} = -\frac{E}{2(1+\nu)}\frac{1}{\sin 2\psi_0 \sin^2\eta} \quad (11)$$

Residual stresses (compressive or tensile) will influence the mechanical behavior of the welded joint. It can reduce brittle fracture strength, buckling strength and cracking in the weldment. Residual stress also influenced the prediction of brittle failure and affected the lifetime prediction of the component [55–57]. Residual stress contributes both negative and positive effects to the weldment. Generally, the tensile residual stress leads to negative effect to the weldment [45, 50]. A mini portable x-ray diffraction apparatus (Pulstec μ-X360) at Delhi technological university, Delhi, India, was used to determine the residual stresses in weldment of AA6061 and AA7075 by the cosα method. The x-ray incident angle was set at 35°, and ± 5 oscillation was applied. The variation of residual stresses in transverse direction of TIG and TIG + FSP welded joint is shown in Fig. 15.

The base material AA7075 on LHS of the weldment shows a minimum compressive residual stress; however, the residual stress gradually increases from base material to toward the weldment and then decreases till second base material AA6061. The maximum compressive residual stress 77 MPa was located at the fusion zone (FZ) of the TIG weldment with filler ER4043, whereas minimum compressive residual stress 29 MPa was obtained at stir zone (SZ) of the TIG + FSP. The residual stresses profile of TIG + FSP was not symmetrical about the centerline of the weldment for both the cases. The left side peak value of the weldment was greater than the right side, because the forming processes are different for retreating side (RS) and advancing side (AS).

Influence of Process Parameters on Residual Stress

The value of residual stress is obtained by the experimental methods. Figure 15 shows a comparison between tool rotation and the compressive residual stress, grouped by traverse speed and tilt angle. It can be seen that the compressive residual stresses decrease for a given value of pitch as the traverse speed and tilt angle increase. According to [14], although measured peak residual stress increases with as traverse speed decreases, tensile stress appears to be limited to the softened weld zone instead, resulting in a narrowing of the tensile region [14]. The minimum value of compressive residual stress (29 MPa) is obtained at tool rotational speed of 1300 rpm, traverse speed 45 mm/min and tilt angle 1°. A 54.6% variation is observed in TIG welding and TIG + FSP welding. The trend in variation of maximum residual stress to the processing parameters is seen in Fig. 15.

Conclusions

The tensile strength, % elongation, micro-hardness at nugget zone and residual stress at nugget zone have been analyzed. The following conclusions are as follows:

- The FSP over the TIG welding completely modified the microstructure and mechanical properties of the TIG welded joint.
- The maximum tensile strength, percentage elongation and micro-hardness at nugget zone are 255 MPa, 29.2 and 105 HV at tool rotation 1300 rpm, traverse speed 45 mm/min and tilt angle 1°, whereas maximum residual stress (12.2 MPa) was found at tool rotation 1000 rpm, traverse speed 60 mm/min and tilt angle 0°.
- The application of FSP on TIG welded joint also improves the ductility of the welded joints; TIG + FSP welded joints are more ductile than the TIG welded joint due to fine grain structure.
- The grains size in the nugget zone in TIG + FSP welded joint was observed much finer (4.2 μm) than the TIG welded joint (19.8 μm).

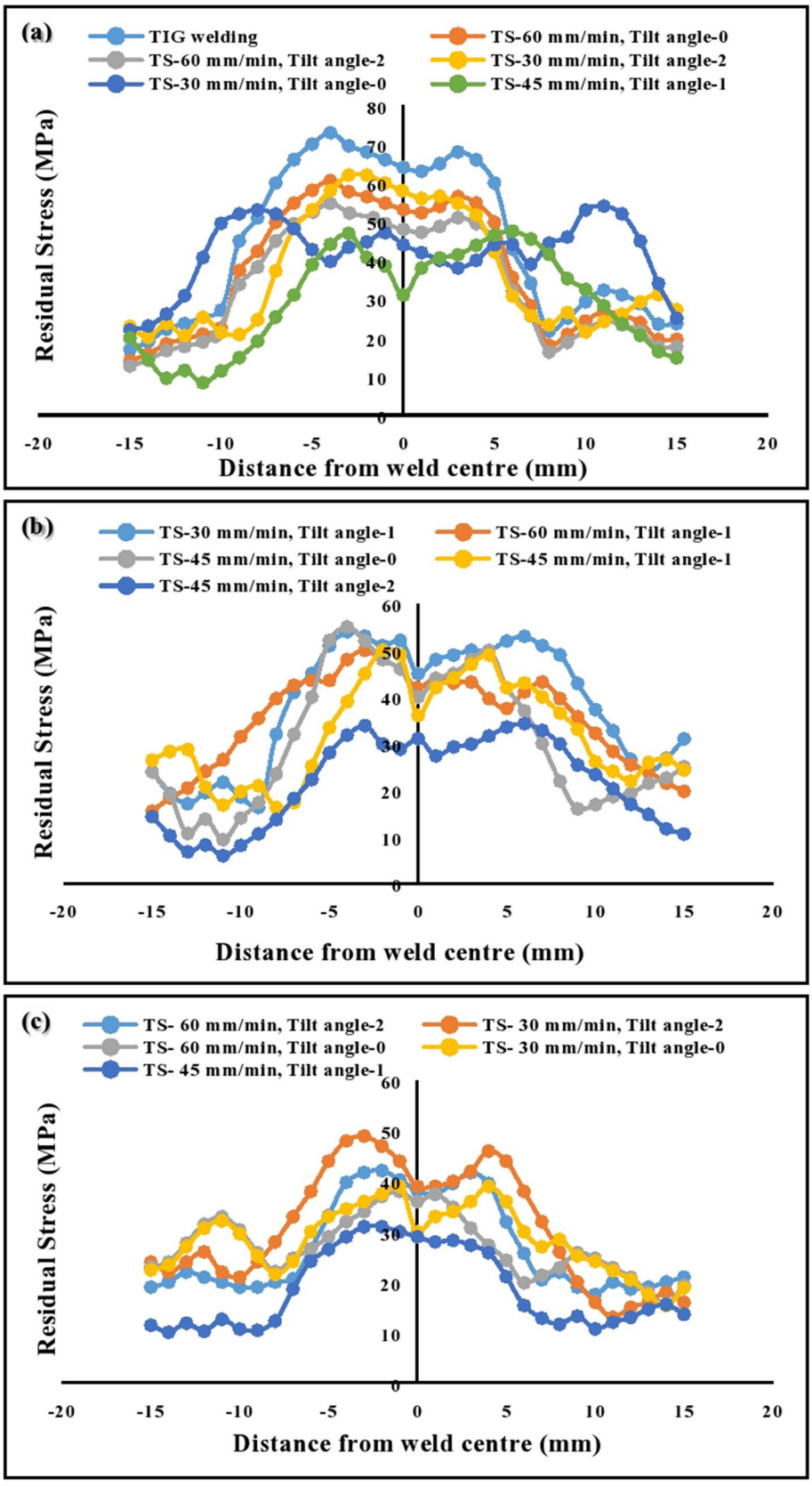


Fig. 15 Variation of residual stress to the processing parameters of TIG + FSP, (**a**) tool rotation 1000 rpm, (**b**) tool rotation 1150 rpm and (**c**) tool rotation 1300 rpm

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

References

1. V.A. Mosneaga, T. Mizutani, T. Kobayashi, H. Toda, Impact toughness of weldments in Al-Mg-Si alloys. Mater. Trans. **43**, 1381–1389 (2002)
2. N. Karunakaran, Effect of pulsed current on temperature distribution and characteristics of GTA welded magnesium alloy. IOSR J. Mech. Civ. Eng. **4**, 1–8 (2013)
3. R.S. Mishra, Z.Y. Ma, Friction stir welding and processing. Mater. Sci. Eng. Res. **50**, 1–78 (2005)
4. M. Zhan, K. Guo, H. Yang, Advances and trends in plastic forming technologies for welded tubes. Chin.J. Aeronaut. **29**, 305–315 (2015)
5. K. Masubuchi, *Analysis of welded structures*, 1st edn. (Pergamon Press, New York, 1980)
6. C. Huang, S. Kou, Partially melted zone in aluminum welds liquation mechanism and directional solidification. Weld. J. **79**, 113–120 (2000)
7. A.P. Biddle, W.A. Wilson, Variable polarity plasma arc welding on the Space Shuttle external tank. Weld. J. **63**, 27–35 (1984)
8. G. Albertini, G. Bruno, B.D. Dunn, F. Fiori, W. Reimers, J.S. Wright, Comparative neutron and x-ray residual stress measurements on Al-2219 welded plate (1997). Mater. Sci. Eng. A **224**, 157–165 (1997)
9. W. Xu, J. Liu, G. Luan, C. Dong, Microstructure and mechanical properties of friction stir welded joints in 2219-T6 aluminum alloy. Mater. Des. **30**, 3460–3467 (2009)
10. R. P. Matrukanitz, in *Selection and weldability of heat-treatable aluminium alloys*. ASM Hand Book Welding, Brazing, and Soldering, vol. 6. (ASM International, Cleveland, 1993) pp. 528–536
11. C.G. Rhodes, M.W. Mahoney, W.H. Bingel, Effects of friction stir welding on microstructure of 7075 aluminium. Scr. Mater. **36**, 69–75 (1997)
12. J.Q. Su, T.W. Nelson, R. Mishra, M. Mahoney, Microstructural investigation of friction stir welded 7050-T651 aluminium. Acta Mater. **51**, 713–729 (2003)
13. M. Zhan, K. Guo, H. Yang, Advances and trends in plastic forming technologies for welded tubes. Chin. J. Aeronaut. **29**, 305–315 (2015)
14. M. Peel, A. Steuwer, M. Preuss, P.J. Withers, Microstructure, mechanical properties and residual stresses as a function of welding speed in aluminium AA5083 friction stir welds. Acta Mater. **51**, 4791–4801 (2003)
15. T. Li, Q. Y. Shi, H. K. Li, W. Wang, and Z. P. Cai, in *Residual stresses of friction stir welded 2024 T-4 joints*. International Welding Joining Conference-Korea (2007).
16. M.A. Sutton, A.P. Reynolds, D.Q. Wang, C.R. Hubbard, A Study of residual stresses and microstructure in 2024-T3 aluminum friction stir butt welds. J. Eng. Mater. Technol. **124**, 215–221 (2002)
17. W.B. Lee, Mechanical properties related to micro structural variation of 6061 Al alloy joints by friction stir welding. J. Mater. Trans. **45**, 1700 (2004)
18. K. Elangovan, V. Balasubramanian, Influences of pin profile and rotational speed of the tool on the formation of friction stir processing zone in AA2219 aluminium alloy. J. Mater. Sci. Eng. A **459**, 7–18 (2007)
19. K. Elangovan, V. Balasubramanian, Influences of tool pin profile and axial force on the formation of friction stir processing zone in AA6061 aluminium alloy. Int. J. Adv. Manuf. Technol. **38**, 285–295 (2008)
20. H. Mehdi, R.S. Mishra, Mechanical properties and microstructure studies in Friction Stir Welding (FSW) joints of dissimilar alloy—a review. J. Achiev. Mater. Manuf. Eng. **77**, 31–40 (2016)
21. H. Mehdi, R.S. Mishra, Influences of process parameter and microstructural studies in friction stir welding of different alloys: a review. Int J. Adv. Prod. Ind. Eng. **5**, 55–62 (2017)
22. H. Mehdi, R.S. Mishra, Mechanical and microstructure characterization of friction stir welding for dissimilar alloy—a review. Int. J. Res. Eng. Innov. **1**, 57–67 (2017)
23. Y.K. Periyasamy, A.V. Perumal, B.K. Periyasamy, Optimization of process parameters on friction stir welding of AA7075-T651 and AA6061 joint using response surface methodology. Mater. Res. Express **6**, 96558–96584 (2019)
24. M. Jayaraman, R. Sivasubramanian, V. Balasubramanian, A.K. Lakshminarayanan, Prediction of tensile strength of friction stir welded A356 cast aluminium alloy using response surface methodology and artificial neural network. J. Manuf. Sci. Prod. Res. **9**, 1–21 (2008)
25. S. Rajakumar, C. Muralidharan, V. Balasubramanian, Optimization of the friction stir- welding process and the tool parameters to attain a maximum tensile strength of AA7075-T6 aluminium alloy. J. Eng. Manuf. **224**, 1175–1191 (2010)
26. A.K. Lakshminarayanan, V. Balasubramanian, Process parameters optimization for friction stir welding of RDE-40 aluminum alloy using Taguchi technique, Transactions of Nonferrous Metals Society of China. Trans. Nonferrous Met. Soc. China **18**, 548–554 (2008)
27. H. G. Neddermeijer, G. J.van Oortmarssen, N. Piersma, R. Dekker, in *A framework for response surface methodology for simulation optimization*. Proceeding Winning Simulation Conference, vol. 128 2000. 129 136
28. C. Pandey, M.M. Mahapatra, P. Kumar, F. Daniel, N. Saini, Dissimilar joining of CSEF steels using autogenous tungsten-inert gas welding and gas tungsten arc welding and their effect on δ-ferrite evolution and mechanical properties. J. Manuf. Proc. **31**, 247–259 (2018)
29. C. Pandey, M.M. Mahapatra, P. Kumar, N. Saini, Comparative study of autogenous tungsten inert gas welding and tungsten arc welding with filler wire for dissimilar P91 and P92 steel weld joint. Mater. Sci. Eng. A **712**, 720–737 (2017)
30. Pk Chaurasia, C. Pandey, A. Giri, N. Saini, M.M. Mahapatra, A comparative study of transverse shrinkage stresses and residual stresses in P91 welded pipe including plasticity error. Arch. Metall. Mater. **63**, 1000–1011 (2018)
31. L. Zhou, R.X. Zhang, X.Y. Hu, N. Guo, H.H. Zhao, Y.X. Huang, Effects of rotation speed of assisted shoulder on microstructure and mechanical properties of 6061-T6 aluminum alloy by dual-rotation friction stir welding. Int. J. Adv. Manuf. Technol. **100**, 199–208 (2018)
32. K. Devireddy, V. Devuri, M. Cheepu, B.K. Kumar, Analysis of the influence of friction stir processing on gas tungsten arc welding of 2024 auminum alloy weld zone. Int. J. Mech. Prod. Eng. Res. Dev. **8**, 243–252 (2018)
33. S. Mabuwa, V. Msomi, Review on friction stir processed TIG and friction stir welded dissimilar alloy joints. Adv. Mater. Sci. Eng. **2019**, 142–159 (2019)
34. S. Kou, *Welding Metallurgy*, 2nd edn. (Willy, New York, 2003)
35. W. Tang, X. Guo, J.V. Mclure, L.E. Murr, Heat input and temperature distribution in friction stir welding. J. Mater. Process. Manuf. Sci. **7**, 163–172 (1998)
36. D. Yi, T. Onuma, S. Mironov, Y.S. Sato, H. Kokawa, Evaluation of heat input during friction stir welding of aluminium alloys. Sci. Technol. Weld. Join. **22**, 41–46 (2016)

37. M.W. Mahoney, C.G. Rhodes, J.G. Flintoff, W.H. Bingel, R.A. Spurling, Properties of friction-stir-welded 7075 T651 aluminum. Metall. Mater. Trans. A **29**, 1955–1964 (1998)
38. H.J. Mohammadzadeh, H. Farahani, M.K.G. Besharati, M.V. Aghaei, Study on the effects of friction stir welding process parameters on the microstructure and mechanical properties of 5086-H34 aluminum welded joints. Int. J. Adv. Manuf. Technol. **83**, 611–621 (2016)
39. Ş. Kasman, Z. Yenier, Analyzing dissimilar friction stir welding of AA5754/AA7075. Int. J. Adv. Manuf. Technol. **70**, 145–156 (2014)
40. G. Liu, R. Xin, J. Li, D. Liu, Q. Liu, Fracture localization in retreating side of friction stir welded magnesium alloy. Sci. Tech. Weld. Join. **205**, 378–384 (2015)
41. S.H. Good, I.M. Brown, The nucleation of cavities by plastic deformation. Acta Mater. **27**, 1–15 (1979)
42. T.C. Tszeng, Interfacial stresses and void nucleation in discontinuously reinforced composites. J. Eng. Mater. Technol. **122**, 86–92 (2000)
43. C. Pandey, M.M. Mahapatra, P. Kumar, P. Kumar, N. Saini, J.G. Thakare, S. Kumar, Study on effect of double austenitization treatment on fracture morphology tensile tested nuclear grade P92 steel. Eng. Fail. Anal. **96**, 158–167 (2018)
44. C. Pandey, N. Saini, M.M. Mahapatrab, P. Kumar, Study of the fracture surface morphology of impact and tensile tested cast and forged (C&F) Grade 91 steel at room temperature for different heat treatment regimes. Eng. Fail. Anal. **71**, 131–147 (2017)
45. V. Patel, V. Badheka, A. Kumar, Influence of friction stir processed parameters on superplasticity of Al-Zn-Mg-Cu alloy. Mater. Manuf. Process. **31**, 1573–1582 (2016)
46. K.T. Senthil, V. Balasubramanian, M.Y. Sanavullah, Influences of pulsed current tungsten inert gas welding parameters on the tensile properties of AA 6061 aluminium alloy. Mater. Des. **28**, 2080–2092 (2007)
47. Y. Li, A.J. Bushby, D.J. Dunstan, The Hall–Petch effect as a manifestation of the general size effect. Proc. R. Soc. A **472**, 20150890 (2016)
48. S.N. Shanmuga, N. Murugan, Tensile behavior of dissimilar friction stir welded joints of aluminium alloys. Mater. Des. **31**, 4184–4193 (2010)
49. W. Xu, J. Liu, G. Luan, C. Dong, Temperature evolution, microstructure and mechanical properties of friction stir welded thick 2219-O aluminum alloy joints. Mater. Des. **30**, 1886–1893 (2009)
50. W. Xu, J. Liu, G. Luan, C. Dong, Microstructure and mechanical properties of friction stir welded joints in 2219–T6 aluminum alloy. Mater. Des. **30**, 3460–3467 (2009)
51. T. Luijendijk, Welding of dissimilar aluminium alloys. J. Mater. Proc. Technol. **103**, 29–35 (2000)
52. M. Ishak, N.F.M. Noordin, L.H.A. Shah, Feasibility study on joining dissimilar aluminum alloys AA6061 and AA7075 by tungsten inert gas (TIG). Jurnal Teknolo **75**, 79–84 (2015)
53. R. Nandan, G.G. Roy, T.J. Lienert, T. Debroy, Three-dimensional heat and material flow during friction stir welding of mild steel. Acta Mater. **55**, 883–895 (2007)
54. S. Taira, K. Tanaka, T. Yamazaki, A method of x-ray microbeam measurement of local stress and its application to fatigue crack growth problem. J. Soc. Mater. Sci. Jpn. **27**, 251–256 (1978)
55. K. Jitender, S. Hari, 2018 Modelling and analysis of process parameters in friction stir welding of AA5083-H321 using response surface methodology. Adv. Mater. Proc. Technol. **4**, 183–199 (2018)
56. W. J. Arbegast, Modeling friction stir joining as a metal working process, in *Symposium, Hot Deformation of Aluminum Alloys*, San Diego, CA, 2003, pp. 313–330
57. M.N.M. Salleh, M. Ishak, L.H. Shah, S.R.A. Idris, The effect of ER4043 and ER5356 filler metal on welded Al 7075 by metal inert gas welding. WIT Trans. Built. Environ. **166**, 213–224 (2016)

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

Journal of Fluorescence
https://doi.org/10.1007/s10895-020-02540-z

ORIGINAL ARTICLE

Flavones Fluorescence-Based Dual Response Chemosensor for Metal Ions in Aqueous Media and Fluorescence Recovery

Nisha Fatma[1] · Mohan Singh Mehata[2] · Nupur Pandey[1] · Sanjay Pant[1]

Received: 2 February 2020 / Accepted: 13 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Responsiveness of sensing materials 3-hydroxyflavone (3HF) towards metal ions in aqueous medium has been explored following photoexcitation. 3HF exhibited both colorimetric and fluorescence (FL) turn–off response towards Cu^{2+} and Fe^{2+} with high sensitivity and selectivity. Meanwhile, the distinct colour change and the rapid quenching of FL intensity provide naked-eye detection. On successive addition of $Cu^{2+}$and Fe^{2+} ions, FL of 3HF was "turned off," whereas there is no change in wavelength of FL bands. Quenching efficiencies for $Cu^{2+}$and Fe^{2+} ions are 88% and 49%, respectively. The detection limit of the sensor towards Cu^{2+} and Fe^{2+} was 1.54 μM and 1.98 μM, respectively. The binding strategy between 3HF and metal ions (Cu^{2+}, Fe^{2+}) and the nature of quenching have been explored with the Benesi-Hildebrand and Stern-Volmer plots, respectively. The FL of 3HF significantly quenched in the presence of Cu^{2+} ions, and then recovered upon addition of HCl, providing the possibility of constructing a sensitive Cu^{2+}-HCl off-on fluorescent probe. Moreover, the proposed simple, quick response and visual test strip-based chemosensor could be used for the detection of Fe^{2+} and Cu^{2+} ions.

Keywords Hydroxyflavone · Chemosensor · Metal ions · Colorimetric · Dual response

Introduction

Fluorescent chemosensor strategy has emerged as a research area of significant importance in a variety of fields such as biology, physiology, pharmacology and environmental sciences. Its characteristics of smooth operation, high sensitivity, instantaneous response and real-time monitoring make it convenient over other optical sensors [1, 2]. Fluorescent chemosensors are renowned kind of molecular systems that react with the analyte to induce a marked change in various photophysical properties, e.g., fluorescence (FL) intensity, wavelength, lifetime and anisotropy [3–5]. Sensing based on intensity is appropriate for excitation and emission spectra because many probes display changes in quantum yield on binding with analyte. Probes with large Stokes-shift (which reduces the light scattering effect and increases the subtlety of detection) are favorable for intensity-based sensing. Intensiometric sensors are recommended for qualitative and semi-quantitative detection technologies where high sensitivity is desirable. Intensity-based sensing offers the merit of absolute sensitivity compared to other methods based on spectrum resolution, lifetime, or anisotropy, where a significant loss of intensity occurs [6, 7].

Metal selective fluorescent probes serve as a potential research domain for their explicit effects in ecological, biological and clinical applications [1, 8]. During past decade numerous sensors based on fluorescent dyes such as Schiff bases [1, 14, 15], quinolines [2, 9, 10], pyrene [4], rhodamine [11, 12] and metal nanoclusters [13] have been developed for transition metal ions. Owing to their real-time chemical analysis and inherent high responsiveness to stimuli, the sensing materials exhibit various photophysical phenomena, viz., excited-state intramolecular proton transfer (ESIPT), metal to ligand charge transfer (MLCT), photoinduced electron transfer (PET), twisted intramolecular charge transfer (TICT), intramolecular

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10895-020-02540-z) contains supplementary material, which is available to authorized users.

✉ Mohan Singh Mehata
msmehata@gmail.com

✉ Sanjay Pant
sanjayphotophys@gmail.com

[1] Photophysics Laboratory, Department of Physics, Centre of Advance Study, DSB Campus, Kumaun University, Nainital 263002, India

[2] Laser Spectroscopy Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 15 May 2020

charge transfer (ICT), fluorescence resonance energy transfer (FRET) and excimer/exciplex formation [6, 7, 16]. Apart from these phenomena, colorimetric sensing has received enormous attention because it allows naked eye recognition of ions without any sophisticated instrumentation, which leads to on-site and real-time detection [4, 17–20].

Given the crucial role of iron and copper ions in biological and environmental applications, probes for sensing of Cu^{2+} and Fe^{2+} are of particular interest. Copper and iron are essential trace elements by their versatility and cytotoxicity [20, 21]. Copper is the third element that exists in a large amount in human body. Its insufficiency, along with excursive levels, stimulates several problems to human health, such as Menke's, Wilson's and Alzheimer's diseases. Higher concentration level causes ischemic heart disease, bone disorder, kidney disease. Also, the existence of copper provides aid in cardiovascular disease, gastrointestinal disorder and diabetes [2, 3, 16, 17, 22]. However, Fe^{2+} actively participates in oxygen uptake, oxygen metabolism and electronic transfer. The inadequacy of Fe^{2+} in a body causes prion diseases, liver and kidney damage, Parkinson's disease, heart failure, etc. [2, 4, 24]. Therefore, trace amount determination of Cu^{2+} and Fe^{2+} ions in the life systems and the environment has attracted general attention.

In recent literature, a large number of probes have been reported and investigated for Cu^{2+} [8, 11, 13–15, 17, 22, 23] and Fe^{2+} [24–27]. Mostly, fluorescent probes focus on detecting individual analyte selectively. Simple sensors capable of recognizing multiple target ions in environmental and biological instances have the advantages of being highly efficient, cost-effective and easy to operate [28–33]. Recent reports on fluorescence sensors for contemporary detection of Cu^{2+} and Fe^{2+}consist of different fluorophores based on quinoline [2], pyrene [4], rhodamine [18], aldazine [19], carbazole [20], coumarin [21], etc.

Flavones framework consists of 2-phenyl-1-benzopyran-4 one moiety. Flavones and its derivatives are a conscientious group of organic molecules in the sphere of natural medicines, biosciences and pharmaceutical activities. The inherent properties of flavones such as antioxidant, antiviral, antibacterial, anti-inflammatory, anticancer, angioprotective and low toxicity are of paramount importance for many scientific centers [34–36]. Among variously substituted flavones derivatives, photophysical characteristics of hydroxyl flavones have been studied extensively over the past few years. Hydroxyflavone (HF) is a captivating array of organic fluorophores that play a prominent role in the analytical chemistry and designing of probes [37–41].

3-hydroxyflavone is a prototype for excited-state intramolecular phototautomerization in various kinds of solvents. The –OH group, including the effects of the environment (polarity, hydrogen bond, pH, and temperature), influences the dual fluorescence emission [42–49]. Studies on excited-state relaxation processes show that 3HF is an exquisitely sensitive environmental probe to microenvironments (micelles, RTIL's, model membranes) [40–52]. 3HF inhibits chelating capability to form stable complexes with transition metal ions since oxygen of carbonyl and –OH group behave as a donor atom. Metal complexes of hydroxyl-substituted flavones possess different physicochemical properties (increased solubility, hydrophilicity) and biological features as compared to that of ligand alone. Literature reported about 3HF explored the relationships between its metal complexes, physicochemical properties and biological character and further expanded the applications of metal complexes in DNA, proteins and lipids [53–56]. There are few reports on fluorescence recognition of d-block metal ions with hydroxyflavones. 3HF has been reported as a sensor for Zn^{2+} ions in methanol and one of its derivatives was synthesized and used as a sensor for $Zn^{2+}$and Cu^{2+} ions [57–59]. In the view of sensing practices, emphasis should be placed on reducing the effect of organic solvents in the development of a probe.

In the present work, we have employed 3HF as a dual response probe for Cu^{2+} and Fe^{2+} ions in the aqueous medium. Spectroscopic measurements confirmed the sensing mechanism of probe 3HF in the presence of Cu^{2+} and Fe^{2+} ions. With the addition of Cu^{2+} and Fe^{2+} ions, 3HF shows the colorimetric response in visible light. Cu^{2+} and Fe^{2+} ions quench the fluorescence intensity of 3HF by ~74 fold and ~ 36 fold, respectively. Test-strips pre-stained with 3HF were prepared. They also demonstrate sensing ability towards Cu^{2+} and Fe^{2+} ions with the naked eye. Interestingly, on adding HCl, the quenched FL of 3HF metal complex was recovered immediately and stabilized within a few minutes, indicating a fast recovery of the system.

Experimental Section

Materials

3-Hydroxyflavone of 98–99% purity was acquired from Sigma Aldrich, USA and used as received. Copper sulphate pentahydrate and all the metal salts (99% purity) were purchased from Loba Chemie and Sisco Research Laboratories (SRL) Pvt. Ltd. and all reagents were used without additional distillation. All the solvents used were of spectroscopic grades and checked for their fluorescence purity. Distilled-deionized water was used throughout the test. All the inorganic salts solutions were dissolved in distilled water to make 1 mM concentration. The stock solution of 3HFwas prepared in MeOH: H_2O by diluting 60 μM of 3HF in MeOH with distilled water (1:2, v/v) to obtain 20 μM of 3HF. Titration experiments of 3HF (20 μM) were performed in MeOH: H_2O by the addition of microliters of copper sulphate solutions (1 mM). The absorption and FL spectra of these solutions

were recorded after each addition of Cu^{2+} (λ_{ex} = 345 nm). The same process was followed for the titration of Fe^{2+} with a probe.

Instrumentation

UV–visible absorption and fluorescence spectroscopy is the most commonly used mode for identifying physiologically and environmentally active probes, taking in to account their high susceptibility and smooth functioning. UV/VIS/NIR spectrophotometer (Lambda-750 Perkin Elmer) has been used to record absorption spectra at room temperature. FL and FL-excitation spectra were measured by using Horiba Fluorolog-3 spectrofluorometer equipped with double grating for both excitation and emission path and 450 W xenon lamp for excitation and data were analyzed by Origin software. Transparent quartz cuvettes of 1 cm path length have been used for the UV-visible absorption and FL measurements. Slit widths for excitation and emission were chosen as 2 nm and 1 nm for all the measurements, respectively. However, for the titration of Fe^{2+} ions, the slit widths were selected as 2 nm for both excitation and emission.

Results and Discussion

Selectivity

3HF has a strong tendency to bind with metal ions. As a result, the α-hydroxyl carbonyl group of hydroxyflavone forms metal chelates following the deprotonation of hydroxyl [57]. UV-visible absorption and fluorescence spectra of 3HF have been measured in MeOH:H_2O (1:2) upon addition of 50 μM of different metal ions (Na^+, Ca^{2+}, Mg^{2+}, Al^{3+}, Zn^{2+}, Fe^{2+}, Pb^{2+}, Ag^+, Mn^{2+}, K^+ and Cu^{2+}) to examine the selectivity.

The absorption spectrum of 3HF exhibits a strong band centered at 345 nm with a shoulder at 310 nm and a minor hump at 410 nm, and is inconsistent with the reported values [42]. Among the distinct metal ions studied, variations in the absorption spectra were observed in the presence of selected metal ions with reduced optical density (of absorption maximum) followed by the appearance of new absorption bands in a longer wavelength region, 400–500 nm, as shown in Fig. S1 of supplementary information (SI). The variation in the absorption intensity in the presence of selected metal ions, Cu^{2+} and Fe^{2+} is shown in Fig. 1(a). These absorption bands of 3HF-Cu^{2+} and 3HF-Fe^{2+}induces the colour change of solution. However, $Zn^{2+}$and Pb^{2+} do not display any visible colour change in a solution of 3HF. Thus, the results indicate that probe 3HF can be utilized for the visible bare eye detection of Cu^{2+} and Fe^{2+} ions, as evidenced by appearing obvious yellow and brown colors, respectively (Fig.1 (b)). The color change noticed in the probe solution upon binding to metal

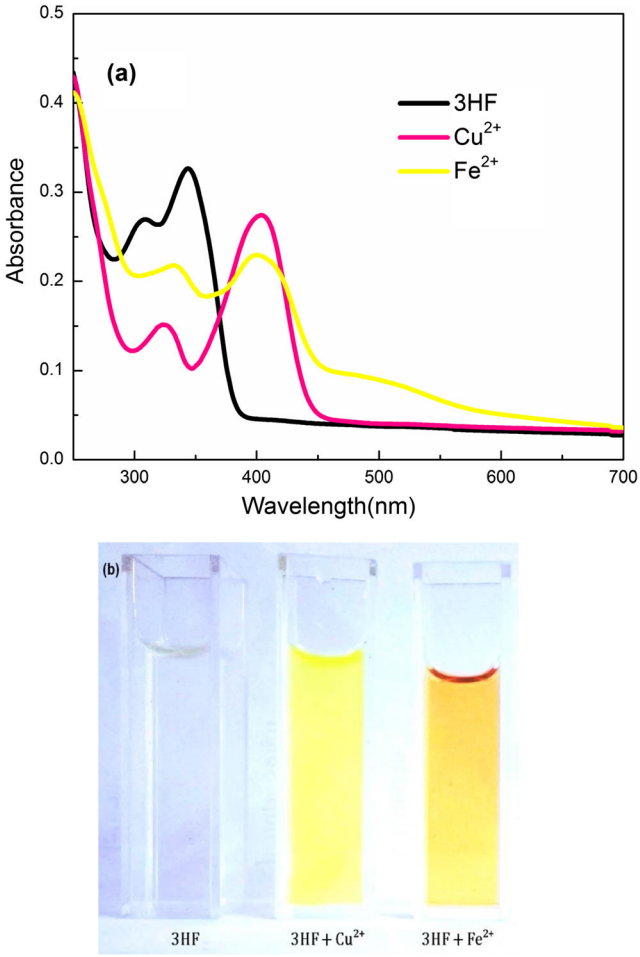


Fig. 1 Absorption spectra (**a**) and images under visible light (**b**) of 3HF (20 μM) in the presence of Cu^{2+} and Fe^{2+} ions (50 μM)

ions (Cu^{2+} and Fe^{2+}) are due to the detachment of proton(H^+) of phenolic OH- which allows the charge transfer from the probe to metal ions.

Excitation of 3HF by 345 nm showed two FL bands peaked at 409 nm corresponds to the N^* (normal form) and 513 nm to the T^* (tautomer form) (Fig.2a). The FL properties of 3HF were also studied upon addition of alkali and transition metal ions Na^+, Ca^{2+}, Mg^{2+}, Al^{3+}, Zn^{2+}, Fe^{2+}, Pb^{2+}, Ag^+, Mn^{2+}, K^+ and Cu^{2+} (50 μM) (Fig. S2 of SI). The FL of the 3HF was almost completely quenched on the addition of 50 μM of Cu^{2+} and Fe^{2+} ions (Fig. 2(b)). It is worth mentioning that other metal ions (at 50 μM concentration) have a negligible effect at 513 nm (Fig. S2 of SI). The FL maximum of 3HF was slightly blue-shifted in the vicinity of $Zn^{2+}$and Pb^{2+}. The Zn^{2+} and Pb^{2+}cation's with close shell d-orbitals do not introduce low-energy metal-centered or charge-separated excited energy states due to which electron transfer processes not feasible [3]. The FL-excitation spectrum of probe resembles the absorption spectrum; even in the presence of Cu^{2+} and Fe^{2+} the peak remains unaltered and there is no variation in the excitation spectrum (Figs. S3 and S4 of SI).

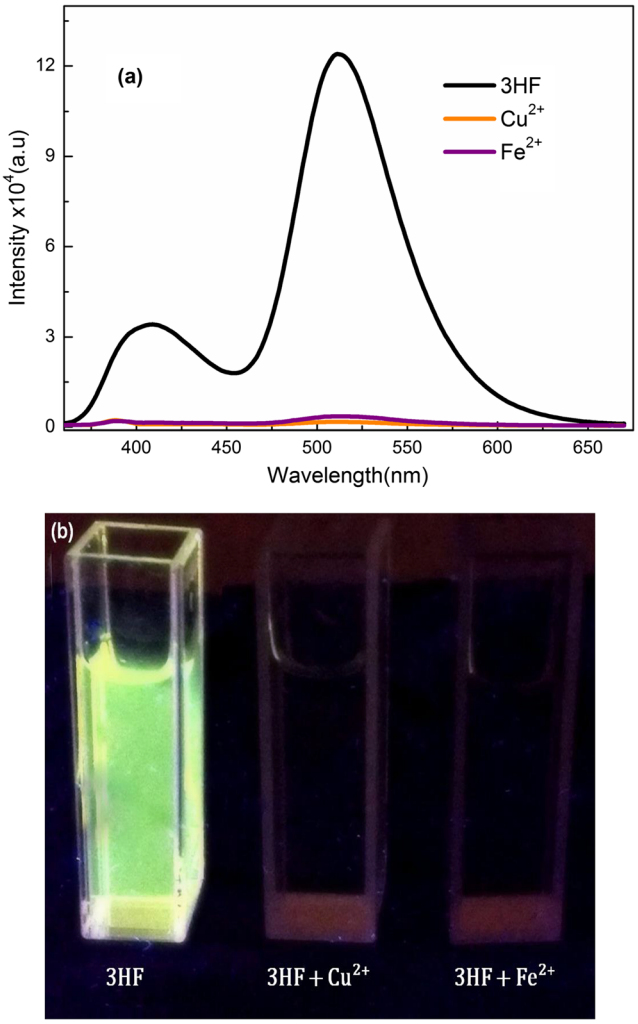


Fig. 2 Fluorescence spectra (**a**) and the response of the test solution under UV-light (**b**) of 3HF (20 μM) in the presence of Cu^{2+} and Fe^{2+}ions. The excitation wavelength (λ_{ex}) was 345 nm

Figure 3 illustrates the comparative response of different metal ions on the FL of 3HF. It indicates that the magnitude of quenching by Cu^{2+} and Fe^{2+} is much higher than changes observed for Zn^{2+} and Pb^{2+}. It is remarkable that 50 μM (0.124 *ppm*) of Cu^{2+} and Fe^{2+} quenches the FL intensity of 3HF ~74 fold and ~ 36 fold, respectively. Therefore, 3HF could also act as a fluorescence turn off probe for Cu^{2+} and Fe^{2+} ions.

Sensitivity

The sensitivity of 3HF towards Cu^{2+} and Fe^{2+} is accessed by titration of 3HF on adding micromolar conentration of Cu^{2+} and Fe^{2+} and recording its absorption and FL spectra at room temperature. The UV-visible absorption spectra of 3HF exhibit a continuous decrease of optical density (O.D.) at 345 nm on a progressive increment of Cu^{2+} ions concentration (Fig.4) along with the increase in absorbance of the long-wavelength region from 370 to 700 nm. The change in absorbance and formation of the isosbestic point at 372 nm indicates the formation of the ground state complex, i.e., towards the existence of two ground-state species. Moreover, similar effects on the absorption spectrum of 3HF were also observed upon the addition of Fe^{2+} (Fig. 5). On the inclusion of micromolar concentration of Fe^{2+} ions, decrease in absorbance of 3HF, the formation of the isosbestic point at 372 nm and a newly developed band at 403 nm were observed. The broad absorption band between 400 and 450 nm is assigned to the solvent-mediated deprotonation reaction in alcoholic solvents in the ground state [3]. In the present study, two complexes could be formed in the ground state due to the deprotonation between the probe and metal ions.

To analyze the complexation of 3HF with Cu^{2+} and Fe^{2+} the plots of $[1/(A_0\text{-}A)]$ as a function of 1/[MI] were obtained using the well-known Benesi-Hildebrand (B-H) expression:

$$\frac{1}{(A_0-A)} = \frac{1}{\{K(A_0-A_{max})[MI^{x+}]^n\}} + \frac{1}{[A_{max}-A_0]} \quad (1)$$

where A_0 is the absorbance of the probe before adding metal ions and A is the absorbance of the probe with different metal ion concentrations and MI represents the concentration of metal ions ($Cu^{2+}$and Fe^{2+}) used throughout the study [9].

The linear relationship obtained in the B-H plot for both the cations indicates that the complexation of Cu^{2+} and Fe^{2+} towards 3HF is 1:1 in nature. The association constant (K_a) of 3HF with Cu^{2+} and Fe^{2+} calculated from a ratio of intercept and slope (intercept/slope) of these linear plots are $7.14 \times 10^4\ M^{-1}$ (R = 0.99) and $2.18 \times 10^4\ M^{-1}$ (R = 0.97), respectively (Fig. 6).The 1:1 stoichiometry of complex between 3HF and Cu^{2+} adequately reflect the reported literature at 343 nm under acidic conditions (pH = 4) [60].

Furthermore, the fluorescence titration studies of 3HF were performed for both Cu^{2+} and Fe^{2+} ions. The FL intensity of 409 nm, as well as the tautomer bands (513 nm) was decreased gradually with an increase in the concentration of Cu^{2+} and Fe^{2+} ions (0 − 48 μM). Almost complete quenching was observed with the addition of 48.3 μM Cu^{2+} ions (Fig.7), whereas for Fe^{2+} ions, this extent of quenching is achieved by adding 64.5 μM of Fe^{2+} ions (Fig. 8). No remarkable spectral shift in FL maximum of 3HF was observed on the formation of 3HF-Cu^{2+} and 3HF-Fe^{2+} complexes; this is also examined by measuring the FL-excitation spectra. The FL detection limit for Cu^{2+} and Fe^{2+} ions are 1.54 μM and 1.98 μM, respectively, calculated by 3σ/k, where the symbols σ and k, respectively represents the standard deviation of the blank measurement and the slope of a plot of fluorescence intensity vs. concentration of metal ions (Figs. S5 and S6 of SI). The calculated LOD's (limit of detections) are well below the safety level set for drinking water by the *Environmental Protection Agency (EPA)*. According to EPA guidelines, maximum contaminant levels for Cu^{2+} and Fe^{2+} in

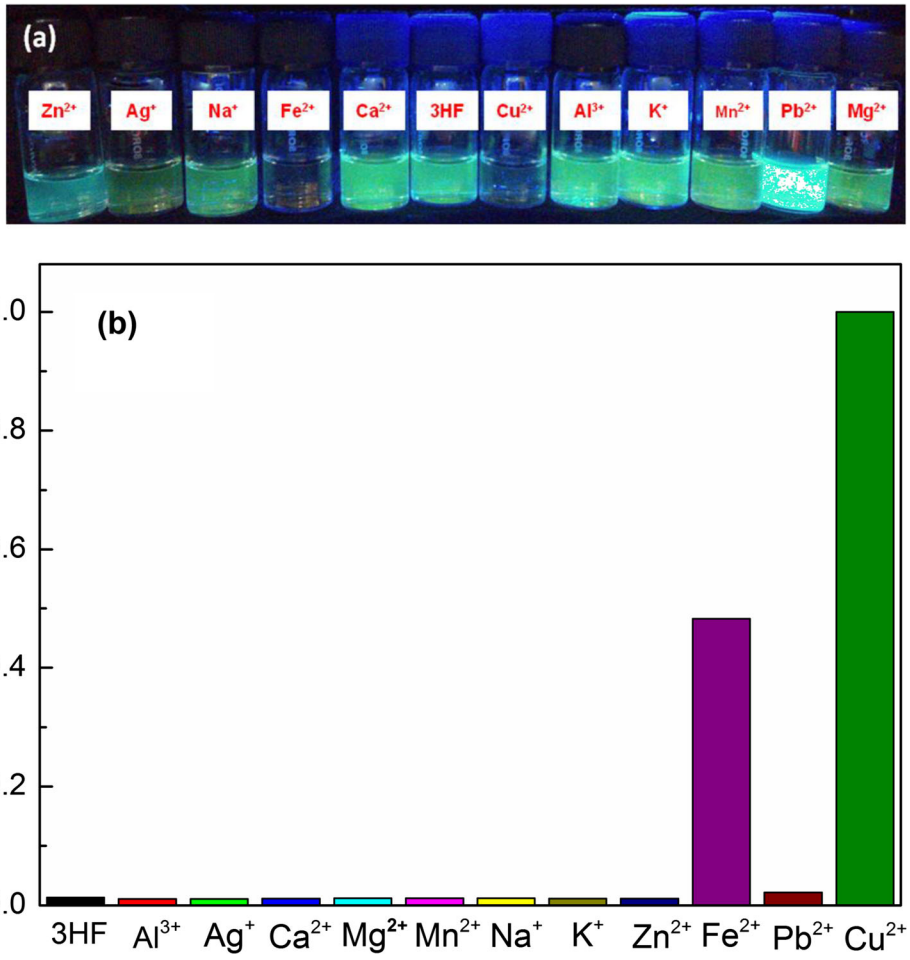


Fig. 3 The response of the test solution of 3HF in the presence of different metals ions of the same concentration under UV-light (**a**) and fluorescence intensity ratio of 3HF in the presence of various metal ions (50 μM) (**b**)

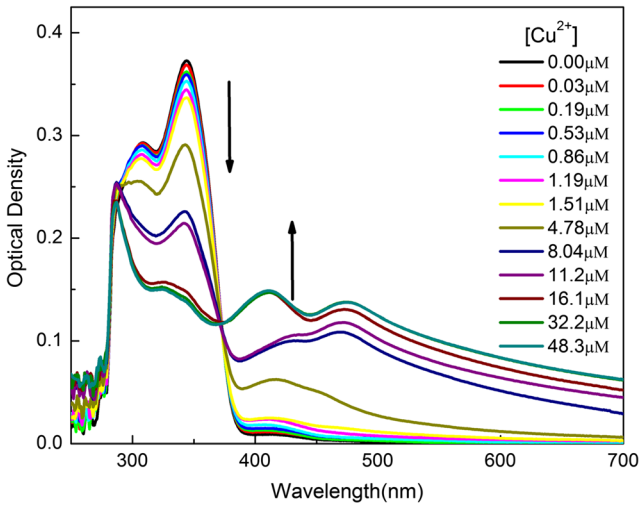


Fig. 4 UV-visible absorption spectra of 3HF on the successive addition of Cu^{2+} ions

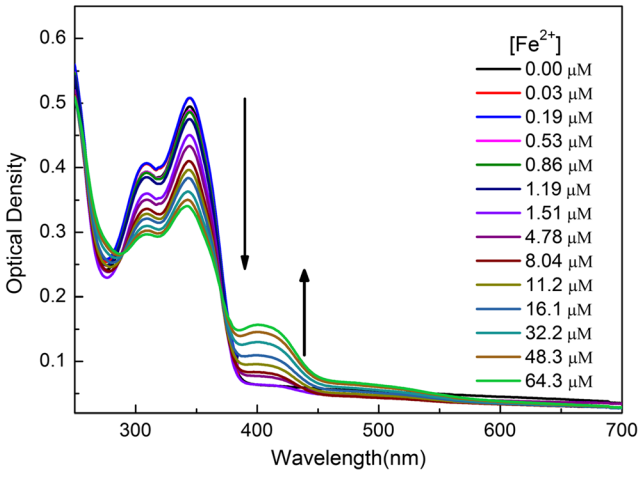


Fig. 5 UV-visible absorption spectra of 3HF on the successive addition of Fe^{2+} ions

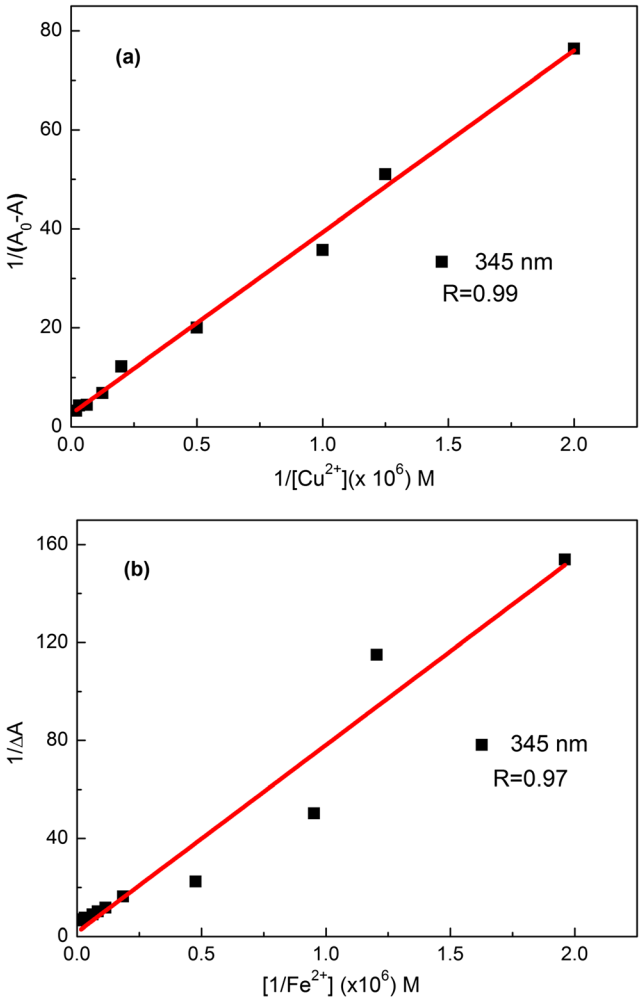


Fig. 6 B-H curves of 3HF in the presence of (**a**) Cu^{2+} and (**b**) Fe^{2+} ions

drinking water are 20 μM and 5.4 μM, respectively. Furthermore, WHO reports suggest that the average concentration of Cu^{2+} in blood should not exceed 15.7 − 23.6 μM [2, 19].

Quenching of 3HF fluorescence in the presence of metal ions was further investigated by Stern-Volmer (S-V) equation.

$$\frac{I_0}{I} = 1 + K_{SV}[Q] \tag{2}$$

where I_0 and I are the FL intensities of probe molecule in absence and presence of the quencher, respectively, Q is the concentration of the quencher and K_{SV} is the Stern -Volmer constant, which may be interpreted in different ways depending upon the process by which quenching is illustrated and estimated from the slope of a linear graph displaying dependence of intensity on [Q], as expressed by Eq. (2) [6, 7, 61–63]. The S-V plot for Cu^{2+} ions is linear at lower as well as higher concentrations of Cu^{2+} with different slopes, whereas a deviation from the linearity is noticed for intermediated Cu^{2+} concentration (Fig. S7 of SI). This could be either due to the differential accessibility of fluorophore to Cu^{2+} or the manifestation of a reverse reaction in the

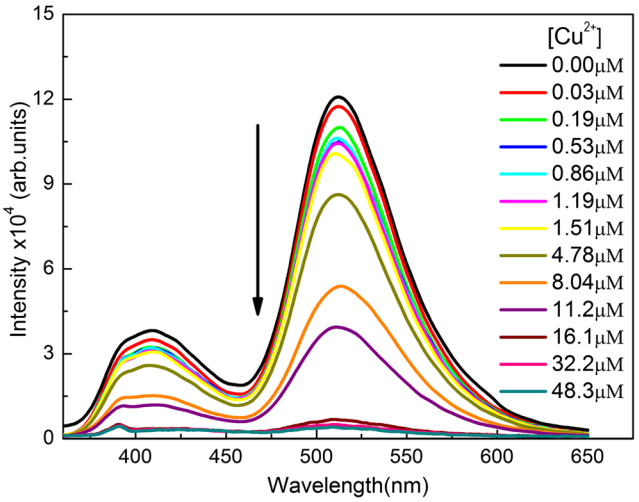


Fig. 7 Fluorescence spectra of 3HF on gradually increasing the concentration of Cu^{2+} ions. The excitation wavelength (λ_{ex}) was 345 nm

photochemical process. Since no new emission band was developed with the addition of Cu^{2+} into the aqueous solution of 3HF, the deviation from linearity is most likely due to the appearance of 3HF in two different ground state conformers (differentially accessible to the metal ions). Figure 9(a,b) represents the liner S-V plot for lower concentrations of Cu^{2+} ions at 409 nm and 513 nm. Furthermore, modified Stern-Volmer relation known as Lehrer equation given below transformed the nonlinear S-V plot into a linear form:

$$\frac{I_0}{I_0 - I} = \frac{1}{f} + \frac{1}{fK_{SV}[Q]} \tag{3}$$

where I_0 and I are the intensities of the solute molecules in the absence and in the presence of quencher, [Q] is the concentration of the quencher, *f* is the fraction of the accessible fluorophores, K_{sv} is the S-V constant [64, 65].

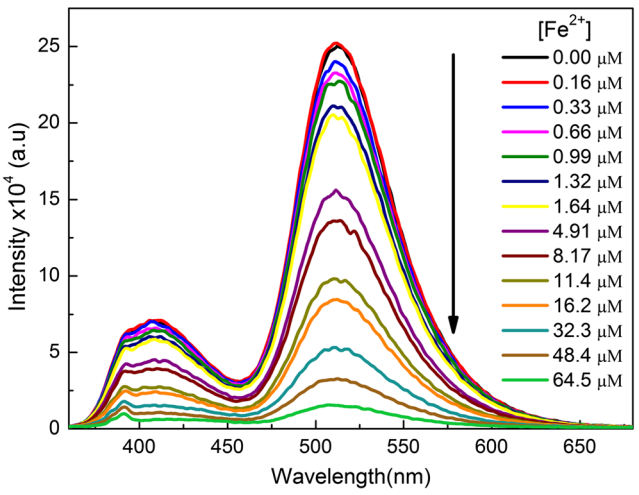


Fig. 8 Fluorescence spectra of 3HF on the successive addition of Fe^{2+} ions. Excitation wavelength (λ_{ex}) was 345 nm

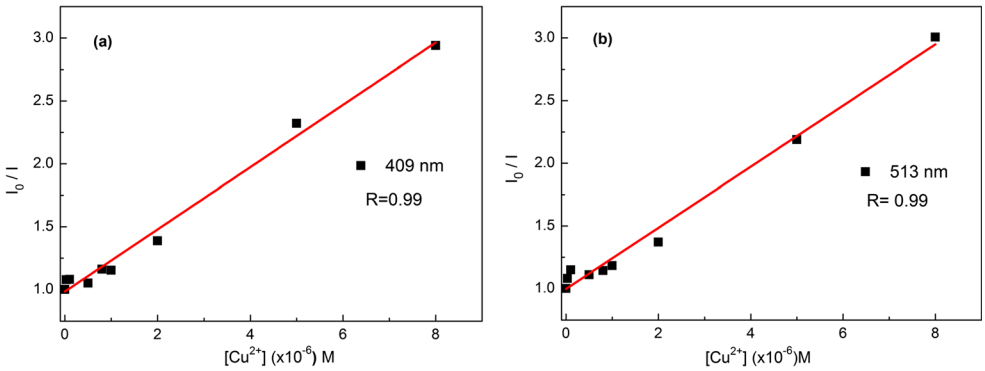


Fig. 9 S-V plots for 3HF for lower concentrations (0–8 μM) of Cu^{2+} ions at (**a**) λ_{em} = 409 nm and (**b**) λ_{em} = 513 nm

The plot of $I_0/(I_0 - I)$ versus 1/[Q] is linear as shown in (Fig.10 (a,b)) with 1/*f* as intercept and K_{sv} = intercept/slope. The value of *f* is nearly equal to 1 and S-V constants are 1.42 × 10^5 M^{-1} (at 409 nm) and 1.32 × $10^5 M^{-1}$ (at 513 nm). The K_{sv} values obtained from the slope of the linear plot of the S-V plot is ~2 fold higher than K_{sv} calculated from Lehrer equation, indicating that the probable quenching mechanism of 3HF in the presence of Cu^{2+} ions is initiated by ground-state complex formation. Initially, the S-V plot is linear with the smaller concentration of metal ions; however, it deviated slightly in the intermediated regions but followed well the Lehrer equation. Thus, static quenching occurs due to the presence of Cu^{2+} ions in the vicinity of fluorophores.

As depicted in (Fig. 11), the S-V plots obtained by using the experimentally determined values of I_0 and I is linear for Fe^{2+}. The K_{sv} from the steady-state experiment was calculated to be 1.41 × 10^5 M^{-1} (409 nm) and 1.31 × 10^5 M^{-1} (513 nm). The changes in the absorption spectrum, development of new absorption band in the longer wavelength regions and the linear S-V plots suggest that the Fe^{2+} ions quenched FL of 3HF statically following the ground state complexation between the ligand and metal ions.

There is a complexation between transition metals and the functional groups of 3HF, which is responsible for declining the fluorescence intensity. The newly formed charge transfer complex is considered to be non-emitting, as we could not observe any additional FL bands corresponds to the complex and the intensity of FL bands does not increase upon excitation with a red edge of the absorption band. The PL-excitation spectra as a function of the metal ions concentration (MI) follow the same trend of decrease of intensity. However, there is no new band in the excitation spectrum too, which hinders the emission from the complex. The mechanism behind the 3HF-MI complexes may be the formation of a strong π-electron bond between the fluorophore/ functional group of 3HF and the d-orbitals of the transition metal. The π-electron bonding system can be formed in two ways first, by donation of electrons from the functional group of 3HF to the partially filled-orbital of the transition metal (MI) or from the partially filled orbitals of metal to the functional group of 3HF. This type of π-electron bond in 3HF-MI complexes reduces the electron density of the fluorophore. It causes the d-orbital's to be either stabilized (of lower energy) or destabilized (of higher energy). A stabilizing effect from a fluorophore lowers the energy of the interacting d-orbital's, which can considerably decrease the electron transfer probability and decrease the fluorescence intensity consequently. It is generally known that the transition metals are excellent Lewis acids and accept electron density from many molecules or ions that act as Lewis bases. As an alternative, upon 3HF-MI formation, the probability of an electron transfer may be enhanced if a destabilizing effect from a fluorophore raises the energy of the d-orbitals. This effect subsequently may lead to an increase in the fluorescence intensity of the fluorophore in the

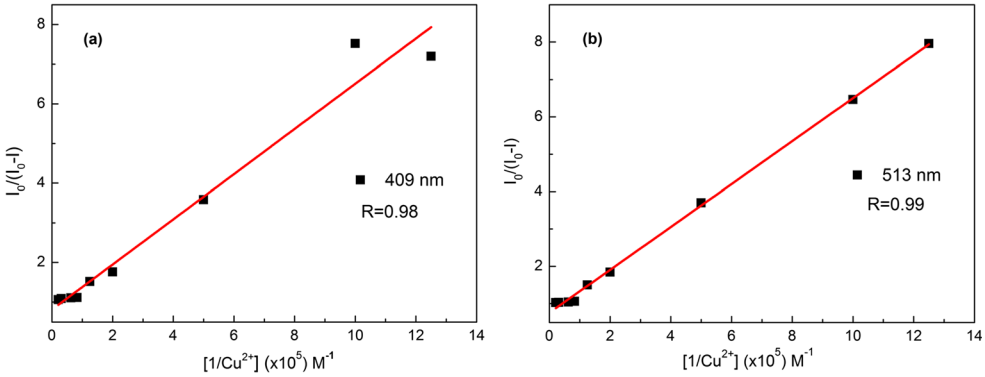


Fig. 10 Plots of $I_0/(I_0-I)$ vs. 1/[Q] for Cu^{2+} ions at (**a**) λ_{em} = 409 nm and (**b**) λ_{em} = 513 nm

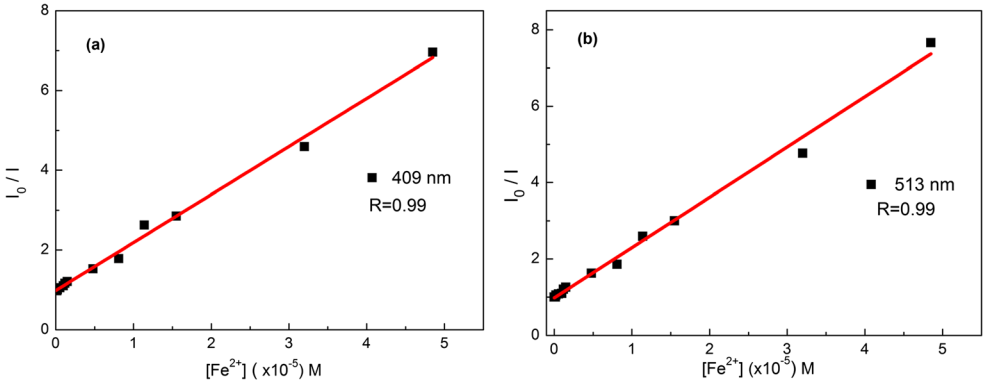
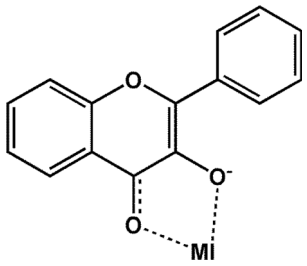


Fig. 11 S-V plots for 3HF in the presence of Fe^{2+} ions at (**a**) λ_{em} = 409 nm and (**b**) λ_{em} = 513 nm

3HF-MI complex [66, 67]. The rich redox activity of 3d metals probably explains that the electron transfer can be carried out directly from the fluorophore to the complexed ion (or from complexed ion to the fluorophore). Also, 3d ions possess empty levels of appropriate energy that can involve in energy transfer processes from the photoexcited fluorophore [68, 69]. The probe coordinates with metal ions via one oxygen atom of the carbonyl group and one oxygen atom from the hydroxyl group (Scheme 1). The FL quenching of 3HF caused by divalent copper and ferrous ions may be attributed to the photoinduced electron transfer (PET). It is assumed that electron transfer undergoes between 3HF and empty d-orbital's of Cu^{2+} and Fe^{2+}, leading to the rapid and efficient deactivation of 3HF through bimolecular non-radiative processes [2, 32, 35, 43].

In the present study, the other cations did not produce such effects under identical conditions, producing only minor changes in fluorescence intensity; this may be due to weak complexation. On the other hand, the fluorescent changes of the 3HF by closed-shell heavy metal ions such as Zn^{2+} and Pb^{2+} does not introduce low-energy metal-centered or charge-separated excited energy states due to which electron transfer processes is not feasible [13]. However, these fluorescence changes are negligible compare to those of paramagnetic metal ions such as Fe^{2+} and Cu^{2+}. Two possibilities are considered for the decreased fluorescence intensities of 3HF in the presence of metal ions. One is the formation of metal complexes with 3HF and the other is electron/charge transfer between



Scheme 1 Complexation between 3HF and metal ions (MI).

3HF and metal ions. Based on reported values of the oxidation potential of 3HF, Fe^{2+} and Cu^{2+} (0.80, 0.161 and − 0.44 V) [70] and applying Rehm Weller equation [71], free energy change of 3HF with Fe^{2+} and Cu^{2+} is calculated to be −1.635 and − 2.161 V, respectively. The more negative values of free energy change suggest electron/charge transfer of 3HF in the presence of Fe^{2+} and Cu^{2+}. It is noteworthy that large oxidation of 3HF was considered in the presence of Cu^{2+} ions. As the standard potential of the Cu^{2+} was much more positive than that of other metals (including K^+, Na^+, Zn^{2+}, Ca^{2+}, Mn^{2+}, Mg^{2+}, Pb^{2+}, and Al^{3+}), no apparent fluorescence quenching could be observed with the addition of other metal ions.

Anti-Interference Study

Competitive analysis was carried out to confirm the distinction of 3HF towards Cu^{2+} ions. As shown in Fig. 12, only 20 μM of Cu^{2+} can quickly quench the fluorescence of 3HF solution in the presence of 50 μM concentration of other relevant metal ions. However, even the presence of 50 μM of other metal ions does not impede the colorimetric recognition of probe to Cu^{2+}, except for Fe^{2+}.

To examine the interference effects of Fe^{2+}, another competitive experiment was performed. It is significant that 50 μM of $Cu^{2+}$and Fe^{2+} quenches the FL intensity of 3HF. Further, the addition of 20 μM Fe^{2+} in 3HF-Cu^{2+} complex reduced the FL intensity of complex by 1.76 fold. Conversely, on adding 20 μM of Cu^{2+}, FL intensity of 3HF-Fe^{2+} decreases upto 23.9 fold. This demonstrates that Cu^{2+} is an effective quencher, even in the presence of Fe^{2+} ions (Fig. 13) for the used probe. The parameters calculated from UV-Vis absorption and FL spectra of 3HF-Cu^{2+} and 3HF-Fe^{2+} complexes are furnished in Table 1. A comparison of obtained data reveals that 3HF shows a lower detection limit, strong binding affinity and high quenching efficiency towards Cu^{2+} ions.

Selectivity is due to the high affinity of Cu^{2+} for oxygen donors and the strong tendency to promote the deprotonation

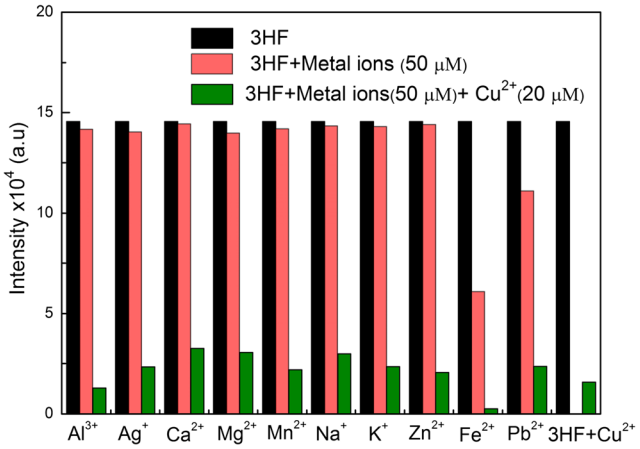


Fig. 12 Anti-interference study of 3HF (20 μM) in methanol-water (1:2) towards Cu^{2+}ions (20 μM) in the presence of other metal ions (50 μM) (λ_{ex} = 345 nm)

of hydroxyl oxygen atoms during complex formation. Metal-ion selectivity, binding kinetics, and proton cross-sensitivity are strongly dependent on the structure of the chelating group of the fluorescent sensor. According to the rule of Irving–Williams, among the relevant paramagnetic metal ions, Cu^{2+} has an unusually high thermodynamic affinity for typical N, O-chelate ligands and fast metal-to-ligand binding kinetics [72]. Cu^{2+} with 3HF demonstrates high binding constant and also it has multiple chelation with 3HF as reported by Monocomble et al. [60]. Although the other metal ions also have the ability to form complexes, it is not as strong as that of Cu^{2+} ions. However, equimolar concentrations of metals (50 μM) do not affect the response of Cu^{2+} (20 μM), the presence of Cu^{2+} quenches the fluorescence, possibly because there is a competitive binding between interfering metals and Cu^{2+} with 3HF. This result also indicates the complex formation between other metal ions and 3HF is feeble that can be

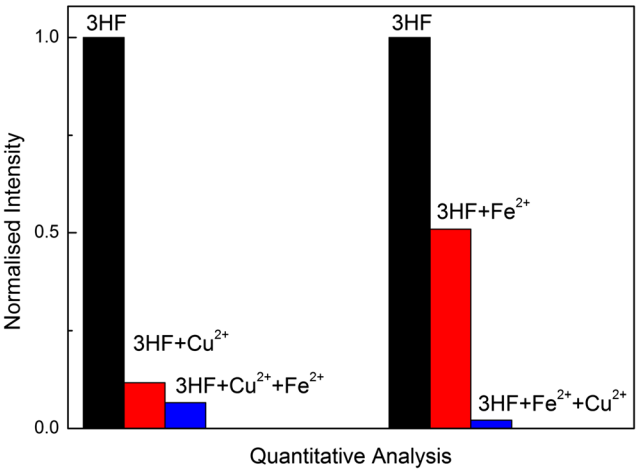


Fig. 13 Normalized fluorescence intensity of probe 3HF in the presence of $Cu^{2+}$and Fe^{2+} ions and vice-versa for competitive analysis

replaced by Cu^{2+}. Hence, the probe 3HF is more sensitive in detecting Cu^{2+} ions in the medium.

Restoring Fluorescence of the Probe

Since the reversible nature of chemosensors is equally essential for its practical applications. To assess its reversibility with metal ions, 0.2 M of HCl were added to the 3HF solutions comprising Cu^{2+} and Fe^{2+} ions. The addition of HCl could immediately fully re-established the FL of 3HF for Cu^{2+} ions (~50 fold from the diminished one) while weakly restored for Fe^{2+} ions (~19 fold from the diminished one) even at an excessive concentration of HCl (Fig. 14(a)). For 3HF + Fe^{2+}, the concentration 0.2 M HCl maybe not sufficient for the substitution of 3HF + Fe^{2+} with HCl or that the substitution reaction was not efficient. As adding 0.2 M HCl in the 3HF + Fe^{2+} immediate increase in the FL intensity to ~19 fold, whereas gradually after that up to ~30 fold. This reflects the weak restoration of 3HF in the presence of Fe^{2+}. To examine the FL recovery process of 3HF, Cu^{2+} ions (50 μM) were added in the 3HF solution to acquire an effectively quenched 3HF-Cu^{2+} fluorescence. Further, upon addition of 0.2 M of HCl, 91% of diminished FL was recovered from 3HF-Cu^{2+} complex, as shown in Fig. 14 (b). Fig. 14(c) represents the response of 3HF+Cu^{2+} solutions in the absence and presence of HCl under UV- light. UV-Visible absorption spectra of a 3HF-Cu^{2+} complex in the presence of increasing concentrations of HCl (Fig. 15) exhibits the formation of free 3HF, as the absorption band at 345 nm redevelops and the band at longer wavelength region disappears due to protonation (H^+). Steadily addition of HCl (Fig. 16), the quenched FL signal of the 3HF-Cu^{2+} complex is significantly enhanced at the higher level indicating the regeneration of free chemosensors. The presence of HCl in a 3HF-Cu^{2+} solution may cause Cu^{2+} to be removed from the 3HF surface; as a result, the quenched fluorescence being restored by protonation of 3HF. The FL increment was discontinued upon the addition of 0.2 M of HCl, indicating complete removal of metal ions from binding sites of sensing probe.

Practical Applications

3HF showed a dual response detection of Cu^{2+} and Fe^{2+} ions; it is not convenient enough to use the probe in the form of a solution for onsite analysis. Test strips like pH strips have been prepared to demonstrate better portability and operational simplicity. The test strips were prepared by immersing filter papers into 2 mM ethyl alcohol solution of 3HF for 10 min and then dried in air. The solution of Cu^{2+} and Fe^{2+}ions was prepared in

Table 1 Comparison of parameters calculated for $Cu^{2+}$and Fe^{2+} ions

Spectral Parameters		3HF+ Cu^{2+}	3HF + Fe^{2+}
Detection Limit		1.54 μM	1.98 μM
Association Constant		7.14×10^4 M^{-1}	2.18×10^4 M^{-1}
S-V constant (Static Quenching)	$\lambda_{em} = 409$ nm	2.47×10^5 M^{-1} *	1.41×10^5 M^{-1}
	$\lambda_{em} = 513$ nm	2.43×10^5 M^{-1}*	1.31×10^5 M^{-1}
I_0/I (at λ_{513} for 50 μM)		73.84	35.66
Quenching efficiency (50 μM)		88.35%	49.07%

*for lower concentrations of Cu^{2+}

water with different concentrations of 0.0 M, 10 μM, 0.1 mM, 0.5 mM and 1.0 mM. To check the response of the strips, they were immersed in the various concentrations of Cu^{2+} and Fe^{2+} ions, each for only 2–3 s. Promptly, a gradual colour change has been observed by the naked eye from white to yellow for Cu^{2+} (Fig.17 (a)) and white to brown for Fe^{2+}(Fig.17 (b)).

The test strips were kept under a UV light; the colour of test strips changed from light blue to dark blue (Fig. 18 (a,b)). The visual detection limit of this method was approximately 10 μM. The test strips of 3HF were therefore found to be useful, simple, quick and handy to detect Cu^{2+} and Fe^{2+} ions in colorimetric and fluorescent mode.

ON-OFF Switching Behaviour of 3HF and its Logical Implementation

As the quencher displacement strategy causes consequent changes in emission intensity at 513 nm; as a result, the system worked as a molecular switch at this fluorescence intensity, which performed at Boolean logic operations. The molecular logic function is performed with 3HF along with Cu^{2+}

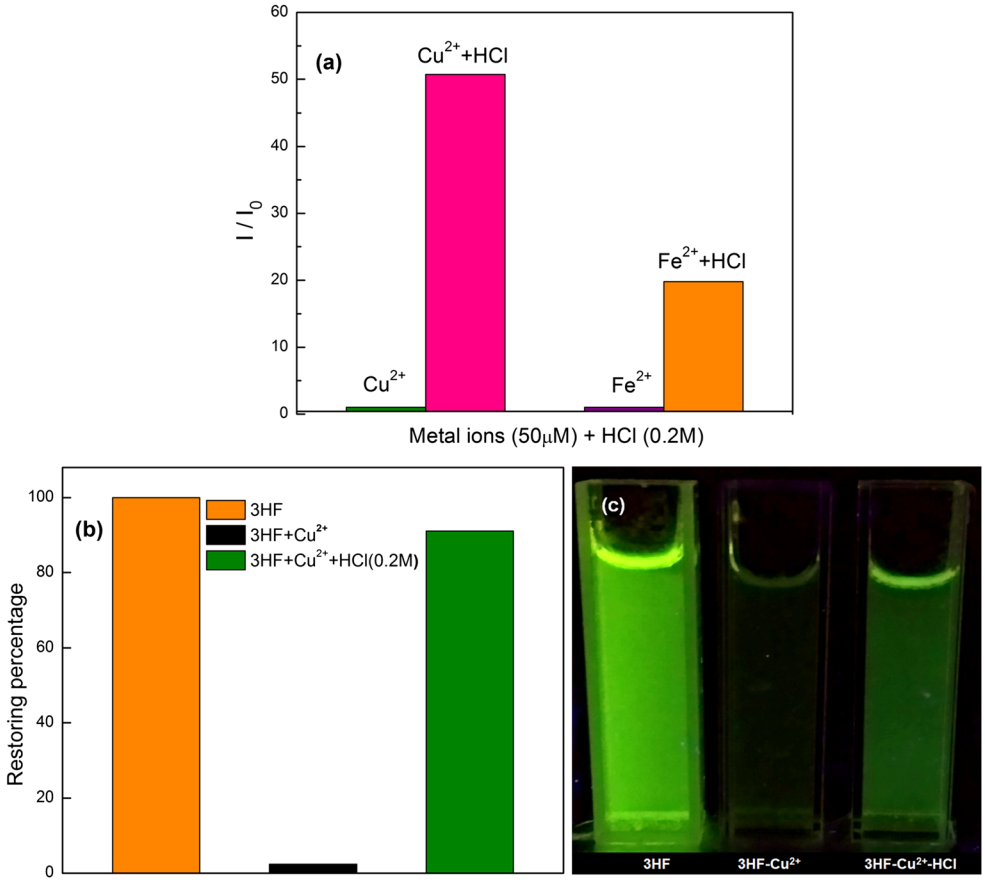


Fig. 14 Comparative response of 3HF-MI complexes in the presence of (0.2 M) HCl (**a**) and restoring percentage of fluorescence intensity of 3HF-Cu^{2+} on the addition of 0.2 M HCl (**b**). Response (image) of 3HF+Cu^{2+} solutions in the absence and presence of HCl under UV- light (**c**)

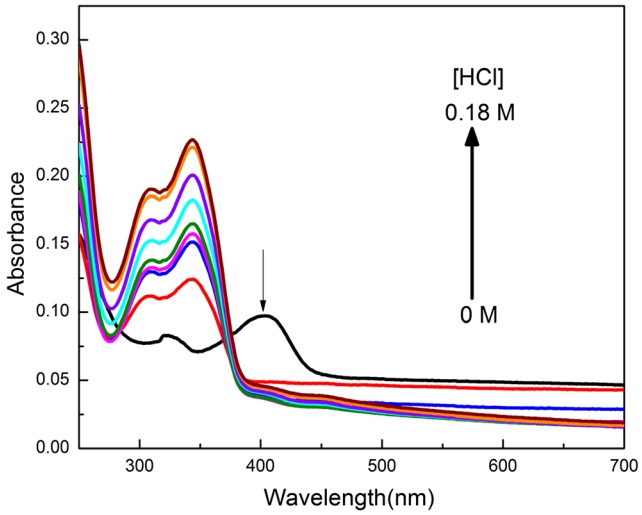


Fig. 15 Absorption spectra of 3HF-Cu^{2+} in the presence of HCl

(In-1) and H^+ (In-2) as inputs in the emission mode. In the existing system, the strong fluorescence at 513 nm is assigned as ON state (output = 1), while the weak fluorescence defined as OFF state (output = 0). The threshold value of fluorescence intensity is specified at 10,000 CPS. The addition of Cu^{2+} in the 3HF solution leads fluorescence quenching below the threshold label, whereas in situ addition of excess H^+ almost regained the fluorescence intensity. The strong fluorescence intensity above the threshold label was recognized in the absence (0, 0) and presence (1, 1) of both inputs and also H^+ alone. These results of the IMPLICATION logic gate implies that the OFF state or 0 output is obtained only when the input is Cu^{2+}. The working principle of IMPLICATION logic gate is outlined in Fig. 19.

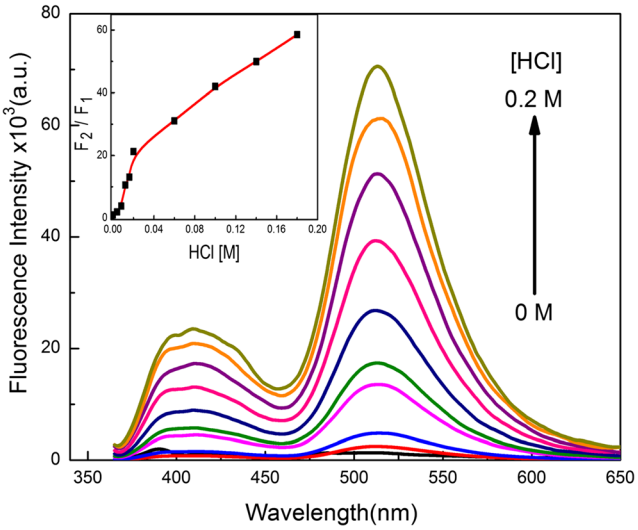


Fig. 16 Fluorescence spectra of 3HF-Cu^{2+} on the successive addition of HCl, λ_{ex} = 345 nm

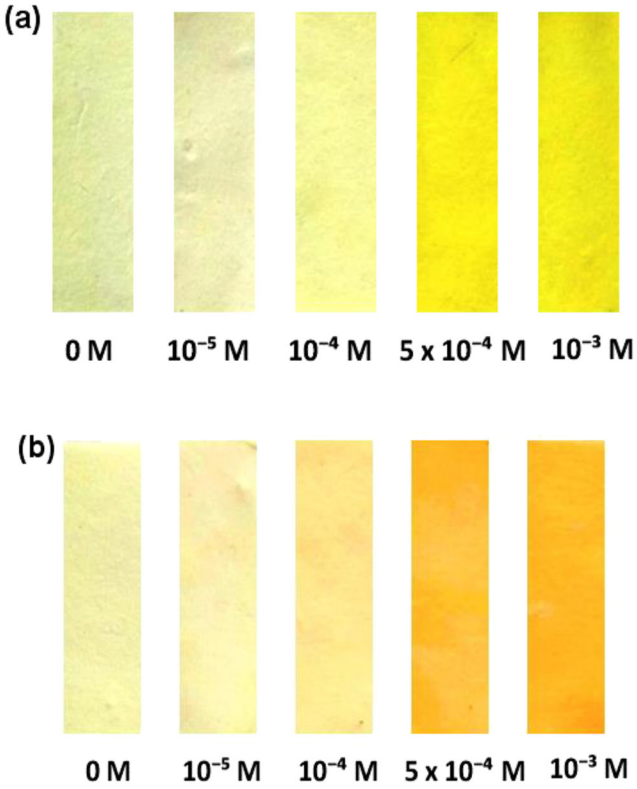


Fig. 17 Color change of test strip coated by 3HF (2 mM) with various concentrations of (**a**) Cu^{2+} and (**b**) Fe^{2+} ions (0 M, 10^{-5} M, 10^{-4} M, 5×10^{-4} M and 10^{-3} M) in visible light

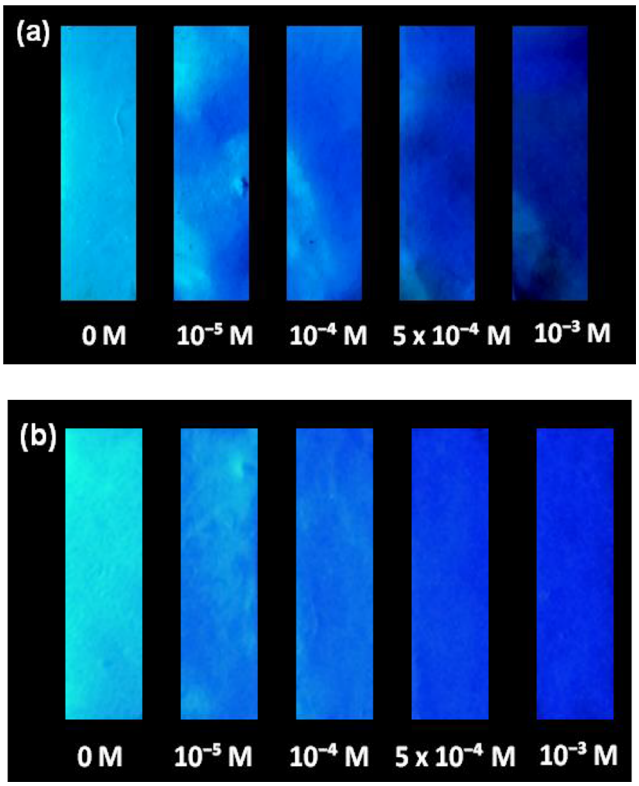


Fig. 18 Color change of test paper coated by 3HF (2 mM) with various concentration of (**a**) $Cu^{2+}$and (**b**) Fe^{2+} (0 M, 10^{-5} M, 10^{-4} M, 5×10^{-4} M and 10^{-3} M) under UV light

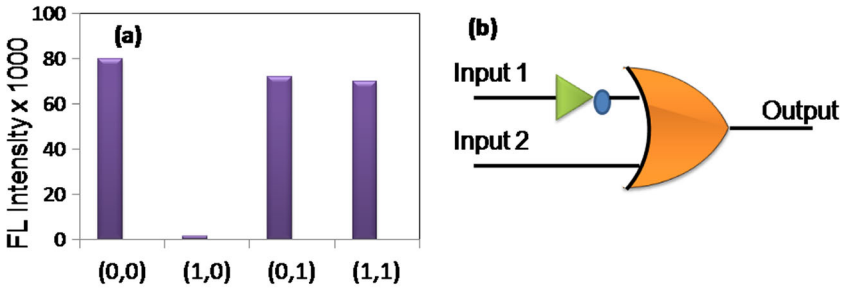


Fig. 19 Arrangement of the IMPLICATION logic gate: (**a**) the changes in fluorescence intensity under different input conditions, (**b**) the switch circuit diagram and (**c**) the truth table corresponding to the logic gate

(c)

Input 1 (Cu^{2+})	Input 2 (H^{+})	Output
0	0	1
1	0	0
0	1	1
1	1	1

Conclusions

As a result, chemosensor based on hydroxyflavone has been studied for colorimetric and turn-off fluorescence detection of Cu^{2+} and Fe^{2+} ions. According to the B-H plots, 1:1 stoichiometry was formed between the probe and Cu^{2+} or Fe^{2+} ions. The Lehrer and Stern-Volmer plots suggest the possibility that the nature of FL quenching of 3HF in the presence of Cu^{2+} (~74 fold) and Fe^{2+} (~36 fold) ions is static. In the presence of common interfering ions, 3HF represented a highly selective and efficient sensing of Cu^{2+} and Fe^{2+} ions in aqueous solutions. Significantly, calculated detection limits of 3HF for Cu^{2+} (1.54 μM) and Fe^{2+} (1.98 μM) ions are much below than the guidelines of EPA (Cu^{2+}~20 μM and Fe^{2+}~5.4 μM).

Moreover, the competition experiment and anti-interference study recommend that 3HF is more proficient and sensitive towards Cu^{2+} ions. Besides, efficient but straightforward test strips could detect Cu^{2+} and Fe^{2+} ions conveniently and rapidly in aqueous media. We consider that inherited ESIPT process of 3HF and its colorimetric, and turn off response towards Cu^{2+} and Fe^{2+} ions are undoubtedly promising features in developing novel chemosensors for the detection of multiple metal ions such as Cu^{2+} and Fe^{2+} in biological and environmental systems. In addition, the response of 3HF towards Cu^{2+} ions was found to be reversible, which further made the present system attractive for sensor applications.

Acknowledgements Authors are thankful to all the members of Photophysics Laboratory, Department of Physics, Centre of Advance Study, DSB Campus, Kumaun University, Nainital and Laser Spectroscopy Laboratory, Department of Applied Physics, Delhi Technological University, Delhi for their support. Authors thank Science and Engineering Research Board (SERB)(EMR/2016/001110) Govt. of India for supporting the research by awarding the project.

References

1. Anbu S, Ravishankaran R, Guedes da Silva MFC, Karande AA, Pombeiro AJL (2014) Differentially selective chemosensor with fluorescence off–on responses on Cu^{2+} and Zn^{2+} ions in aqueous media and applications in pyrophosphate sensing, live cell imaging, and cytotoxicity. Inorg Chem 53:6655–6664. https://doi.org/10.1021/ic500313m
2. Zhanga B, Liua H, Wu F, Haoc GF, Chend Y, Tana C, Tana Y, Jiang Y (2017) A dual-response quinoline-based fluorescent sensor for the detection of copper (II) and Iron(III) ions in aqueous medium. Sens Actuators B Chem 243:765–774. https://doi.org/10.1016/j.snb.2016.12.067
3. Helal A, Kim HS (2009) Thiazole-based chemosensor: synthesis and ratiometric fluorescence sensing of zinc. Tetrahedron Lett 50:5510–5515. https://doi.org/10.1016/j.tetlet.2009.07.078
4. Bhorge YR, Tsai HT, Huang KF, Pape AJ, Janaki SN, Yen YP (2014) A new pyrene-based Schiff-base: a selective colorimetric and fluorescent chemosensor for detection of cu(II) and Fe(III). Spectrochim Acta - Part A Mol Biomol Spectrosc 13:7–12. https://doi.org/10.1016/j.saa.2014.03.110
5. Wu J, Liu W, Ge J, Zhang H, Wang P (2011) New sensing mechanisms for design of fluorescent chemosensors emerging in recent years. Chem Soc Rev 40:3483–3495. https://doi.org/10.1039/C0CS00224K
6. Demchenko A P (2015) Introduction to fluorescence sensing, Second Edition, Springer. https://doi.org/10.1007/978-3-319-20780-3
7. Lackowicz J R (2002) Topics in fluorescence spectroscopy, Probe Design and Chemical, Springer Volume 4. https://doi.org/10.1007/B112911
8. Swamy KMK, Ko SK, Kwon SK, Lee HN, Mao C, Kim JM, Lee KH, Kim J, Shin I, Yoon J (2008) Boronic acid-linked fluorescent and colorimetric probes for copper ions. Chem Commun 45:5915–5917. https://doi.org/10.1039/b814167c
9. Pandey N, Mehata MS, Fatma N, Pant S (2019) Efficient fluorescence quenching of 5-aminoquinoline: silver ion recognition study. J Lumin 205:475–481. https://doi.org/10.1016/j.jlumin.2018.09.062
10. Juan W, Guan QJ, Wei TB, Yan GT, Lin Q, Zhang YM (2016) A turn–on fluorescent sensor for relay recognition of two ions: from

F^- selective sensor to highly Zn^{2+}–selective sensor by tuning electronic effects. RSC Adv 6:35804–35808. https://doi.org/10.1039/C6RA05381E

11. Chen X, Jia J, Ma H, Wang S, Wang X (2009) Characterization of Rhodamine B Hydroxylamide as a highly selective and sensitive fluorescence probe for copper(II). Anal Chim Acta 632:9–14. https://doi.org/10.1016/j.aca.2007.08.025
12. Chen X, Nam S-W, Jou MJ, Kim Y, Kim S-J, Park S, Yoon J (2008) Hg^{2+} selective fluorescent and colorimetric sensor: its crystal structure and application to bioimaging. Org Lett 10:5235–5238. https://doi.org/10.1021/ol8022598
13. Shanmugaraj K, I lanchelian M (2016) A "turn-off" fluorescent sensor for the selective and sensitive detection of copper(ii) ions using lysozyme stabilized gold nanoclusters. RSC Adv 6:54518–54524. https://doi.org/10.1039/c6ra08325k
14. Tiwari K, Kumar S, Kumar V, Kaur J, Arora S, Mahajan RK (2018) An azine based sensor for selective detection of Cu^{2+} ions and its copper complex for sensing of phosphate ions in physiological conditions and in living cells. Spectrochim Acta Part A Mol Biomol Spectrosc 191:16–26. https://doi.org/10.1016/j.saa.2017.09.072
15. Espada-Bellido E, Galindo-Riaño MD, García-Vargas M, Narayanaswamy R (2010) Selective chemosensor for copper ions based on fluorescence quenching of a schiff-base fluorophore Appl. Spectrosc. 64:727–732. https://doi.org/10.1366/000370210791666282
16. Bhardwaj S, Maurya N, Singh AK, Varshney R, Roy P (2016) Promising ESIPT-based fluorescence sensor for Cu^{2+} and CN^- ions: investigation towards logic gate behaviour, anticancer activities and bioimaging application. RSC Adv 6:102096–102101. https://doi.org/10.1039/c6ra22352d
17. Chen D, Chen P, Zong L, Sun Y, Liu G, Yu X, Qin J (2017) Colorimetric and fluorescent probes for real-time naked eye sensing of copper ion in solution and on paper substrate. R Soc Open Sci 4:161–171. https://doi.org/10.1098/rsos.171161
18. Wang L, Ye D, Li W, Liu Y, Li L, Zhang W, Ni L (2017) Fluorescent and colorimetric detection of Fe(III) and cu(II) by a difunctional rhodamine-based probe. Spectrochim Acta Part A Mol Biomol Spectrosc 183:291–297. https://doi.org/10.1016/j.saa.2017.04.056
19. Narayanaswamy N, Govindaraju T (2012) Aldazine-based colorimetric sensors for Cu^{2+} and Fe^{3+}. Sens Actuators B Chem 161:304–310. https://doi.org/10.1016/j.snb.2011.10.036
20. Park GJ, You GR, Choi YW, Kim C (2016) A naked-eye chemosensor for simultaneous detection of iron and copper ions and its copper complex for colorimetric/fluorescent sensing of cyanide. Sens Actuators B Chem 229:257–271. https://doi.org/10.1016/j.snb.2016.01.133
21. Beltran OG, Cassels BK, Perez C, Mena N, Nunez MT, Martinez NP, Pavez P, Aliaga ME (2014) Coumarin-based fluorescent probes for dual recognition of copper (ii) and iron (iii) ions and their application in bio-imaging. Sensors 14:1358–1371. https://doi.org/10.3390/s140101358
22. Kumawat LK, Mergu N, Singh AK, Gupta VK (2015) A novel optical sensor for copper ions based on phthalocyaninetetrasulfonic acid. Sens Actuators B Chem 212:389–394. https://doi.org/10.1016/j.snb.2015.02.027
23. Warrier S, Kharkar PS (2018) A coumarin based chemosensor for selective determination of cu (II) ions based on fluorescence quenching. J Lumin 199:407–415. https://doi.org/10.1016/j.jlumin.2018.03.073
24. Nawaz H, Tian W, Zhang J, Jia R, Chen Z, Zhang J (2018) Cellulose-based sensor containing phenanthroline for the highly selective and rapid detection of Fe^{2+} ions with naked eye and fluorescent dual modes. ACS Appl Mater Interfaces 10:2114–2121. https://doi.org/10.1021/acsami.7b17342
25. Hirayama T, Nagasawa H (2017) Chemical tools for detecting Fe ions. J Clin Biochem Nutr 60:39–48. https://doi.org/10.3164/jcbn.16_70
26. Basoglu A, Tosun G, Ocak M, Alp H, Yayl N, Ocak U (2015) Simple time-saving method for iron determination based on fluorescence quenching of an azaflavanon-3-ol compound. J Agric food Chem 63:2654–2659. https://doi.org/10.1021/jf505336d
27. Lee JA, Eom GH, Park HM, Lee JH, Song H, Hong CS, Yoon S, Kim C (2012) Selective Fe^{2+} ion recognition using a fluorescent pyridinyl-benzoimidazole-derived ionophore. Bull Kor Chem Soc 33:3625–3628. https://doi.org/10.5012/bkcs.2012.33.11.3625
28. Kumawat LK, Asif M, Gupta VK (2017) Dual ion selective fluorescence sensor with potential applications in sample monitoring and membrane sensing. Sens Actuators B Chem 241:1090–1098. https://doi.org/10.1016/j.snb.2016.10.031
29. Choi YW, Park GJ, Na YJ, Jo HY, Lee SA, You GR, Kim C (2014) A single schiff base molecule for recognizing multiple metal ions: a fluorescence sensor for Zn(II) and Al(III) and colorimetric sensor for Fe(II) and Fe(III). Sens Actuators B Chem 194:343–352. https://doi.org/10.1016/j.snb.2013.12.114
30. Zhang S, Wu X, Niu Q, Guo Z, Li T, Liu H (2016) Highly selective and sensitive colorimetric and fluorescent chemosensor for rapid detection of Ag^+, Cu^{2+} and Hg^{2+}based on a simple Schiff base. J Fluoresc 27:729–737. https://doi.org/10.1007/s10895-016-2005-y
31. Nguyen DM, Frazer A, Rodriguez L, Belfield KD (2010) Selective fluorescence sensing of zinc and mercury ions with hydrophilic 1,2,3-triazolyl fluorene probes. Chem Mater 22:3472–3481. https://doi.org/10.1021/cm100599g
32. Lee HG, Kim KB, Park GJ, Na YJ, Jo HY, Lee SA, Kim C (2014) An anthracene-based fluorescent sensor for sequential detection of zinc and copper ions. Inorg Chem Comm 39:61–65. https://doi.org/10.1016/j.inoche.2013.10.049
33. Qin J-C, Yang Z-Y (2015) Design of a novel coumarin-based multifunctional fluorescent probe for Zn^{2+}/Cu^{2+}/S^{2-} in aqueous solution. Mater Sci Eng C 57:265–271. https://doi.org/10.1016/j.msec.2015.07.064
34. Sisa M, Bonnet SL, Ferreira D, Van der Westhuizen JH (2010) Photochemistry of flavonoids. Molecules 15:5196–5245. https://doi.org/10.3390/molecules15085196
35. Jovanovic SV, Steeden S, Tosic M, Marjanovic B, Simicg MG (1994) Flavonoids as antioxidants. J Am Chem Soc 116:4846–4851. https://doi.org/10.1021/ja00090a032
36. Samsonowicz M, Regulska E, Kalinowska M (2017) Hydroxyflavone metal complexes – molecular structure, antioxidant activity and biological effects. Chem Biol Interact 273:245–256. https://doi.org/10.1016/j.cbi.2017.06.016
37. de Rijke E, Out P, Niessen WMA, Ariese F, Gooijer F, Brinkman UAT (2006) Analytical separation and detection methods for flavonoids. J Chromatogr A 1112:31–63. https://doi.org/10.1016/j.chroma.2006.01.019
38. Voicescu M, Ionescu S, Gatea F (2014) Photophysical properties of some flavones probes in homogeneous media. J Fluoresc 24:75–83. https://doi.org/10.1007/s10895-013-1272-0
39. Wolfbeis OS, Leiner M, Hochmuth P (1984) Absorption and fluorescence spectra, pK_a values and fluorescence lifetimes of monohydroxyflavones and monomethoxyflavones. BerBunsenges Phys Chem 88:759–767. https://doi.org/10.1002/bbpc.19840880817
40. Kumar S, Jain S K, Rastogi R C (2001) An experimental and theoretical study of excited-state dipole moments of some flavones using an efficient solvatochromic method based on the solvent polarity parameter E^T_N. Spectrochim Acta Part A Mol BiomolSpectrosc 57:291–298. https://doi.org/10.1016/S1386-1425(00)00372-3
41. Zhu A, Wang B, White JO, Drickamer HG (2004) The effect of pressure on the excited-state intramolecular proton transfer of

polyhydroxyflavone. J Phys Chem B 108:895–898. https://doi.org/10.1021/jp036312g
42. Mandal PK, Samanta A (2003) Evidence of ground-state proton-transfer reaction of 3-hydroxyflavone in neutral alcoholic solvents. J Phys Chem A 107:6334–6339. https://doi.org/10.1021/jp027613d
43. Sengupta P K, Kasha M (1979) Excited state proton-transfer spectroscopy of 3-hydroxyflavone and quercetin. Chem Phys Lett 68: 382–385. https://doi.org/10.1016/0009-2614(79)87221-8
44. Dereka B, Letrun R, Svechkarev D, Rosspeintner A, Vauthey E (2014) Excited-state dynamics of 3-hydroxyflavone anion in alcohols. J Phys Chem B 119:2434–2443. https://doi.org/10.1021/jp507311n
45. Strandlord AJG, Courtney SH, Friedrich DM, Barbara PF (1983) Excited-state dynamics of 3-hydroxyflavone. J Phys Chem 87: 1125–1133. https://doi.org/10.1021/j100230a008
46. Beg SA, Ormson SM, Brown RG, Matousek P, Towrie M, Nibbering ETJ, Neuwahl FVR (2001) The ultrafast measurements of excited state intramolecular proton transfer (ESIPT) in room temperature solutions of 3-hydroxyflavone and derivatives. J Phys Chem A 105:3709–3718. https://doi.org/10.1021/jp0031101
47. McMorrow D, Kasha M (1984) Intramolecular excited-state proton transfer in 3-hydroxyflavone hydrogen-bonding solvent perturbations. J Phys Chem 88:2235–2243. https://doi.org/10.1021/j150655a012
48. Swinney TC, Kelley DF (1991) Proton-transfer and solvent polarization dynamics in 3-hydroxyflavone mixed solvents. J Phys Chem 95:10369–10373. https://doi.org/10.1021/j100178a024
49. Protti S, Mezzetti A (2015) Solvent effects on the photophysics and photoreactivity of 3-hydroxyflavone: a combined spectroscopic and kinetic study. J Mol Liq 205:110–114. https://doi.org/10.1016/j.molliq.2014.12.001
50. Voicescu M, Ionescu S (2014) Fluorescence characteristics of some flavones probes in different micellarmedia. J Fluoresc 24:735–743. https://doi.org/10.1007/s10895-013-1346-z
51. Mondal S, Basu S, Mandal D (2009) Ground- and excited-state proton-transfer reaction of 3-Hydroxyflavone in aqueous micelles. Chem Phys Lett 479:218–223. https://doi.org/10.1016/j.cplett.2009.08.026
52. Ghosh S, Chattopadhyay N (2015) Modification of the photophysics of 3-hydroxyflavone in aqueous solutions of imidazolium-based room temperature ionic liquids: a comparison between micelle-forming and non micelle-forming ionic liquids. RSC Adv 5:49054–49061. https://doi.org/10.1039/c5ra03833b
53. Protti S, Mezzetti A, Lapouge C, Cornard JP (2008) Photochemistry of metal complexes of 3-hydroxyflavone: towards a better understanding of the influence of solar light on the metal–soil organic matter interactions. Photochem Photobiol Sci 7:109–119. https://doi.org/10.1039/b709682h
54. Engelmann MD, Hutcheson R, Cheng IF (2005) Stability of ferric complexes with 3-Hydroxyflavone (Flavonol), 5,7-Dihydroxyflavone (Chrysin), and 3′,4′-Dihydroxyflavone. J Agric Food Chem 53:2953–2960. https://doi.org/10.1021/jf048298q
55. Mira L, Fernandez MT, Santos M, Rocha R, Florencio MH, Jennings KR (2002) Interactions of flavonoids with iron and copper ions: a mechanism for their antioxidant activity. Free Rad Res 36: 1199–1208. https://doi.org/10.1080/1071576021000016463
56. Thompson M, Williams CR, Elliot GEP (1976) Stability of flavonoid complexes of copper(ii) and flavonoid antioxidant activity. Anal Chim Acta 85:375–381
57. Al-Kindy SMZ, Al-Hinai KH, Suliman FO, Al-Lawati HJ, Pillay A (2011) Development of a selective fluorimetric technique for rapid trace determination of zinc using 3-hydroxyflavone. Arab J Chem 4:147–152. https://doi.org/10.1016/j.arabjc.2010.06.029
58. Gupta VK, Mergu N, Singh AK (2014) Fluorescent chemosensors for Zn^{2+} ions based on flavonol derivatives. Sens Actuators B Chem 202:674–682. https://doi.org/10.1016/j.snb.2014.05.133
59. Gao H, Wu X (2018) A 3-hydroxyflavone derivative as fluorescence chemosensor for copper and zinc ions. Chem Heterocycl Compd 54:125–129. https://doi.org/10.1007/s10593-018-2243-9
60. Moncomble A, Falantin C, Cornard J –P (2018) Electronic spectroscopies combined with quantum chemistry calculations study of the interactions of 3-hydroxyflavone with copper ions. J Phys Chem B https://doi.org/10.1021/acs.jpcb.8b06062, 122, 8943, 8951
61. Joshi NK, Tewari N, Arora P, Rautela R, Pant S, Joshi HC (2015) Photophysical behaviour and fluorescence quenching by halides of quinidine di cation : steady state and time resolved study. J Lumin 158:412–416. https://doi.org/10.1016/j.jlumin.2014.10.041
62. Mehata MS, Tripathi HB (2002) Fluorescence quenching of 6-methoxyquinoline: an indicator for sensing chloride ion in aqueous media. J Lumin 99:47–52. https://doi.org/10.1016/S0022-2313(02)00294-6
63. Mishra H, Mishra V, Mehata MS, Pant TC, Tripathi HB (2004) Fluorescence studies of salicylic acid doped poly(vinyl alcohol) film as a water/humidity sensor. J Phys Chem A108:2346–2352
64. Htun T (2004) A negative deviation from stern–Volmer equation in fluorescence quenching. J Fluoresc 14:217–222. https://doi.org/10.1023/B:3AJOFL.0000016294.96775.fd
65. Wang QX, Berglund KA (2016) Pyranine fluorescence quenching for the characterization of solutions. Am J Analyt Chem 7:43–56. https://doi.org/10.4236/ajac.2016.71005
66. Mostofa KMG, Liu CQ, Feng XB, Yoshioka T, Vione D, Pan XL, Wu FC (2013) Photobiogeochemistry of organic matter: principles and practices in water environments. Springer. https://doi.org/10.1007/978-3-642-32223-5
67. Mostofa KMG, Liu CQ, Wu FC, Yoshioka T, Sakugawa H, Tanoue E (2011) Photochemical, microbial and metal complexation behavior of fluorescent dissolved organic matter in the aquatic environments. Geochem J 45:235–254
68. Fabbrizzi L, Licchelli M, Pallavicini P, Perotti A, Taglietti A, Sacchi D (1996) Fluorescent sensors for transition metals based on electron-transfer and energy-transfer mechanisms. Chem Eur J 2: 75–82. https://doi.org/10.1002/chem.19960020114
69. Warrier S, Kharkar PS (2017) Highly selective on-off fluorescence recognition of Fe^{3+} based on a coumarin derivative and its application in live-cell imaging. Spectrochim Acta - Part A Mol Biomol Spectrosc 188:659–665. https://doi.org/10.1016/j.saa.2017.07.068
70. Güney S, Yildiz G, Capan A, Ozturk T (2010) Evaluation of the electrochemical properties of 3-hydroxyflavone using voltammetric methods. Electrochim Acta 55:3295–3300. https://doi.org/10.1016/j.electacta.2009.12.078
71. Narayanan M, Kodali G, Singh VR, Velvadapu V, Stanley RJ (2012) Oxidation and reduction potentials of 8-vinyladenosine measured by cyclic voltammetry: implications for photoinduced electron transfer quenching of a fluorescent adenine analog. J Photochem and Photobio A: Chemistry 249:9–14. https://doi.org/10.1016/j.jphotochem.2012.08.018
72. Liang Z, Liu Z, Jiang L, Gao Y (2007) A new fluorescent chemosensor for copper(II) and molecular switch controlled by light. Tetrahedron Lett 48:1629–1632. https://doi.org/10.1016/j.tetlet.2006.12.133

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

GA based optimization of tri-bological behaviour of diamond coated tungsten carbide

Ramakant Rana
Department of Mechanical and Automation Engineering, Maharaja Agrasen Institute of Technology, Delhi, India and
Department of Mechanical Engineering, Delhi Technological University, Delhi, India
Qasim Murtaza
Department of Mechanical Engineering, Delhi Technological University, Delhi, India, and
R.S. Walia
Department of Production and Industrial Engineering, Punjab Engineering College, Chandigarh, India

Abstract

Purpose – In this study, the tri-bological behaviour of the un-coated and diamond coated tungsten carbide was evaluated using the pin-on-disc test rig. The same was also tested on a lathe machine tool. This paper aims to compare the tri-bological behaviour of coated tungsten carbide pin with un-coated tungsten carbide pin it also correlates the wear obtained from the two machines used.

Design/methodology/approach – Experiments were performed using L8 orthogonal array and results obtained on a pin-on-disc test rig under dry sliding process were optimized through a modern optimization technique i.e. genetic algorithm (GA). The response surface methodology model (L8 orthogonal array) formed the basis for the development of the GA model, which defines the conditions of minimum wear, minimum coefficient of friction and minimum surface roughness for the sliding process of the pin-on-disc test rig.

Findings – Implementation of the heuristic approach for optimization of input parameters for the combination of tool material used for the turning process. The initial approach involves tri-bological testing considering the same combination. The set of experiments further performed, inferred that the results were similar and that the diamond coating enhances the life of the tool.

Originality/value – Successfully synthesized the diamond coating on tungsten carbide tool material. Implantation of the heuristic approach, i.e. GA to tri-bological tests to identify the optimized level of input variables. Experimentation involves the tri-bological testing whose results were confirmed through performing experiments on the lathe machine tool.

Keywords Genetic algorithm, Wear, Friction, WC, Tool, Diamond coating

Paper type Research paper

1. Introduction

The ever-changing demands of the modern technical system, the environmental sustainability forms the major pillar for any manufacturing process Zhang and Liu (2017). Cooling or lubrication-based machining process or wet machining process have undesirable health issues on humans and ecology as the cooling process is carried out through the use of cutting fluids or lubricating oils Krolczyk *et al.* (2017). Manufacturing industries extensively use lubricating oils and cutting fluids owing to the benefits they provide i.e. long tool life, high surface finish, easy chip removal, low temperature in the cutting zone and reduced friction at the tool-work interface Maruda *et al.* (2017), Rabiei *et al.* (2015), Park *et al.* (2015) and Rana *et al.* (2014). Effective lubrication plays a significant role in high-proficiency cutting operation Maruda *et al.* (2018). Kaynak *et al.* (2013) investigated the tool wear for Nickel-Titanium alloy under dry and minimum quantity lubrication (MQL) conditions during the machining process and yielded that tool life enhanced under MQL condition. Krolczyk *et al.* (2018) investigated the work-piece surface topography under dry turning and with the usage of minimum quantity cooling lubricant (MQCL). They inferred MQCL yielded better results against dry turning Krolczyk *et al.* (2018). However, the lubricants are associated with the numerous environmental problems involving their disposal as waste. A rapid increase has been recorded in the issues related to the environment. In recent times, the enforcement of environmental laws has increased economic problems for the manufacturing industries Klocke and Eisenblätter (1997). Another research claims that dry hard turning is beneficial because of safety concerned with

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/1708-5284.htm**

World Journal of Engineering
17/3 (2020) 335–346
© Emerald Publishing Limited [ISSN 1708-5284]
[DOI 10.1108/WJE-08-2019-0220]

Authors would like to acknowledge the support of Dr Lucky Krishnia, Research Scholar, Nano Research Laboratory of Physics Department of Delhi Technological University for preparing the samples. Authors would also like to give special thanks to Nano Research Facility of Indian Institute Technology, Delhi, India for allowing us to use their facilities for material characterization. They also extend their regards to Precision Manufacturing Lab of Delhi Technological University, Delhi, India, for their kind help in conducting the experiments.

Received 6 August 2019
Revised 24 October 2019
3 December 2019
Accepted 28 December 2019

the environment Noordin *et al.* (2007). Dry turning is advantageous owing to the elimination of cutting fluids cost and cost of fluid disposal system. This urge for the improved machined surface quality without deteriorating effects on the environment has imposed exploitation of a new research area, cutting tool materials. The manufacturing industries are now heading towards the new challenges with the demand for tougher steels Chinchanikar and Choudhury (2014). A required hardness is to be maintained during the high-speed machining of heat-treated material Mia and Dhar (2018), Mia *et al.* (2017) and Mia *et al.* (2018a, 2018b). The range of hardness for a work material if lies between 40 and 65 HRC, then hard cutting tool insert is used for its machining, known as hard turning, a subtractive process Mia *et al.* (2017) and Mia *et al.* (2018a, 2018b).

The unending urge for machining of metals in a more efficient way has led to the development of new techniques for their machining process. The hard-to-machine materials have made hard turning significant over the grinding process as surface finish and higher removal rate of the material is achievable. A huge positive response is claimed by the hard-turning process by defining new standards of quality machining Dogra *et al.* (2011). Hard turning is associated with the surface quality of the work-piece and tool wear. This process demands high hardness and high resistance to oxidation of the tool at elevated temperature Ichijo *et al.* (2007), Rafaja *et al.* (2007), Faga *et al.* (2007) and Dobrzański *et al.* (2004). Hard turning is a beneficial process against grinding process in terms of production cost, productivity, efficiency, product quality and material properties, even when used with single or multi-layer coated tool inserts Mia *et al.* (2018a, 2018b). This process demands the right tool with high tool life for a work material to reduce power consumption. Presently, the forthcoming demands of the manufacturing industries are superior cutting tool inserts for nearly all machining operations Chinchanikar and Choudhury (2014). Ozel *et al.* (2005) concluded that cutting forces are dependent on cutting conditions and geometry of the cutting edges and work-piece hardness in hard turning. It was inferred that work surface softens at high temperature while low cutting-edge radius yielded low radial and tangential force components. Hard coating for cutting tools are being extensively researched in past decades. The hardness of the coatings is now explicitly examined in terms of the range of coatings and their thickness varying from a few nanometers to a few millimetres. These help to enhance tool life as much as 10 times through a reduction in wear rate and further increase the productivity of the firm. Kupczyk and Komolka *et al.* (2015) deduced coated tungsten carbide-cobalt tool insert to be very efficient in numerous processes such as turning, milling, etc. The strength of the tool material lies in the refined grain structure along with enhanced mechanical properties Kupczyk and Komolka (2015). Development of thin films through various coating deposition techniques has revolutionized the field of tool coating materials. Ferreira *et al.* (2016) discussed the conventional and wiper ceramic tool on hard turned AISI H13 steel inferring that wiper ceramic tool is better in terms of the functioning of the tool. Sahoo *et al.* (2012) investigated performance comparison between coated (both TiN and ZrCN) and un-coated carbide tool inserts yielding that coated surfaces perform better during the machining operations and among TiN and ZrCN coatings, TiN concluded better results. Optimization was carried out by Kumar *et al.* (2017) for surface roughness, energy consumption during the machining process and material removal rate considering four input parameters viz. nose radius, cutting speed, depth of cut and feed rate through Taguchi approach. The model considered the effects of the interaction of the process parameters Kumar *et al.* (2017).

The extensive research of the previous literature on hard turning, hard coating, tool coating materials and parameter optimization, it is distinctly visible that ceramic, CBN and PCBN tools are readily used for machining of hard-to-machine materials. Research is still in demand for the successful utilization of coated carbide inserts due to their low thermal stability and high rate of oxidation at high temperature specifically in the range 900-1,200°C (Singh and Singhal, 2018; Singh *et al.*, 2018). Different treatments such as coatings or surface modification processes can be performed to improve the properties of WC. Diamond or diamond like carbon (DLC) coatings has good chemical inertness Grill (1999) and Donnet and Erdemir (2008). Wear behaviour of DLC coatings deposited on WC substrate or metallic substrates has been studied in different sliding situations but not much for diamond coatings Donnet and Erdemir (2008), Erdemir and Donnet (2006) and Pech *et al.* (2009). The application of diamond coatings on WC is growing but issues about wear and resistance of coatings remain poorly researched. In this work, tri-biological tests were carried out on diamond coating deposited on WC, with different loads and speed keeping the goal of adding some scientific and technological knowledge to this field Navaneethakrishnan *et al.* (2009). It is crucial to optimize the tri-biological properties of the coated material, study the relationship between process parameters, structure and properties and evaluate the behaviour of the coatings Lal and Singh (2018), Lal and Singh (2019) and Lata *et al.* (2018).

In the present work, a diamond coating is prepared on Tungsten Carbide pins by thermal chemical vapour deposition method. The tri-biological properties and wear mechanism of diamond coating were analyzed. The diamond coatings so produced is analyzed in term of coefficient of friction (COF), wear loss and surface roughness produced. The aim of the work is to find out the optimal conditions for low coefficient of friction, good wear resistance and less surface roughness using a heuristic approach i.e. genetic algorithm (GA). Diamond coating on WC substrate pins is selected keeping in mind their present application as tool inserts. The present study is relevant in exploring the opportunity of diamond coating to achieve a lower coefficient of friction, lower wear loss and lesser surface roughness.

2. Experimental details

2.1 Selection of material

The selected work material for the experimental investigation was IS: 2062 steel specified as hot-rolled structural steels available in three grades, low, medium and high tensile strength. High tensile strength steel was used in this study. The sample was in the form of a solid disc, which was tested for the identification of its chemical composition. The dimensions of

the disc were 100 mm diameter and 10 mm in height. Table I presents the chemical composition of the steel shaft. The elemental weight percentage of tungsten carbide pin is shown in Figures 1 and Table II.

Tungsten carbide pins were cut and pin holder was designed and fabricated to hold the pins properly on the pin-on-disc test rig. In this experimental work, un-coated and diamond coated tungsten carbide pins were used during the dry sliding process. The Schematic diagram of a un-coated pin and diamond coated pin are shown in Figure 2 (a) and (b).

2.1.1 Coating process

A film of diamond was deposited on the tungsten carbide substrate material using thermal-chemical vapour deposition method. The coating was coated at reduced atmospheric pressure and very controlled temperature. Table II shows the chemical composition of tungsten carbide substrate pin. Initially, the tungsten carbide substrates were pre-treated for 7-10 min, this pre-treatment consists of cleaning samples with acetone and concentrated sulphuric acid (H_2SO_4). At the same time, sugarcane bagasse (SBg) produced from the sugarcane was put in the split furnace while maintaining the 550°C temperature at atmospheric pressure for producing the pyrolyzed SBg (*p*-SBg). The pyrolyzed SBg (*p*-SBg), was chemically treated to get activated carbon. After that the produced activated carbon and pre-treated substrate material was put in the reaction chamber for the growth of the diamond coating. Also, a mixture of Ar/H_2 gases was supplied during the thermal CVD process to remove oxygen contents at atmospheric pressure at 900°C temperature. This process develops a diamond film on the tungsten carbide substrate Krishnia *et al.* (2018), Krishnia and Tyagi (2018) and Tyagi *et al.* (2019).

2.2 Pin-on-disc test rig

The tri-bological behaviour of un-coated tungsten carbide pin and diamond coated tungsten carbide pin against structural steel was investigated under dry sliding conditions in terms of friction, wear and surface roughness on a pin-on disc test rig in accordance with ASTM Standard G99 (2017). IS: 2062 discs (ϕ 100 mm) were ground to the roughness of ranging from 0.15 to 0.45 μm and were secured to the rotating axis of the machine using a lock screw. Surface roughness was measured by Talysurf; eight readings along the radial direction at an interval of 45° and eight readings perpendicular to it from the middle of the stroke of Talysurf. The experiments were performed at ambient temperature under a dry condition at the range of 19°C to 21°C, to record the coefficient of friction and wear. A wear sensor and friction sensors were used to continuously measure the wear and coefficient of friction. The pin-on-disc test rig, corresponding tribo-pairs and the accessories are shown in Figure 3 Singh *et al.* (2016).

Tests were carried out using the pin-on-disc tribometer under boundary lubrication conditions. The disc was rotated against the stationary pin. The test conditions are described in Table III. For each material couple, the test was carried out more than three times to check the repeatability. Before testing, the test samples were cleaned with acetone in ultrasonic bath for 10 min Stachowiak and Batchelor (2005).

2.3 Selection of process parameters

The tests parameters selected for carrying out the dry sliding process of tungsten carbide pin on the pin-on-disc test rig were pin diameter (mm), load (N) and sliding speed (m/s). The pins of diameter 5.5 mm and 3.5 mm were used, and the sliding speed of the pin and disc at their interface was maintained at 2.4 m/s (144 m/min) and 3.2 m/s (192 m/min). The load applied on the pins was 20 N and 30 N. The complete L8 orthogonal array used for the experimentation is given in Table IV. The experiments using a coated and un-coated pin were run on the same L8 orthogonal array so that the results can be compared and shown easily. The selection of speed values was dependent on the speed range of surfaces under investigation. The available range of the parameters used on pin-on-disc test rig is mentioned in Table V.

2.4 Development of predictive model

GA being a heuristic optimization technique develops its model based on the regression equations generated through regression analysis. As the main effects of process parameters are focused in this experimentation, only the linear regression equations were generated. These equations are further transformed into predictive equations considering specific range of the process parameters. This section discusses a general form of GA based predictive equations. Hence, the lowest value of range defines the lower bound while highest value defines the upper bound for these predictive equations.

The GA parameters considered for the optimization are selected as population type: double vector, selection: Stochastic uniform, Elite count: 20 per cent, crossover fraction: 0.8, mutation: 20 per cent and crossover function: scattered. The wear, coefficient of friction and surface roughness were expressed in terms of diameter, load and speed. In general form, the response function can be expressed as follows:

$$Y = \Phi(V, L, D) \quad (1)$$

where

Y = Response (wear/coefficient of friction/surface roughness);
V = Sliding speed (m/s);
L = Load applied (N);
D = Diameter (mm).

Using the following regression equation:

$$Y = \beta_0 + \beta_1 V + \beta_2 L + \beta_3 D \quad (2)$$

where β_0 is constant while β_1, β_2 and β_3 are partial regression coefficients of the model from which the response functions were generated as given below:

Table I Chemical composition of steel shaft (**IS: 2062**)

C	Si	Mn	P	S	Cr	Mo	Ni	Al	Cu	Ti	V
0.213	0.196	0.760	0.036	0.037	0.161	0.007	0.094	0.032	0.169	0.003	0.008

Figure 1 EDAX of tungsten carbide cutting insert

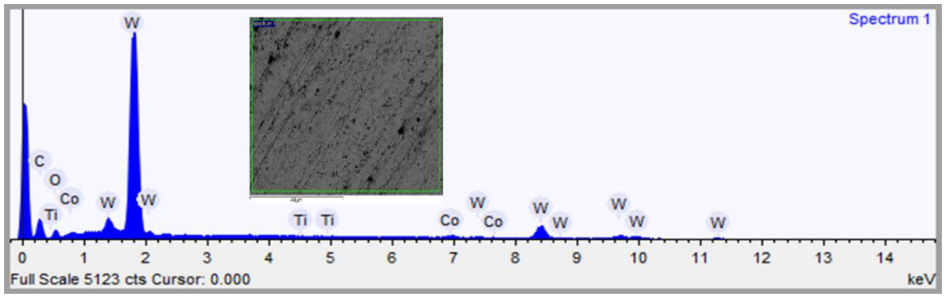


Figure 2 Schematic diagram of (a) un-coated pin (b) diamond coated pin

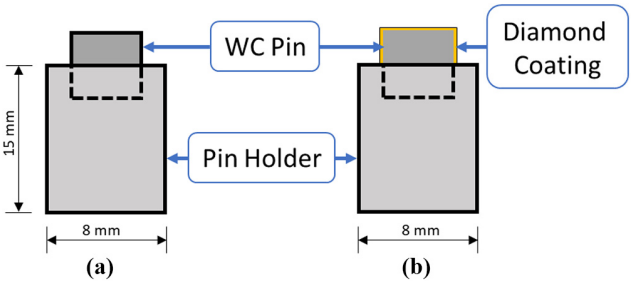


Figure 3 Set-up of pin-on-disc test rig for experimentation

Table II Chemical composition of tungsten carbide cutting insert

Element	Wt.%
Carbon	20.046
Oxygen	5.307
Titanium	0.000
Cobalt	1.607
Tungsten	73.041

$$W = C_1 * V^{x1} * L^{x2} * D^{x3} \quad (3)$$

$$COF = C_2 * V^{x1} * L^{x2} * D^{x3}$$

$$SR = C_3 * V^{x1} * L^{x2} * D^{x3}$$

where

W = Wear;
COF = Coefficient of Friction;
SR = Surface Roughness
C_1, C_2, C_3, = Constant
V = Sliding speed (m/s)
L = Load applied (Kg)
D = Diameter (mm)

The unknown coefficients, namely, exponents of sliding speed, load applied and diameter and constants involved in above equations were determined by minimizing the least-squares error between experimental and predicted results obtained within the range of cutting conditions.

Table III Tribological test conditions used for the experimentation

Parameters	Conditions
Load	20 and 30 (N)
Sliding speed	2.4 and 3.2 (m/s)
Pin diameter	3.5 and 5.5 (mm)
Lubrication	Dry sliding
Sliding distance	1,000 m

Table IV L8 Orthogonal array used for the experimentation

S. no.	Pin diameter (mm)	Load (N)	Speed (m/s)
1	5.5	30	3.2
2	5.5	30	2.4
3	5.5	20	3.2
4	5.5	20	2.4
5	3.5	30	3.2
6	3.5	30	2.4
7	3.5	20	3.2
8	3.5	20	2.4

Table V Available range of the parameters used on pin-on-disc test rig

S. no.	Parameters	Range	
		From	To
1	RPM	300	3,000
2	Load	20 N	200 N

2.5 Optimization using genetic algorithm

GA is one of the modern optimization techniques, which is heuristic in nature. It creates the population through reproduction via three methods; selection, crossover and mutation. After setting the fixed parameters such as elite population, crossover function, mutated population, etc. the optimized output parameters were yielded. The population count considered was 50 and the elite, as well as the mutated population, was taken as 20 per cent from the previous population. Through the M-file, the GA model was run for the considered output parameters i.e. wear, coefficient of friction and surface roughness, which have been discussed below.

3. Results and discussion

3.1 Material characterization

The Hitachi made scanning electron microscope (SEM) was used to analyze the surface morphology (Figure 4). To evaluate the presence of a carbon peak of diamond coating, the micro-Raman spectrometer was performed with a spectral range of 100 cm^{-1} to 3,500 cm^{-1}. Through Figure 5, the D-Band peak was found at 1,338.54 cm^{-1} and G-Band peak was found at 1,590.99 cm^{-1} Polini *et al.* (2003).

3.2 Record of results

As per the L8 orthogonal array, the variation of wear, coefficient of friction and surface roughness for coated and un-coated pins are presented in Table VI. It is observed that the wear, coefficient of friction and surface roughness have mostly decreased for diamond coated pins. The results obtained from the pin-on-disc testing were optimized used GA and have presented in this section. Figure 6 presents the variation of wear for the different experimental runs. The minimum amount of wear was obtained at 3.2 m/s sliding speed, 20 N load and 3.5 mm pin diameter input factors. Figure 7 shows the variation of the coefficient of friction. The minimum coefficient of friction was obtained at a sliding speed of 3.2 m/s, 20 N load and 5.5 mm pin diameter. Similarly, Figure 8 gives the variation of surface roughness. The surface roughness obtained was least at 2.4 m/s sliding speed, 20 N load and 3.5 mm pin diameter Tyagi *et al.* (2019).

Figure 4 Scan electron microscopic image of diamond coated pin

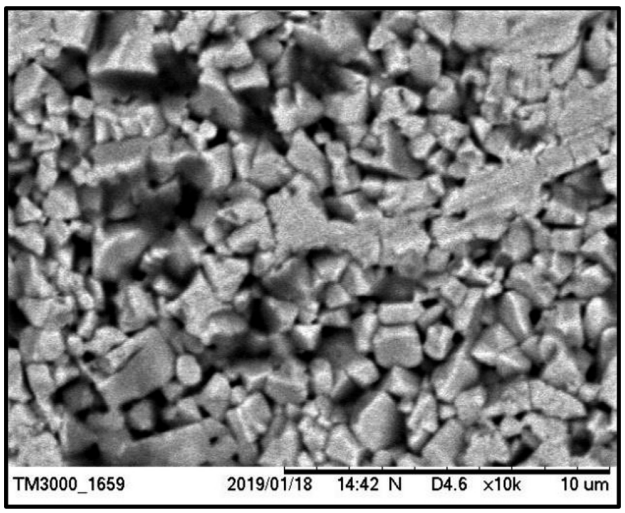
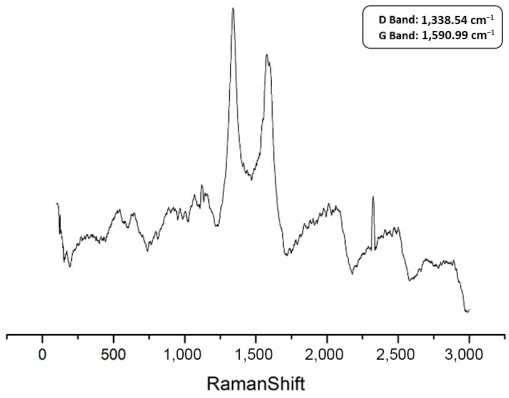


Figure 5 Raman spectra graph of diamond coated pin



3.3 Optimization by genetic algorithm

GA predictive model was considered for optimization talking into account all the process parameters individually. This approach was adopted to analyze the individual effect of process parameters on the considered output parameters. Also, the experimentation on pin-on-disc test rig yielded results considering individual parameters. This helped in building a similar comparison of turning and tribological tests.

3.3.1 Wear of coated and un-coated pin

The equation developed using regression analysis for the output parameter "Wear" is given as follows Rana *et al.* (2018):

$$Wear = 0.2 - 0.363\ Diameter + 0.039\ Load + 1.13\ Speed \quad \text{(Un-Coated)}$$

$$Wear = 1.28 - 0.164\ Diameter + 0.0457\ Load + 0.246\ Speed \quad \text{(Coated)}$$

The predictive equation developed for the optimization model is as follows:

$$Wear = 1.221 \times x_1^{-0.363} \times x_2^{0.039} \times x_3^{1.13} \quad \text{(Un-Coated)}$$

$$Wear = 3.596 \times x_1^{-0.164} \times x_2^{0.0457} \times x_3^{0.246} \quad \text{(Coated)}$$

x_1, x_2 and x_3 are pin diameter, load and sliding speed, respectively.

After setting the fixed parameters, the yielded optimized output parameter for the un-coated pin was 3.71 μm and for

Table VI Variation of wear, coefficient of friction and surface roughness

S. No.	Wear (μm)		Coefficient of friction		Surface roughness (Ra)	
	Un-coated	Diamond coated	Un-coated	Diamond coated	Un-coated	Diamond coated
1	22.32	17.63	0.08	0.07	0.302	0.259
2	10.61	9.01	0.22	0.19	0.142	0.130
3	11.08	8.08	0.06	0.07	0.389	0.302
4	6.94	6.31	0.19	0.17	0.175	0.163
5	20.56	21.18	0.30	0.37	0.172	0.210
6	15.28	12.04	0.22	0.20	0.398	0.324
7	8.60	7.39	0.07	0.08	0.221	0.175
8	18.39	16.55	0.18	0.16	0.135	0.158

Figure 6 Variation of wear of the diamond coated pin and un-coated pin specimens as per L8 orthogonal array

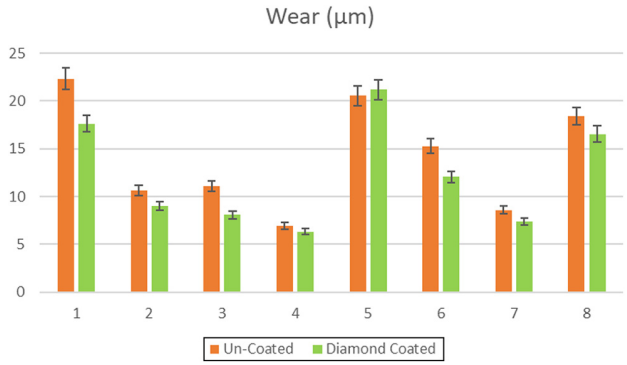
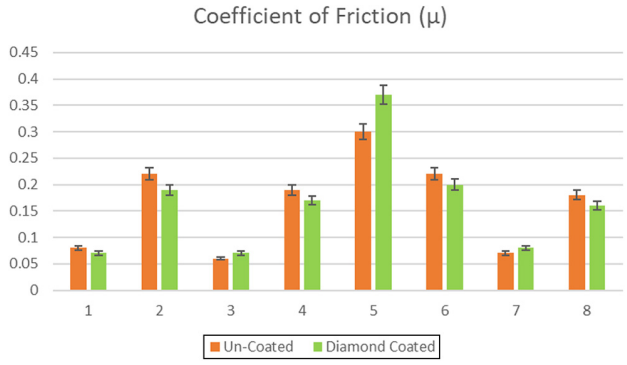


Figure 7 Variation of coefficient of friction of the diamond coated pin and un-coated pin as per L8 orthogonal array

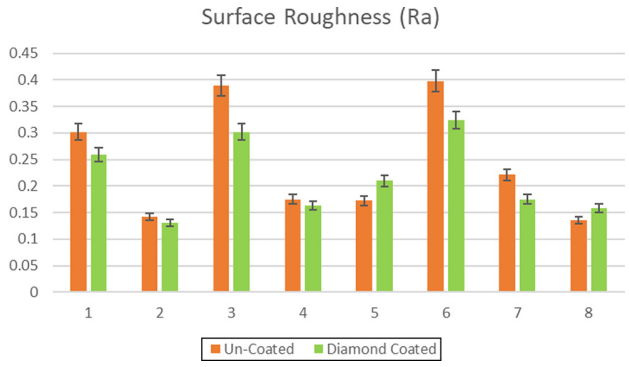


diamond coated pin was 1.99 μm, which is shown in Figure 9. The optimized value of wear was obtained at a sliding speed of 3.2 m/min, a load of 20.01 N and diameter of 3.50 mm for un-coated pin and the optimized value of wear was obtained at a sliding speed of 2.4 m/min, a load of 20.16 N and diameter of 5.5 mm for coated pin (Figure 10).

3.3.2 Coefficient of friction of coated and un-coated pin

The equation developed using regression analysis for the output parameter "Coefficient of Friction" is given as follows Rana *et al.* (2018):

Figure 8 Variation of surface roughness of the diamond coated pin and un-coated pin as per L8 orthogonal array



$$Coefficient\ of\ Friction = -0.78 - 0.054\ Diameter + 0.0272\ Load - 0.51\ Speed \quad \text{(Un-Coated)}$$

$$Coefficient\ of\ Friction = -1.19 - 0.149\ Diameter + 0.0466\ Load - 0.427\ Speed \quad \text{(Coated)}$$

The predictive equation developed for the optimization model is:

$$Coefficient\ of\ Friction = 0.458 \times x_1^{-0.054} \times x_2^{0.0272} \times x_3^{-0.51} \quad \text{(Un-Coated)}$$

$$Coefficient\ of\ Friction = 0.3042 \times x_1^{-0.149} \times x_2^{0.0466} \times x_3^{-0.427} \quad \text{(Coated)}$$

x_1, x_2 and x_3 are pin diameter, load and sliding speed, respectively.

During the optimization process, the optimized value of the coefficient of friction of un-coated pin was yielded as 0.196 and for diamond coated pin it was 0.165, which is shown in Figure 11. These optimized values of coefficient of friction were obtained at sliding speed of 3.19 m/min, a load of 20.35 N and a diameter of 5.5 mm for un-coated pin and at sliding speed of 3.2 m/min, a load of 20.00 N and a diameter of 5.5 mm for diamond coated pin (Figure 12).

Figure 9 Genetic algorithm graph for the best value of wear

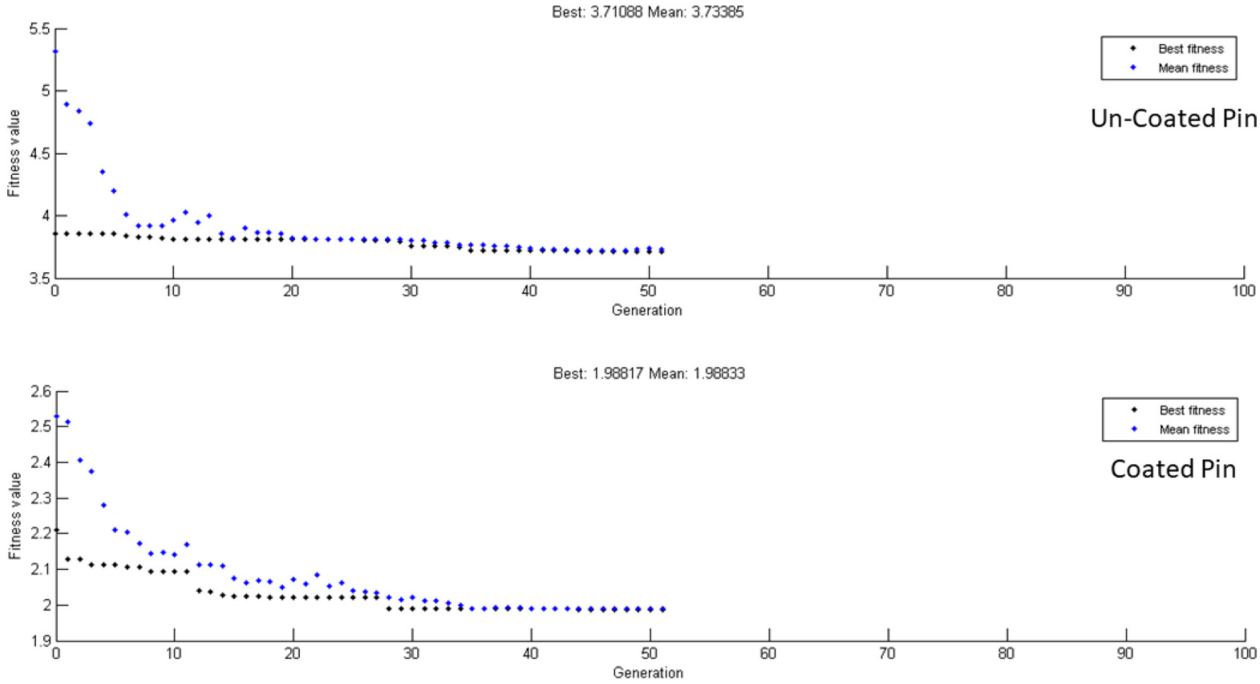


Figure 10 Genetic algorithm graph for the average value of wear

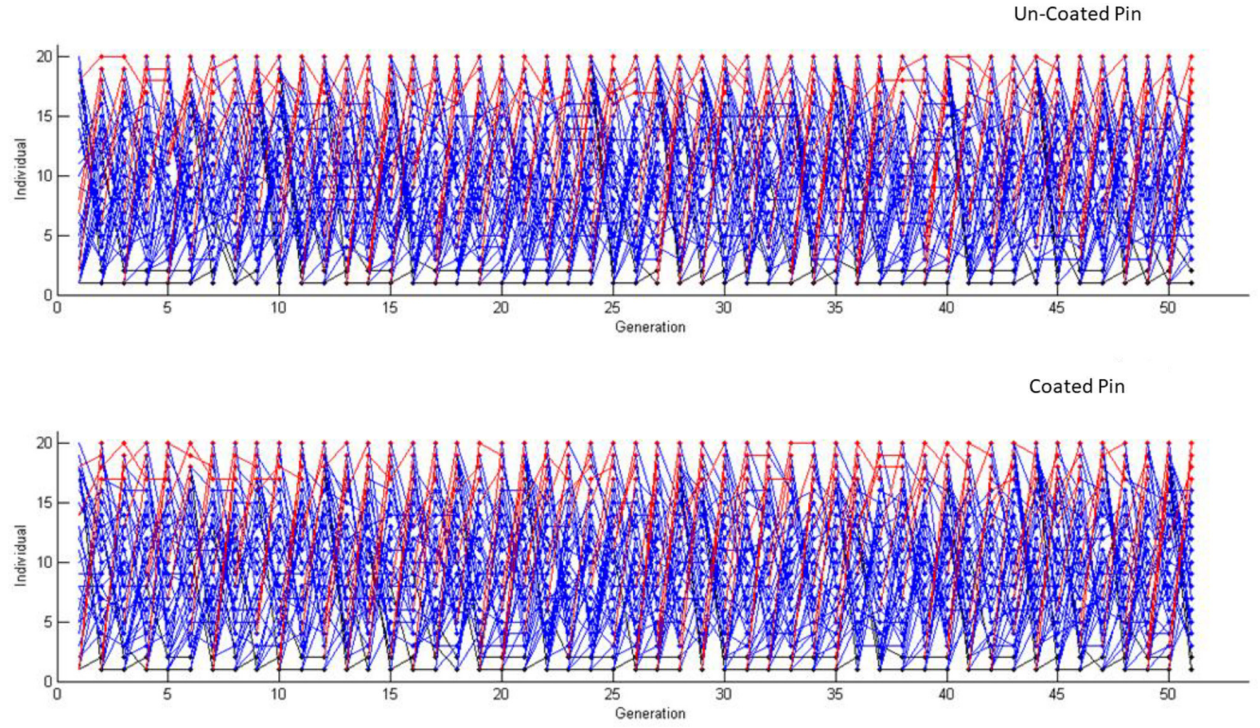


Figure 11 Genetic algorithm graph for the best value of coefficient of friction

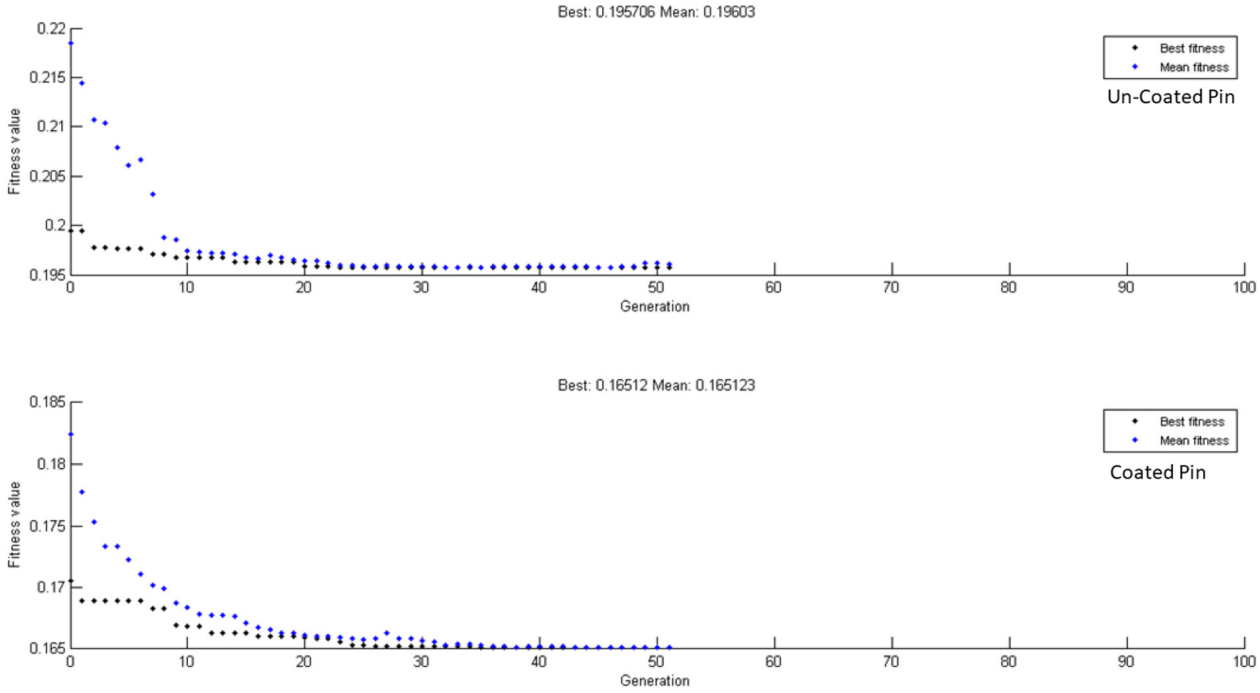
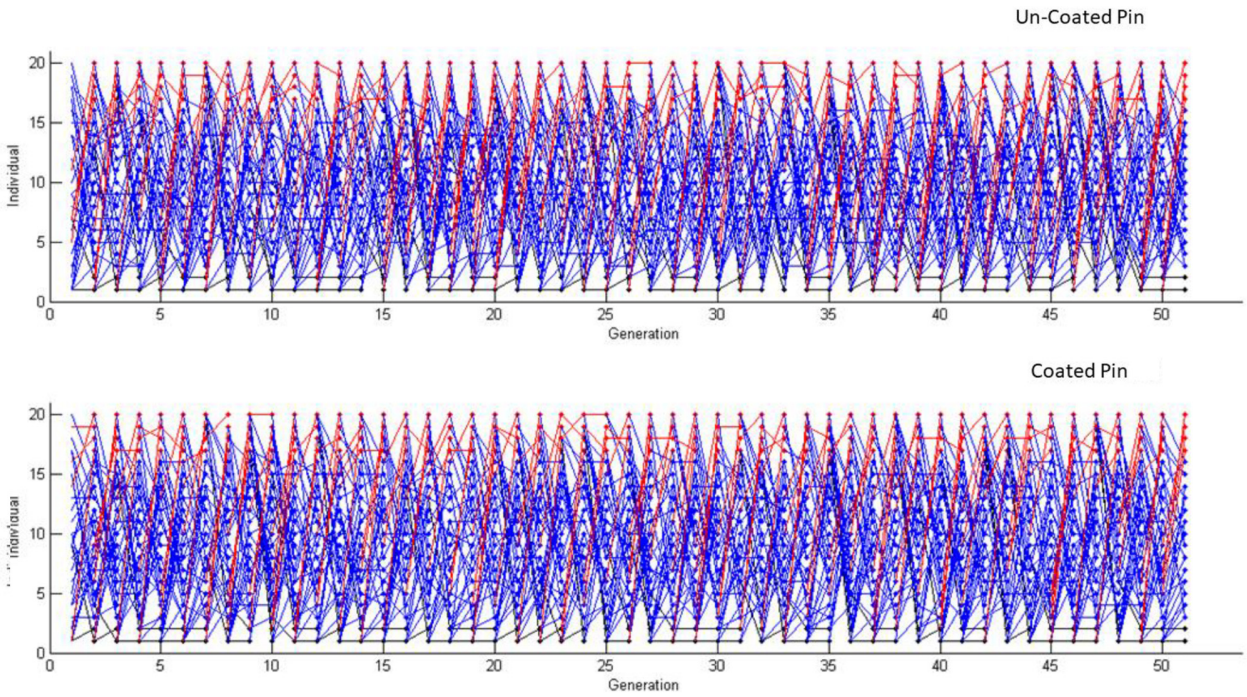


Figure 12 Genetic algorithm graph for the average value of coefficient of friction



3.3.3 Surface roughness of coated and un-coated pin

The equation developed using regression analysis for the output parameter "surface roughness" is given as follows Rana *et al.* (2018):

$$Surface\ Roughness = -2.57 + 0.022\ Diameter + 0.0036\ Load + 0.321\ Speed \quad \text{(Un-Coated)}$$

$$Surface\ Roughness = -2.79 - 0.057\ Diameter + 0.0205\ Load + 0.305\ Speed \quad \text{(Coated)}$$

The predictive equation developed for the optimization model is:

$$Surface\ Roughness = 0.0765 \times x_1^{-0.022} \times x_2^{0.0036} \times x_3^{-0.321} \quad \text{(Un-Coated)}$$

$$Surface\ Roughness = 0.0614 \times x_1^{-0.057} \times x_2^{0.0205} \times x_3^{0.305} \quad \text{(Coated)}$$

During the optimization process, the optimized value of surface roughness was obtained as 0.118 μm for the un-coated pin and 0.105 μm for diamond coated pin, which is shown in Figure 13. The optimized value for un-coated pin was obtained at a sliding speed of 2.4 m/min, a load of 20 N and a diameter of 3.5 mm and for diamond coated pin it was obtained at a sliding speed of 2.4 m/min, a load of 20 N and a diameter of 3.5 mm (Figure 14).

4. Confirmatory experiments

A set of confirmatory experiments were performed on completion of experimentation. The tri-biological behaviour was studied between steel and tungsten carbide pin (un-coated and diamond coated) using a pin-on-disc test rig. The results obtained were further optimized by a heuristic approach to identify the optimum input parameters for a minimum value of wear, coefficient of friction and surface roughness. On determining the optimum levels, a set of experiments were conducted during the turning process at equivalent cutting speed. It was inferred that the tool wear value was similar to the wear value obtained during tri-biological test and was within tolerance limits. It was also concluded that the amount of wear in coated tool was high as compared to coated tool. These results were similar to tri-biological results Mishra *et al.* (2019). Figures 15 and 16 shows the wear of un-coated and coated tungsten carbide tool inserts used during the experimentation.

5. Conclusions

An attempt was made to test the material combination of tool and work-piece in the tri-biological testing equipment in the form of disc and pins. Similar to tool coatings, pins were also coated and were tested against the uncoated pins. Following conclusions were drawn from the experimental investigations:

- A comparison was conducted between the experimental values of un-coated and diamond coated pins, leading the study in the favour of diamond coated pins.
- A heuristic approach, i.e. GA was successfully applied in the tri-biological experimentation.
- The optimized value of wear after yielding the parameter for un-coated pin was 3.71 μm and for diamond coated pin was 1.98 μm.
- The optimized value of wear was obtained at a sliding speed of 3.2 m/min, a load of 20.085 N and diameter of 3.502 mm for un-coated pin and the optimized value of wear was obtained at a sliding speed of 2.4 m/min, a

Figure 13 Genetic algorithm graph for the best value of surface roughness

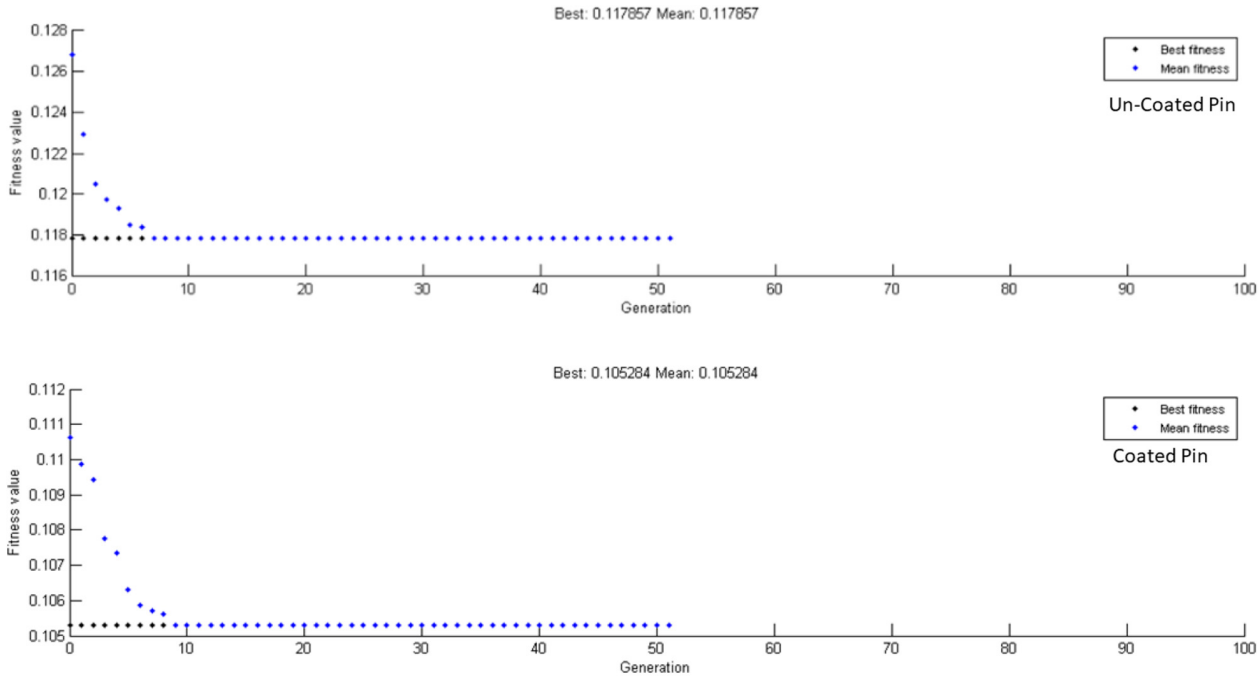


Figure 14 Genetic algorithm graph for the average value of surface roughness

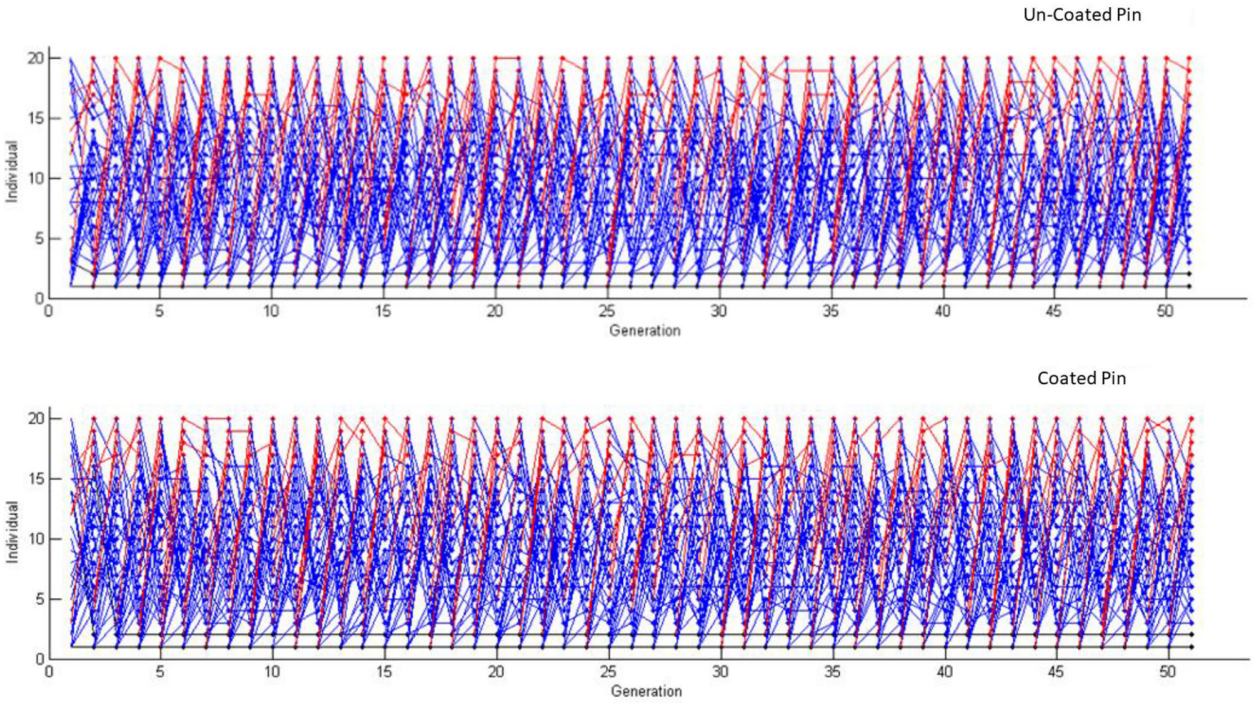
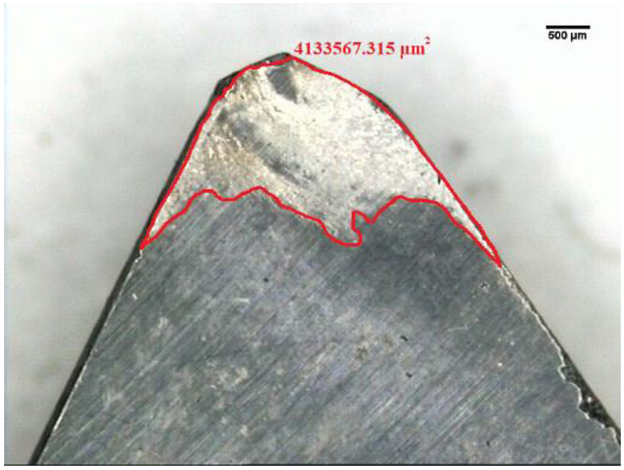


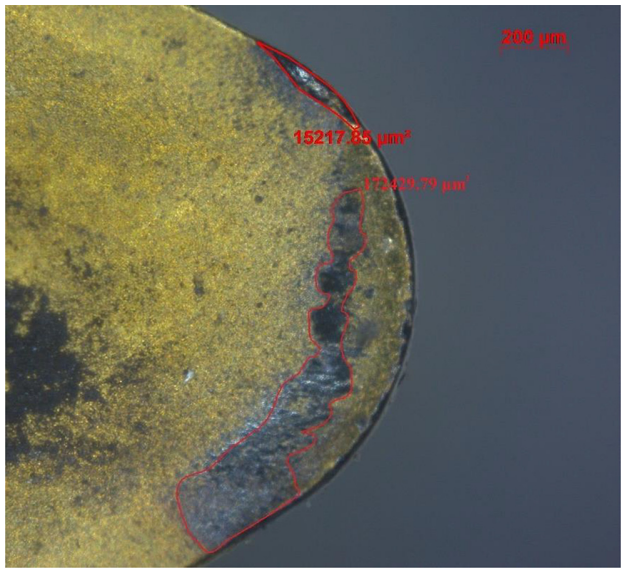
Figure 15 Tool wear of un-coated tungsten carbide tool insert during turning process



load of 20.164 N and diameter of 5.5 mm for coated pin.

- The optimized value of the coefficient of friction for un-coated pin was 0.196 and for coated pin, it was 0.165.
- The optimized values of coefficient of friction were obtained at sliding speed of 3.19 m/min, a load of 20.35 N and a diameter of 5.5 mm for un-coated pin and for diamond coated pin they were obtained at sliding speed of 3.2 m/min, a load of 20 N and a diameter of 5.5 mm.

Figure 16 Tool wear of diamond coated tungsten carbide tool insert during turning process



- The optimized value of surface roughness for un-coated pin was 0.118 μm for un-coated pin and 0.105 μm for diamond coated pin.
- The optimized values of surface roughness for un-coated pin was obtained at a sliding speed of 2.4 m/min, a load of 20 N and a diameter of 3.5 mm and for un-coated pin, it was obtained at a sliding speed of 2.4 m/min, a load of 20 N and a diameter of 3.5 mm.

- The optimized results were further confirmed by performing the experiments on lathe machine tools during tuning operations. The results so, obtained were in accordance to the tri-biological results obtained on pin-on-disc test rig within the tolerance limit.
- Based on the outcomes from the current studies, diamond coating can be recommended for tungsten carbide for better wear resistance, less friction and low surface roughness by current coating technique.
- The diamond coating on tungsten carbide inserts can be used on various machining processes while varying the process parameters.

References

ASTM Standard G99 (2017), *Standard Test Method for Wear Testing with a Pin-on-Disk Apparatus*, ASTM International, West Conshohocken, PA, doi: 10.1520/G0099-17.

Chinchanikar, S. and Choudhury, S.K. (2014), "Hard turning using HiPIMS-coated carbide tools: wear behavior under dry and minimum quantity lubrication (MQL)", *Measurement*, Vol. 55, pp. 536-548.

Dobrzański, L.A., Pakuła, D. and Hajduczek, E. (2004), "Structure and properties of the multi-component TiAlSiN coatings obtained in the PVD process in the nitride tool ceramics", *Journal of Materials Processing Technology*, Vols 157/158, pp. 331-340.

Dogra, M., Sharma, V.S., Sachdeva, A., Suri, N.M. and Dureja, J.S. (2011), "Performance evaluation of CBN, coated carbide, cryogenically treated uncoated/coated carbide inserts in finish-turning of hardened steel", *The International Journal of Advanced Manufacturing Technology*, Vol. 57 Nos 5/8, pp. 541-553.

Donnet, C. and Erdemir, A. (2008), *Tribology of Diamond-like Carbon Films, Fundamentals and Applications*, Springer.

Erdemir, A. and Donnet, C. (2006), "Tribology of diamond-like carbon films: recent progress and future prospects", *J Phys D: Appl Phys*, Vol. 39, pp. 311-327.

Faga, M.G., Gautier, G., Calzavarini, R., Perucca, M., Boot, E.A., Cartasegna, F. and Settineri, L. (2007), "AlSiTiN nanocomposite coatings developed via arc cathodic PVD: evaluation of wear resistance via tribological analysis and high speed machining operations", *Wear*, Vol. 263 Nos 7/12, pp. 1306-1314.

Ferreira, R., Řehoř, J., Lauro, C.H., Carou, D. and Davim, J.P. (2016), "Analysis of the hard turning of AISI H13 steel with ceramic tools based on tool geometry: surface roughness, tool wear and their relation", *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, Vol. 38 No. 8, pp. 2413-2420.

Grill, A. (1999), "Diamond-like carbon: state of the art", *Diamond and Related Materials*, Vol. 8 Nos 2/5, pp. 428-434.

Ichijo, K., Hasegawa, H. and Suzuki, T. (2007), "Microstructures of (Ti, Cr, Al, Si)N films synthesized by cathodic arc method", *Surface and Coatings Technology*, Vol. 201 Nos 9/11, pp. 5477-5480.

Kaynak, Y., Karaca, H.E., Noebe, R.D. and Jawahir, I.S. (2013), "Tool-wear analysis in cryogenic machining of NiTi shape memory alloys: a comparison of tool-wear performance with dry and MQL machining", *Wear*, Vol. 306 Nos 1/2, pp. 51-63.

Klocke, F. and Eisenblätter, G. (1997), "Dry cutting", *CIRP Annals*, Vol. 46 No. 2, pp. 519-526.

Krishnia, L. and Tyagi, P.K. (2018), "Growth and characterization of polycrystalline diamond films on silicon using sugarcane bagasse as carbon precursor at atmospheric pressure by thermal chemical vapor deposition", *Diamond and Related Materials*, Vol. 87, pp. 18-26.

Krishnia, L., Brajesh, Y.S., P., Umesh, Satyam, P.V., G.B., Kumar, K., Nikhil, A. and Tyagi Pawan, K. (2018), "As-pyrolyzed sugarcane bagasse possessing exotic field emission properties", *Applied Surface Science*, Vol. 443, pp. 184-190.

Krolczyk, G.M., Nieslony, P., Maruda, R.W. and Wojciechowski, S. (2017), "Dry cutting effect in turning of a duplex stainless steel as a key factor in clean production", *Journal of Cleaner Production*, Vol. 142, pp. 3343-3354.

Krolczyk, G.M., Maruda, R.W., Krolczyk, J.B., Nieslony, P., Wojciechowski, S. and Legutko, S. (2018), "Parametric and nonparametric description of the surface topography in the dry and MQCL cutting conditions", *Measurement*, Vol. 121, pp. 225-239.

Kumar, R., Bilga, P.S. and Singh, S. (2017), "Multi objective optimization using different methods of assigning weights to energy consumption responses, surface roughness and material removal rate during rough turning operation", *Journal of Cleaner Production*, Vol. 164, pp. 45-57.

Kupczyk, M.J. and Komolka, J. (2015), "High durability of cutting insert edges made of nanocrystalline cemented carbides", *International Journal of Refractory Metals and Hard Materials*, Vol. 49, pp. 225-231.

Lal, R. and Singh, R.C. (2018), "Experimental comparative study of chrome steel pin with and without chrome plated cast iron disc in situ fully flooded interface lubrication", *Surf. Topogr.: Metrol. Prop*, Vol. 6, p. 35001.

Lal, R. and Singh, R.C. (2019), "Investigations of tribodynamic characteristics of chrome steel pin against plain and textured surface cast iron discs in lubricated conditions", *World Journal of Engineering*, Vol. 16 No. 4, pp. 560-568.

Lata, S., Rana, R. and Hitesh, R. (2018), "Investigation of chip-tool interface temperature: effect of machining parameters and tool material on ferrous and non-ferrous metal", *Materials Today: Proceedings*, Vol. 5 No. 2, pp. 4250-4257, doi: 10.1016/j.matpr.2017.11.689.

Maruda, R.W., Krolczyk, G.M., Michalski, M., Nieslony, P. and Wojciechowski, S. (2017), "Structural and microhardness changes after turning of the AISI 1045 steel for minimum quantity cooling lubrication", *Journal of Materials Engineering and Performance*, Vol. 26 No. 1, pp. 431-438.

Maruda, R.W., Krolczyk, G.M., Wojciechowski, S., Zak, K., Habrat, W. and Nieslony, P. (2018), "Effects of extreme pressure and anti-wear additives on surface topography and tool wear during MQCL turning of AISI 1045 steel", *Journal of Mechanical Science and Technology*, Vol. 32 No. 4, pp. 1585-1591.

Mia, M. and Dhar, N.R. (2018), "Modeling of surface roughness using RSM, FL and SA in dry hard turning", *Arabian Journal for Science and Engineering*, Vol. 43 No. 3, pp. 1125-1136.

Mia, M., Rifat, A., Tanvir, M.F., Gupta, M.K., Hossain, M.J. and Goswami, A. (2018a), "Multi-objective optimization of chip-tool

interaction parameters using grey-Taguchi method in MQL-assisted turning", *Measurement*, Vol. 129, pp. 156-166.
Mia, M., Dey, P.R., Hossain, M.S., Arafat, M.T., Asaduzzaman, M., Shoriat Ullah, M. and Zobaer, S.T. (2018b), "Taguchi S/N based optimization of machining parameters for surface roughness, tool wear and material removal rate in hard turning under MQL cutting condition", *Measurement*, Vol. 122, pp. 380-391.
Mia, M., Razi, M.H., Ahmad, I., Mostafa, R., Rahman, S.M. S., Ahmed, D.H., Dey, P.R. and Dhar, N.R. (2017), "Effect of time-controlled MQL pulsing on surface roughness in hard turning by statistical analysis and artificial neural network", *The International Journal of Advanced Manufacturing Technology*, Vol. 91 Nos 9/12, pp. 3211-3223.
Mishra, S.K., Ghosh, S. and Aravindan, S. (2019), "Physical characterization and wear behavior of laser processed and PVD coated WC/Co in dry sliding and dry turning processes", *Wear*, Vol. 428, pp. 93-110.
Navaneethakrishnan, P., Ganesh Sundara Raman, S., Pathak, S. D., Gnanamoorthy, R. and Ravi, N. (2009), "Fretting wear studies on diamond-like carbon coated Ti–6Al–4V", *Surface and Coatings Technology*, Vol. 203 No. 9, pp. 1205-1212.
Noordin, M.Y., Venkatesh, V.C. and Sharif, S. (2007), "Dry turning of tempered martensitic stainless tool steel using coated cermet and coated carbide tools", *Journal of Materials Processing Technology*, Vol. 185 Nos 1/3, pp. 83-90.
Ozel, T., Hsu, T.-K. and Zeren, E. (2005), "Effects of cutting edge geometry, workpiece hardness, feed rate and cutting speed on surface roughness and forces in finish turning of hardened AISI H13 steel", *The International Journal of Advanced Manufacturing Technology*, Vol. 25 Nos 3/4, pp. 262-269.
Park, K.-H., Yang, G.-D., Suhaimi, M.A., Lee, D.Y., Kim, T.-G., Kim, D.-W. and Lee, S.W. (2015), "The effect of cryogenic cooling and minimum quantity lubrication on end milling of titanium alloy Ti-6Al-4V", *Journal of Mechanical Science and Technology*, Vol. 29 No. 12, pp. 5121-5126.
Pech, D., Schupp, N., Steyer, P., Hack, T., Gachon, Y., Héau, C., Loir, A.S., Carlos, J. and Lópezd, S. (2009), "Duplex SiCN/DLC coating as a solution to improve fretting –corrosion resistance of steel", *Wear*, Vol. 266 Nos 7/8, pp. 832-838.
Polini, R., Casadei, F., D'Antonio, P. and Traversa, E. (2003), "Dry turning of alumina/aluminum composites with CVD diamond coated Co-cemented tungsten carbide tools", *Surface and Coatings Technology*, Vol. 166 Nos 2/3, pp. 127-134.
Rabiei, F., Rahimi, A.R., Hadad, M.J. and Ashrafijou, M. (2015), "Performance improvement of minimum quantity lubrication (MQL) technique in surface grinding by modeling and optimization", *Journal of Cleaner Production*, Vol. 86, pp. 447-460.
Rafaja, D., Poklad, A., Klemm, V., Schreiber, G., Heger, D. and Šíma, M. (2007), "Microstructure and hardness of nanocrystalline Ti1−x−yAlxSiyN thin films", *Materials Science and Engineering: A*, Vol. 462 Nos 1/2, pp. 279-282.
Rana, R., Walia, R.S. and Lata, S. (2018), "Development and investigation of hybrid electric discharge machining electrode process", *Materials Today: Proceedings*, Vol. 5 No. 2, pp. 3936-3942, doi: 10.1016/j.matpr.2017.11.650.
Rana, R., Kunal, R., Rohit, S. and Roop, L. (2014), "Optimization of tool wear: a review", *International Journal of Modern Engineering Research*, Vol. 4 No. 11, pp. 35-42.
Singh, R.C., Pandey, R.K., Roop, L., Ranganath, M.S. and Maji, S. (2016), "Tribological performance analysis of textured steel surfaces under lubricating conditions", *Surf. Topogr.: Metrol. Prop*, Vol. 4, p. 34005.
Singh, R.P. and Singhal, S. (2018), "Experimental study on rotary ultrasonic machining of alumina ceramic: microstructure analysis and multi-response optimization", *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, Vol. 232 No. 12, pp. 967-986.
Singh, R.P., Kataria, R., Kumar, J. and Verma, J. (2018), "Multi-response optimization of machining characteristics in ultrasonic machining of WC-Co composite through Taguchi method and grey-fuzzy logic", *AIMS Materials Science*, Vol. 5 No. 1, pp. 75-92.
Stachowiak, G.W. and Batchelor, A.W. (2005), "Elastohydrodynamic Lubrication", *Engineering Tribology*, 3rd ed., Butterworth-Heinemann, 287-362.
Tyagi, A., Walia, R.S. and Murtaza, Q. (2019), "Tribological behavior of temperature dependent environment friendly thermal CVD diamond coating", *Diamond and Related Materials*, Vol. 96, pp. 148-159.
Zhang, P. and Liu, Z. (2017), "On sustainable manufacturing of Cr-Ni alloy coatings by laser cladding and high-efficiency turning process chain and consequent corrosion resistance", *Journal of Cleaner Production*, Vol. 161, pp. 676-687.

Further reading

Kupczyk, M.J. (2015), "Cutting edges with high hardness made of nanocrystalline cemented carbides", *International Journal of Refractory Metals and Hard Materials*, Vol. 49, pp. 249-255.
Mia, M. and Dhar, N.R. (2017), "Optimization of surface roughness and cutting temperature in high-pressure coolant-assisted hard turning using Taguchi method", *The International Journal of Advanced Manufacturing Technology*, Vol. 88 Nos 1/4, pp. 739-753.
Rech, J., Giovenco, A., Courbon, C. and Cabanettes, F. (2018), "Toward a new tribological approach to predict cutting tool wear", *CIRP Annals*, Vol. 67 No. 1, pp. 65-68.
Rocha, L.C.S., de Paiva, A.P., Rotela Junior, P., Balestrassi, P. P. and da Silva Campos, P.H. (2017), "Robust multiple criteria decision making applied to optimization of AISI H13 hardened steel turning with PCBN wiper tool", *The International Journal of Advanced Manufacturing Technology*, Vol. 89 Nos 5/8, pp. 2251-2268.
Rana, R., Walia, R.S., Murtaza, Q. and Mohit, T. (2016), "Parametric optimization of hybrid electrode EDM process", *AES-ATEMA 29th Int. Conference, July 04-08, Toronto*, pp. 151-162.

Corresponding author

Qasim Murtaza can be contacted at: qasimmurtaza@dce.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**

2020 the 5th International Conference on Control and Robotics Engineering

HandAid: A Seven DoF Semi-Autonomous Robotic Manipulator

Vipul Garg
Department of Mechanical Engineering Delhi Technological University
Delhi, India
e-mail: garg.vipul7@gmail.com

Tript Sharma
Department of Mechanical Engineering Delhi Technological University
Delhi, India
e-mail: triptsharma22@gmail.com

Aakash Kumar
Department of Mechanical Engineering Delhi Technological University
Delhi, India
e-mail: aakash_bt2k16@dtu.ac.in

Vikas Rastogi
Department of Mechanical Engineering Delhi Technological University
Delhi, India
e-mail: rastogivikas@gmail.com

***Abstract*—For a person with only one arm, it is not very easy to grasp objects from their surroundings. Human beings have evolved to use both hands to carry out even ordinary tasks. Hence, specially-abled people with only one hand suffer difficulties in their day to day life. We present the development of HandAid, equipment designed to assist single-handed people. HandAid is a semi-autonomous device consisting of a redundant robotic manipulator attachment that can be attached to the shoulder and serves as a prosthetic arm to aid the user. The arm solely cannot identify the object to interact with; hence it is assisted by a headgear containing a stereo-camera and a laser pointer. The user can move his head and guide the laser to the concerned object, the located object is extracted using segmentation, and the point cloud of the object is created to obtain the collision-free path which is traced by the arm to grasp the object. This novel concurrent approach is implemented on ROS, integrated with OpenCV and moveit! for precise recognition and efficient manipulation of the object. Besides, the bot has been simulated on the gazebo to validate the results.**

Keywords—manipulator; point cloud; path tracking; computer vision

I. INTRODUCTION

Human-robot interaction is an active area in robotics research. Research in robotics followed in the development of two domains: Robotic manipulation and the input feeding system. In the present day, robots have become an assisting hand for human beings as a robot can perform many tasks that are difficult to complete for a human being in very little time with comparable accuracy. However, apart from the use of robots for industrial as well as military purposes, it is proposed that robotics arm can find application in a variety of fields. One such field is biomedical engineering as smart prosthetics for specially-abled people which would have the capability to operate semi-autonomously as per the instruction of humans. While the robotic systems keep progressing in terms of motor capacities owing to the advancement in mechatronics, developing control procedures and interfaces providing humans to harness the full potential of an advanced robotic arm.

It also proves to be a critical challenge in the field of humanoid robotics and precisely, rehabilitation engineering.

Even in the case of a typically abled human, the gap between robotic devices' complexity and command signals highlights the need for useful and practical control interfaces and maneuverings. To bridge this gap, researchers have reviewed techniques to retrieve additional input data from humans. One of those solutions is the sensor fusion strategy, which intends to consolidate measurements from multiple sensors running at once.

It is vital to instruct the arm accurately and precisely to achieve the action smoothly. The arms can be commanded by the mechanical action, voice commands or brain directly. Electroencephalography (EEG) headset equipped with multiple sensors had been used in [1] and [2] to interpret the brain commands and actuate the arm using feedback control. Electrical signals of muscles had been measured in [3] with the help of deep learning and electromyography (EMG) in response to neural stimulations. Similarly, voice command based control mechanism has been discussed in [4]. They have utilized a ZigBee transmitter and receiver to increase the range to 50 meters and successfully implemented various actions such as PICK and PLACE. The domain of the voice and gesture commanded manipulators is not limited to the surface only and hence, they also have been implemented for underwater operations [5]. The robust coordination between the manipulator and stereo camera is discussed in [6] and grasping of objects by compensating the hardware errors using fuzzy control has been presented in [7].

In this paper, we used a head-gear equipped with a laser pointer and stereo camera to point the concerned objects and localize them. The proposed approach

978-1-7281-6791-6/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 26,2020 at 04:57:57 UTC from IEEE Xplore. Restrictions apply.

proposes a general 7 seven DoF design with adjustable link length. Furthermore, the approach provides a robust laser detection technique with lateral invariance using computer vision that facilitates the semi-autonomous nature of the arm. Thereby, ensuring the grasping of the objects without any ambiguous voice commands which may arise when objects with similar features are in front.

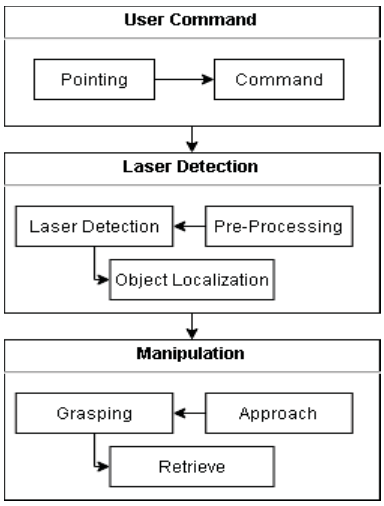


Figure 1. Flow chart of the methodology

The paper is distributed into three segments to discuss the proposed approach in which section II illustrates the methodology in detail. It outlines the proposed design, machine vision technology, and manipulation algorithm. The algorithm has been implemented in the gazebo simulator for its validation and discussed in section III. The article is concluded in section IV and mentions the challenges and future work.

II. METHODOLOGY

In this section, we will discuss the end-to-end pipeline from user input to grasping of an object, as shown in Fig. 1. The semi-autonomous arm requires a laser pointer position in the three-dimension space as input. To facilitate user command to grasp an object, the user has to confirm the object to be picked up. Following the command, the laser is detected and localized using the RGBD camera. The coordinates are transmitted to the manipulator to approach, grasp and retrieve the object.

We further explain the design of the arm and the complete grasping mechanism in the following subsections.

A. Mechanical Design

1) Arm Design: The robot arm is designed based on a robot manipulator with similar functions to a human arm. It consists of links, joints, actuators, sensors, and controllers. The links of the manipulator are connected through joints forming revolute pairs, and the links of the manipulator are arranged to form an open kinematic chain. One end of the chain is attached to the robot base, and another end is equipped with a gripping tool that is analogous to the human hand to perform grasping tasks and to interact with the environment. A brief review of the various design methodologies, mechanism, and control systems had been provided in [8] [9].A robotic arm with seven degrees of freedom is designed because it is adequate for most of the necessary movement. At the same time, it is competitive by its complexity and cost-saving as the number of actuators in the robotic arm increases with degrees of freedom. In a robotic system, the number of degrees of freedom is determined by the number of independent joint variables. The robotic arm has seven DC servo motors, each motor for one DOF and the motors operate with 12 volts.

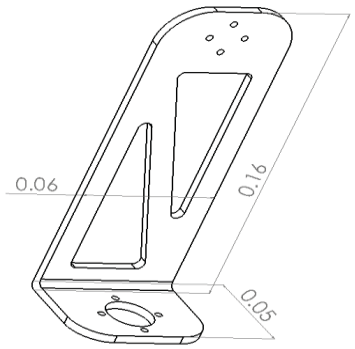


Figure 2. Drawing sheet of the primary link

The area that the gripper or the end-effector can reach is known as robot workspace. It depends on the DOF and translation limitation, the arm link lengths, the angle at which something must be picked up at and robot configuration. To adapt the modular approach in robotic arm design is essential because it aids in finding a replacement of robot parts or components during maintenance of the piece. Besides that, good design helps in providing a comprehensible design concept. As a result, machines and devices are more comfortable to assemble and manage.

Apart from the component designing selection of a suitable material for fabrication is also a vital task. The arm designed here is used for low weight applications. Hence, an Aluminium alloy is used to give adequate strength to the links and fulfil the lightweight application of the arm.

2) Calculations: To design and structure of a machine, the first thing one must apply is the principles of statics to write the force acting on its members. Selection of a suitable motor for the application requires a very detailed calculation to avoid sizing or under-sizing of the motor. Thus, load torque and speed which are converted to those at the motor output shaft have to be calculated. Positioning accuracy, holding of position, speed range and the other environmental resistance need to be checked afterwards. Let the weight of the object to be lifted, and the end parts on the last link are 0.25 kg.

Thus, we can say, the torque required at the nearest joint will be

$$T = r \times F = L(m) \times w\,(N) \quad (1)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 26,2020 at 04:57:57 UTC from IEEE Xplore. Restrictions apply.

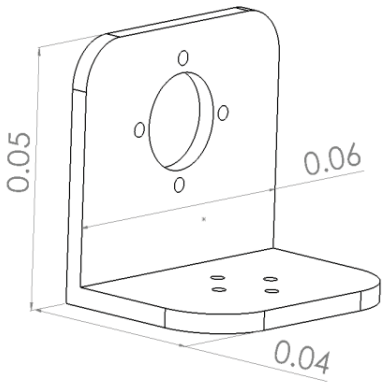


Figure 3. Drawing sheet of the intermediate link

Similarly, the maximum torque on the motor joined at the base link,

$$T = \Sigma[Li(m) \times Wi(N)] \quad (2)$$

Torque on the base link is calculated

Let T_L, T_m, T_I be the net torque on the base joint due to three primary links, five Motors and two intermediate links respectively. Let W_L, W_m, and W_I be the weight of the primary link, Motors and intermediate link respectively To calculate the torque we deduced the center of mass of the links C.O.M. of the primary link (refer Fig. 2) is calculated to be 0.062m from the bend, and C.O.M. of the intermediary link (refer Fig. 3) is calculated to be 0.015m from the bend The torque equation is written for each of the three proceedings from the joint about which the torque is calculated Torque due to primary links

$$T_L = W_L(0.10 + 0.2825 + 0.465) \quad (3)$$

Torque due to motor on the base joint

$$T_m = W_m(0.162 + 0.1825 + 0.3445 + 0.365 + 0.527) \quad (4)$$

Torque due to intermediate joint links

$$T_I = W_I(0.1825 + 0.365) \quad (5)$$

$$T = W_L(0.8475) + W_m(1.581) + W_I(0.5475) + 0.25 * g * (0.90) \quad (6)$$

Material Density (ρ) Aluminium 3000 series alloy is 2750 Kg/m^3. Therefore,

$$W_L = \rho(0.2071 * 0.06 * 0.003) * g = 1.006N$$

$$W_I = \rho(0.09 * 0.06 * 0.003) * g = 0.437N$$

$$W_m = 0.17 * g = 1.6677N$$

$$T = 1.006(0.8475) + 1.6677(1.581) + 0.437(0.5475) + 2.4525(0.90)$$

$$T = 5.94Nm$$

Thus, the torque on the joints can be calculated to decide for the motors to be required.

B. Object/Laser Detection

HandAid utilizes a laser pointer to direct the arm to the desired object. This synchronization between the laser pointer and arm is facilitated using a stereo-camera with 100Hz. The camera detects the laser pointer in the surroundings and transmits its coordinates in the three-dimensional space to the manipulator.

The proposed laser detection methodology requires the user to confirm the object to be picked up. Following the command, two images are taken consecutively, one without the laser pointer and other with the laser being pointed at the object. The former serves as the reference image, while the latter is the laser image. Since the images are taken with a 0.01 second time difference, it is assumed that the objects do not undergo rotation or scalability changes. The reference image is used to identify the salient features after segmentation, as shown in Fig. 4. Utilizing the color information of the segmented objects, contours are identified. These contours are filtered based on their area and convexity. Next, the difference between the respective contours of the two images is used to identify the laser pointer. This is because all contours in the two images are the same except the one with laser because of the split-second difference leading to the identification of the laser pointer. Practically, the images are pre-processed to remove the noise using morphological operations and median filtering with a kernel size of 5 units. The smoothed reference image is segmented using K-Means clustering [10]. Using the cluster centers' color information, salient features in the reference image are dynamically identified. These features are matched against the original image using the Normalized Cross-Correlation Matching Method for template matching [11]. Circular shapes are identified in the differences between the template and detected images and filtered on the basis of area and convexity. Thereby resulting in the identification of the laser pointer in the image.

C. Manipulation

The relative position of the laser pointed object in the three-dimensional space is used to align the arm gripper with respect to it to perform grasping. The coordinates of the extracted object are described in the camera's optical frame and transformed into the manipulator's base reference frame using (7).

$$P_m = RP_c \quad (7)$$

Where R is the rotation matrix, P_c and P_m are the positions of laser pointed object with respect to camera optical and manipulator's base reference frame respectively. The object to be grasped is checked for its feasibility by passing it to feasibility filters.

- *Position Filter*: It estimates the work envelope of the arm dynamically and checks if the objects lie inside it or not. If the object is out of the range, the user is alerted to move towards the object until the work envelope encloses it.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 26,2020 at 04:57:57 UTC from IEEE Xplore. Restrictions apply.

(a) Original Image (with green laser at bottom center)

(b)Image Segmentation

(c) Contour Detection on Reference Image

(d) Laser Pointer Identification after Template Matching

(e)Laser Pointer Localization

Figure 4. Laser detection

- *Size Filter*: It creates a rectangular contour around the object and uses the point cloud to determine the height and width of the object. If both of the object dimensions are greater than the grasp width, then the object cannot be grasped.

Once the object is checked to be feasible, a motion planning algorithm is implemented to perform the function. Filters based collision avoidance algorithm [12] has been used to avoid all the collisions between the arm and the nearby objects. The algorithm eliminates all those arm configurations, which may lead to the collision. The remaining configurations are utilized to generate a path to reach the object, grasp it and retrieve it. This approach is divided into three major actions which are discussed below.

1) Approach: The gripper is moved freely towards the object until it covers 60% of the path and aligns itself according to the concerned object. Then the gripper is moved slowly with the increment of 1cm and covers 50% of the remaining path and leaves the object untouched.

2) Grasping: In this process, the object is again detected for its final location before grasping to include the disturbances and deviations and reflect the changes in the final positions of the gripper. The gripper opens its fingers to the maximum and approaches the object.

Once, the object lies inside the grasping region of the gripper; the gripper closes its fingers until the object is grasped.

3) Retrieve: The gripper retraces the path it went through and moves the object to the desired location, which can be any surface or suspend it in the air until it receives another command.

III. Implementation and Results

The methodology discussed above has been tested in the gazebo to simulate the grasping of a can by pointing a laser at it through a headgear. The camera and the arm are mounted on the headgear so that they can mimic the real person's head and hand. After receiving a grasping command, the images are captured and processed for the concerned object and extract it from the background. The coordinates of the object are the goal state for the gripper mounted on the arm, and hence, a motion plan has been generated to perform the actions. The manipulator initiates the approach action and moves from its initial state (refer Fig. 5a). Fig. 5b shows that the gripper has aligned itself in the middle of the approach and headed towards the can. At the end of approach action (refer Fig. 5c), It initiates the grasping action by recapturing the image and checking for the required

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 26,2020 at 04:57:57 UTC from IEEE Xplore. Restrictions apply.

deviations. The grasping action is completed with the closing of the fingers (refer Fig. 5d) and retrieving the object and moving it away from the table (refer Fig. 5e).

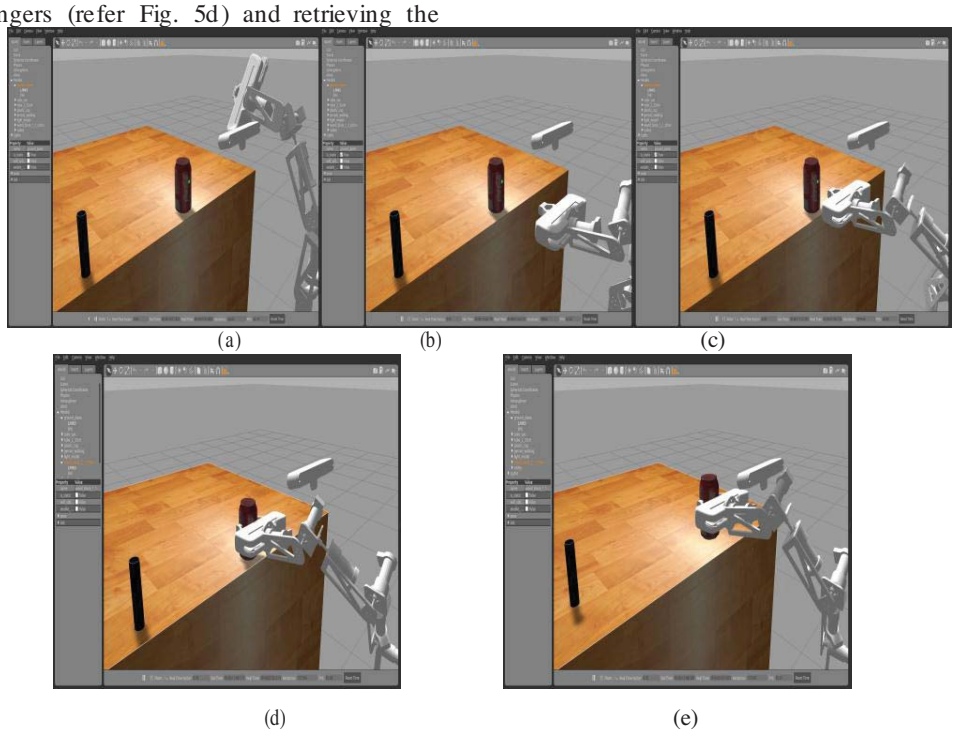


Figure 5. HandAid manipulator under simulation

IV. Conclusion

In this article, we proposed a semi-autonomous robotic arm to grasp and move objects. The arm design enables a seven DoF motion to provide better and robust path planning. The proposed methodology employs a laser and a stereo- camera to guide the arm using computer vision techniques to dynamically locate the laser pointer in the environment. Thus, providing real-time grasping which can aid single-handed people in day-to-day life tasks. They are more precise and consistent than human hands. The robot arm aids the user in multitasking by employing both their human and robotic arm simultaneously.

The research also raises some challenges in the domain. HandAid can detect and grasp static objects, but the grasping of dynamic objects in real-time still needs to be answered. Furthermore, the robot manipulator is designed to lift objects within a small range of weight without real-time adjustment of motor torque as a function of object weight. Future works should acknowledge the effect of object weight on grasping applications.

References

[1] T. Beyrouthy, S. Al Kork, J. Korbane, and A. Abdulmonem, "Eeg mind controlled smart prosthetic arm," 08 2016, pp. 404–409.

[2] D. Salvekar, A. Nair, D. Bright, and S. Bhisikar, "Mind controlled robotic arm," *IOSR Journal of Electronics and Communication Engi- neering (IOSR-JECE)*, pp. 36–44, 2015.

[3] M. Jafarzadeh, D. C. Hussey, and Y. Tadesse, "Deep learning approach to control of prosthetic hands with electromyography signals," in *2019 IEEE International Symposium on Measurement and Control in Robotics (ISMCR)*. IEEE, 2019, pp. A1–4.

[4] R. Aswinbalaji and A. Arunraja, "Wireless voice controlled robotics arm," *International Journal of Emerging Technology in Computer Sci- ence & Electronics (IJETCSE)*, vol. 12, no. 4, pp. 1–6, 2015.

[5] A. K. Mishra and O. Meruvia-Pastor, "Robot arm manipulation using depth-sensing cameras and inverse kinematics," in *2014 Oceans-St. John's*. IEEE, 2014, pp. 1–6.

[6] G. D. Hager, W.-C. Chang, and A. S. Morse, "Robot hand-eye coordina- tion based on stereo vision," *IEEE Control Systems Magazine*, vol. 15, no. 1, pp. 30–39, 1995.

[7] J.-W. Chang, R.-J. Wang, W.-J. Wang, and C.-H. Huang, "Implementa- tion of an object-grasping robot arm using stereo vision measurement and fuzzy control," *International Journal of Fuzzy Systems*, vol. 17, no. 2, pp. 193–205, 2015.

[8] T. Kulkarni and R. Uddanwadiker, "Overview: Mechanism and control of a prosthetic arm," *Molecular cellular biomechanics: MCB*, vol. 12, 01 2016.

[9] J. J. Craig, *Introduction to robotics: mechanics and control, 3/E*. Pearson Education India, 2009.

[10] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algo- rithm," *Procedia Computer Science*, vol. 54, pp. 764–771, 2015.

[11] A. Kaehler and G. Bradski, *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. " O'Reilly Media, Inc.", 2016.

[12] V. Garg and R. Vikas, "Path tracing and object avoidance algorithm for robotic manipulators incorporating constrained filters," 12 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 26,2020 at 04:57:57 UTC from IEEE Xplore. Restrictions apply.

Hierarchical Clustering of World Cuisines

1st Tript Sharma
Department of Mechanical Engineering
Delhi Technological University
New Delhi, India
triptsharma22@gmail.com

1st Utkarsh Upadhyay
Department of Electronics Engineering
Jamia Millia Islamia University
New Delhi, India
utkarshdhy@gmail.com

2nd Jushaan Kalra
Department of Computer Engineering
Delhi Technological University
New Delhi, India
jushaan18@gmail.com

3rd Sakshi Arora
Department of Computer Science
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
sakshi18133@iiitd.ac.in

4th Saad Ahmad
Department of Computational Biology
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
saad18409@iiitd.ac.in

4th Bhavay Aggarwal
Department of Computational Biology
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
bhavay18384@iiitd.ac.in

Ganesh Bagler
Center for Computational Biology
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
bagler@iiitd.ac.in

arXiv:2004.12283v1 [cs.SI] 26 Apr 2020

***Abstract*—Cultures across the world have evolved to have unique patterns despite shared ingredients and cooking techniques. Using data obtained from RecipeDB, an online resource for recipes, we extract patterns in 26 world cuisines and further probe for their inter-relatedness. By application of frequent itemset mining and ingredient authenticity we characterize the quintessential patterns in the cuisines and build a hierarchical tree of the world cuisines. This tree provides interesting insights into the evolution of cuisines and their geographical as well as historical relatedness.**

***Index Terms*—Hierarchical Clustering, Pattern Mining, Authenticity Correlation, Kmeans Clustering, Food Ontology**

I. Introduction

Cultures across the world have evolved diverse cooking practices over time. Although the underlying fundamentals of cooking remain the same, various factors including geography and climate have affected cooking styles. Cuisines from across the globe have thus acquired their signature styles. Each cuisine has interesting patterns that are inherent to it while sharing some common attributes with others. In this article, we characterize the unique features that typify every cuisine in an attempt to discern the footprint of food on human cultures and inter-relatedness of world cuisines.

II. Literature Survey

With increasing availability of recipes data, there has been much interest in data mining recipes data. One of the focus has been on defining recipe similarity. Attempts have been made to define similarity based on various elements of cooking recipes [15] and ingredients [7], [14].

Among other efforts in data mining recipes data have focused on food pairing phenomena in cuisines. Among one of the early studies, Shidochi et al [11] experimented with the possible replacements of ingredients in a recipe. Jain et al [8] investigated the phenomenon of food pairing which examines compatibility of two ingredients in a recipe in terms of their shared flavor compounds. This study investigated the food pairing phenomena in Indian recipes and proclaimed that spices form the basis of their food pairing. The work was extended by Singh et al [12] to analyze a much larger dataset encompassing 22 cuisines across the world and found interesting food pairing patterns in cuisines from across the world. Tuwani et al [13], on the other hand, considered culinary systems as a function of socio-cultural factors and presented computational models for cuisine evolution. An interesting work by Yokoi et al [16] calculated an ingredient associative metric called 'typicality value', giving out typical recurring ingredient patterns.

In this article, we indulge in pattern analysis in world-wide cuisines by way of association rule discovery and frequent pattern mining [1]. Going beyond the application of pattern mining techniques on cuisines in [10], we propose their use for frequent pattern mining of recipe data and cooking processes, utensils and ingredients for hierarchical clustering of cuisines.

III. Data Collection

Our analysis involved four types of information pertaining to traditional recipes, namely, recipes, ingredients, processes and utensils. A total of 118,071 recipes were obtained from various sources: AllRecipes, Food Network, Epicurious and TarlaDalal. RecipeDB [3], a structured compilation of recipes, was used as the primary source of information. All data are available at 'RecipeDB: A resource for exploring recipes'.

For each recipe, details such as its name and the list of ingredients and processes involved while cooking are

TABLE I
SIGNIFICANT PATTERNS MINED FROM CUISINES ACROSS THE WORLD

Region	Number of Recipes	Pattern	Support	Number of patterns
Australian	5,823	Butter	0.24	29
Belgian	1,060	Butter + salt	0.24	51
Canadian	6,700	Onion	0.20	31
Caribbean	3,026	Garlic Clove	0.24	32
Central American	460	Onion	0.30	38
Chinese and Mongolian	5,896	Soy sauce + add + heat	0.27	88
Deutschland	4,323	Onion	0.29	54
Eastern European	2,503	Cream	0.30	60
French	6,381	skillet	0.21	60
Greek	4,185	Olive Oil	0.40	43
Indian Subcontinent	6,464	Onion + add + heat + salt	0.22	119
Irish	2,532	Butter	0.32	41
Italian	16,582	Parmesan cheese	0.31	63
Japanese	2,041	Soy Sauce	0.45	45
Mexican	14,463	cilantro	0.25	33
Rest Africa	2,740	Onion + add + heat	0.20	51
South American	7,176	Onion + salt	0.21	62
Southeast Asian	1,940	Fish sauce	0.24	69
Spanish and Portuguese	2,844	Olive Oil	0.31	67
Thai	2,605	Fish sauce + add + heat	0.23	73
Korean	668	Soy sauce + sesame oil green onion + sesame oil	0.34 0.24	85
Middle Eastern	3,905	Salt + bowl Lemon Juice	0.22 0.22	46
Northern Africa	1,611	cumin + cinnamon cumin + olive oil cumin + Salt	0.21 0.22 0.22	134
Scandinavian	2,811	Butter + Salt Salt + Sugar	0.22 0.21	52
UK	4,401	Butter Oven	0.37 0.46	45
US	5,031	Bake + preheat+ oven + bowl Onion	0.22 0.25	67

available. Each recipe was treated as an unordered list of ingredients, processes and utensils. We integrated recipes from all the sources and grouped them into 26 distinct geo-cultural 'cuisines' while ensuring that each region had enough recipes attributed to it to distinguish it as a cuisine. Please refer to Table I for the list of regions. Due to insufficient information about the region for many recipes, they were aggregated together on the basis of their geological similarities with the prefix 'Rest'. For example, recipes without 'region' information belonging to Africa were put in 'Rest Africa' category.

The database consists of 20,280 unique ingredients, 268 unique processes and 69 unique utensils. The data are sparse in the list of utensils and 14,601 recipes don't have information regarding the preferred utensils required for cooking. An average recipe in a cuisine has $\sim$10 ingredients, $\sim$12 processes and $\sim$3 utensils. This is intuitive as too many ingredients would impede the success/propagation of a recipe, whereas too few would lead to it being modified easily. Thus recipes needs to maintain a balance between complexity and simplicity to survive successive iterations of evolution [13].

IV. PATTERN MINING

To investigate the ontology of food, we mined rules from the data to understand the patterns that are prevalent in a given cuisine. The methodology employed for mining patterns is explained in Section V. The mined patterns consist of ingredients, processes and utensils permutations that have a frequency greater than the defined threshold support. According to [1], [6], support represents the frequency with which the collection of items co-occur as a percentage of all transactions. A high support threshold represents high confidence in the pattern being mined, whereas with a low support threshold noise can creep into the mined patterns, leading to false identification of cuisine features. Hence, a trade off support of 20% was chosen as the threshold. Since the mined patterns are the most frequent ones, it is safe to say that most of the recipes follow the observed patterns and essentially define the cooking practices of the cuisine.

Among the patterns obtained in all the recipes for the Korean region as shown in Table I, the pattern "Soy Sauce + sesame oil" occurs with a support of 0.34, i.e. the pattern is found in 34% of all the recipes in the Korean region. Table I contains the topmost significant patterns in the 26 cuisines.

The pattern depicts set of words occurring in a particular recipe. The patterns mined are highly skewed, with most regions containing patterns having generic ingredients such as 'salt', 'onion' and processes such as 'add' and 'cook', which is justified as they have a high frequency among all cuisines and are fundamental to cooking in many cuisines.

V. Methodology

We implemented two approaches namely Frequent Itemset-based Hierarchical Clustering (FIHC) [5] and Authenticity based Clustering [2] to extract relationships between various cuisines of the world. The hypothesis is that some patterns which are common across a subset of cuisines would be found, which defines their 'similarity/closeness' with each other.

A. Frequent Pattern Mining

Frequent Itemset Mining refers to discovering interesting patterns in databases such as association rules from a set. Since we treat a recipe as a combination of ingredients, processes and utensils, it can be treated as an unordered set of these entities. For the frequent itemset mining, the FP-Growth Algorithm [6] was used as it is an efficient and scalable method for mining the complete set of frequent patterns by pattern fragment growth. The data extracted from RecipeDB was pre-processed to make it compatible with the input form of FP-Growth Algorithm. Ingredients, utensils and processes were concatenated and the FP-Growth Algorithm was applied. This approach was applied to all 26 regions present in the data extracted from RecipeDB. The support was kept at 0.2 so that the pattern was mined across a reasonable number of recipes.

B. Authenticity Based Clustering

We propose that a cuisine can be represented as a set of ingredients, process and utensils which can thus be utilized to define the relationships among the cuisines. Using the authenticity metric described in [2] we calculate the prevalence P_i^c of an item i in a cuisine c according to equations 1 as a function of number of recipes, n_i^c in a cuisine over total number of recipes in the dataset, N_C. This is used to calculate the authenticity of the item for a cuisine using equation 2.

$$P_i^c = n_i^c / N_C \tag{1}$$

$$p_i^c = P_i^c - (P_i^k)_{c \neq k} \tag{2}$$

In order to obtain the contribution of the item in uniquely identifying a cuisine, a relative prevalence matrix is created by subtracting the average prevalence of the item, say i for all cuisines from the prevalence for cuisine c. Accordingly, the most prevalent and least prevalent items in a cuisine can be identified. It should be noted that both the most prevalent and least prevalent items would contribute towards the culinary fingerprint of a cuisine as the former indicates the items having a relatively higher utility in the cuisine while the latter indicates items that are least used in the cuisine versus the rest of the world cuisines.

VI. Clustering Techniques

A. Hierarchical Clustering

Application of FP-Growth Algorithm on the prepared dataset results in 26 files, each containing patterns in a 'frozenset' along with their respective support to remove redundant patterns. These patterns were extracted from the 'frozenset' and appended together in a list in a sorted fashion. All the elements of this list are appended and converted into a string resulting in a 'string pattern'. All the 'string patterns' are compiled into a set resulting in unique set of patterns across all the 26 regions. Since the data is in string form and each element is a unique entity, it can be classified as a category. Therefore, unique set of 'string patterns' are fit for using Label Encoding (because the strings are categorical data) to get a transformer and the 'string patterns' in the rules are transformed using the derived transformer across all the regions. All the 'string patterns' in the rules from all the regions are appended in an array. The data from this array is thus vectorized to form a feature vector which is thus fed to the cluster as the linkage matrix.

Three different approaches were applied in order to cluster the linkage matrix data and to generate subsequent dendrograms. The linkage matrix is converted into a condensed distance matrix (pdist) in order to calculate the distance between all the cuisines based on the rules mined and is then fed into the hierarchical clustering model. To analyze the clusters we have used three distance metrics:

$$\text{Jaccard Distance} = \frac{c_i \cup c_j}{c_i \cap c_j} \tag{3}$$

$$\text{Cosine Distance} = \frac{c_i . c_j}{|c_i||c_j|} \tag{4}$$

$$\text{Euclidean Distance} = \sqrt{c_i^2 + c_j^2} \tag{5}$$

where cuisines c_i, $c_j \in C$, the universal set of cuisines in the dataset. To calculate the distance between two cuisines, they must be quantified. This was done by vectorizing the patterns obtained by the above-mentioned pre-processing technique.

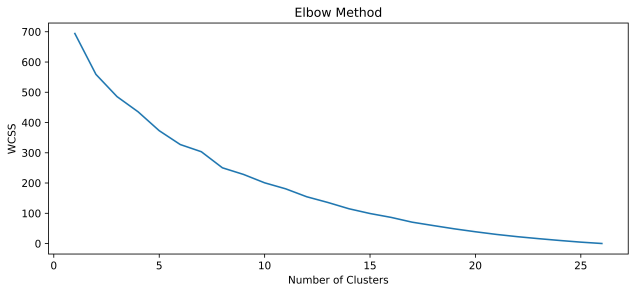


Fig. 1. Elbow Method for cluster identification

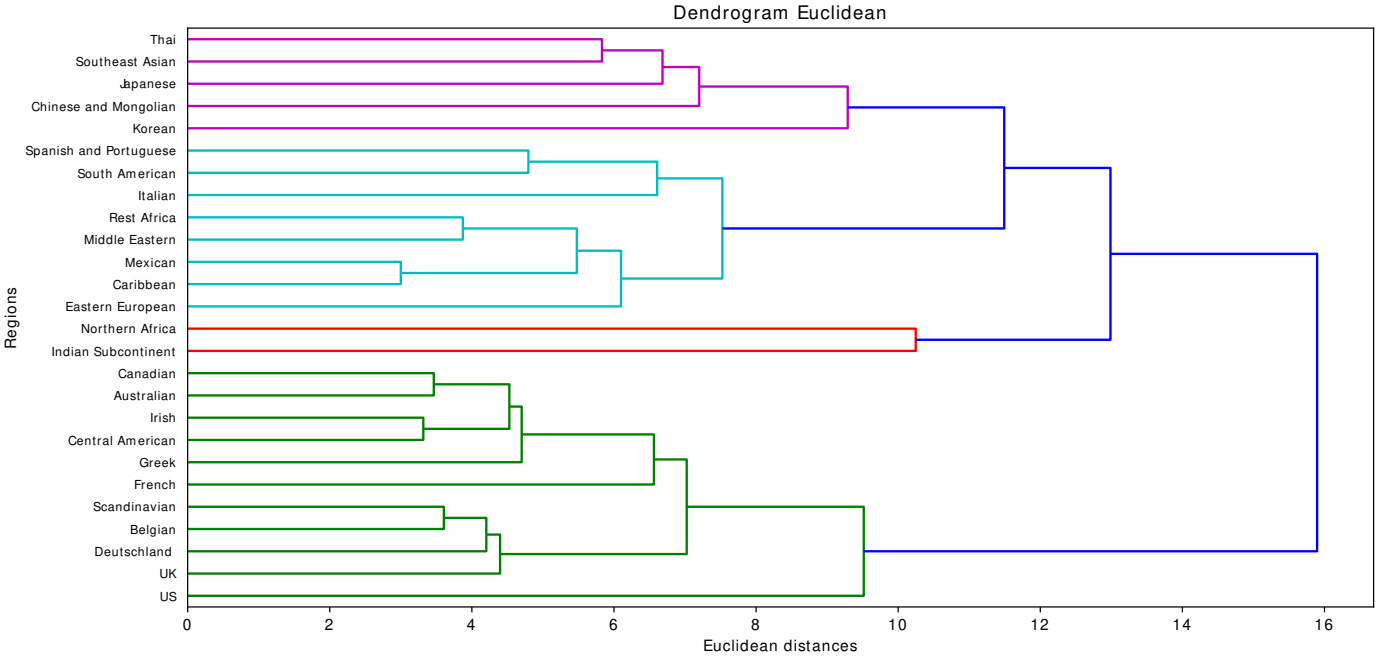


Fig. 2. Hierarchical Agglomerative Clustering based on Patterns Mined using Euclidean distance

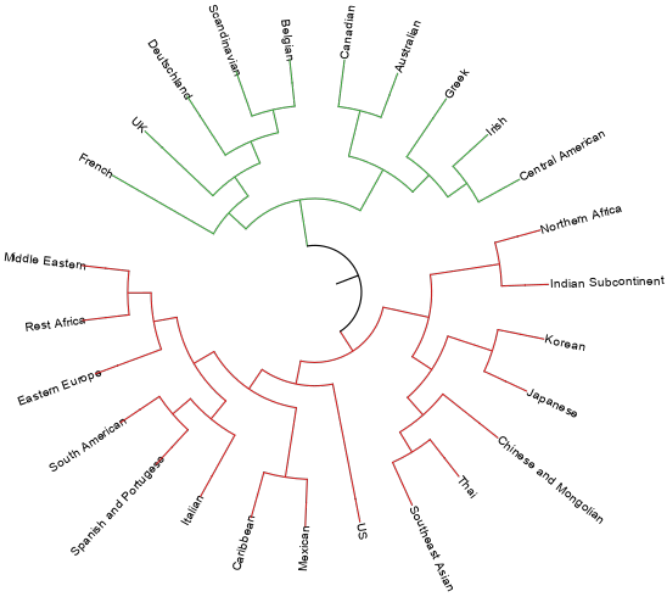


Fig. 3. Hierarchical Agglomerative Clustering based on Patterns Mined using Cosine distance

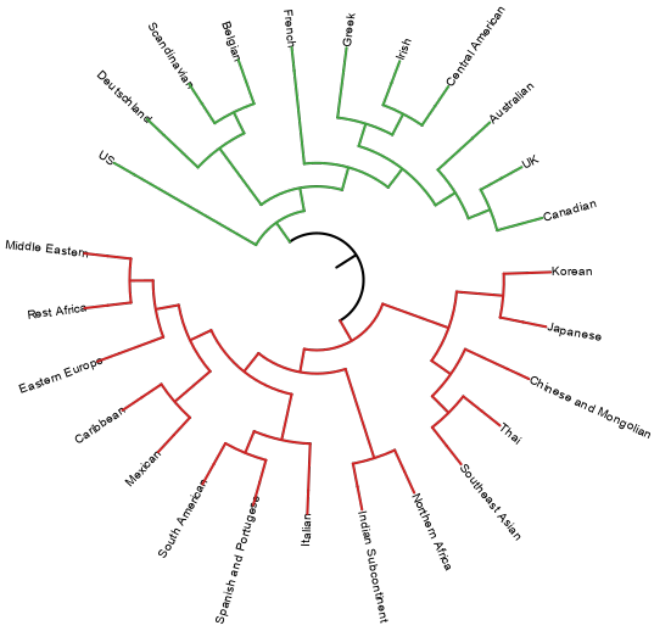


Fig. 4. Hierarchical Agglomerative Clustering based on Patterns Mined using Jaccard distance

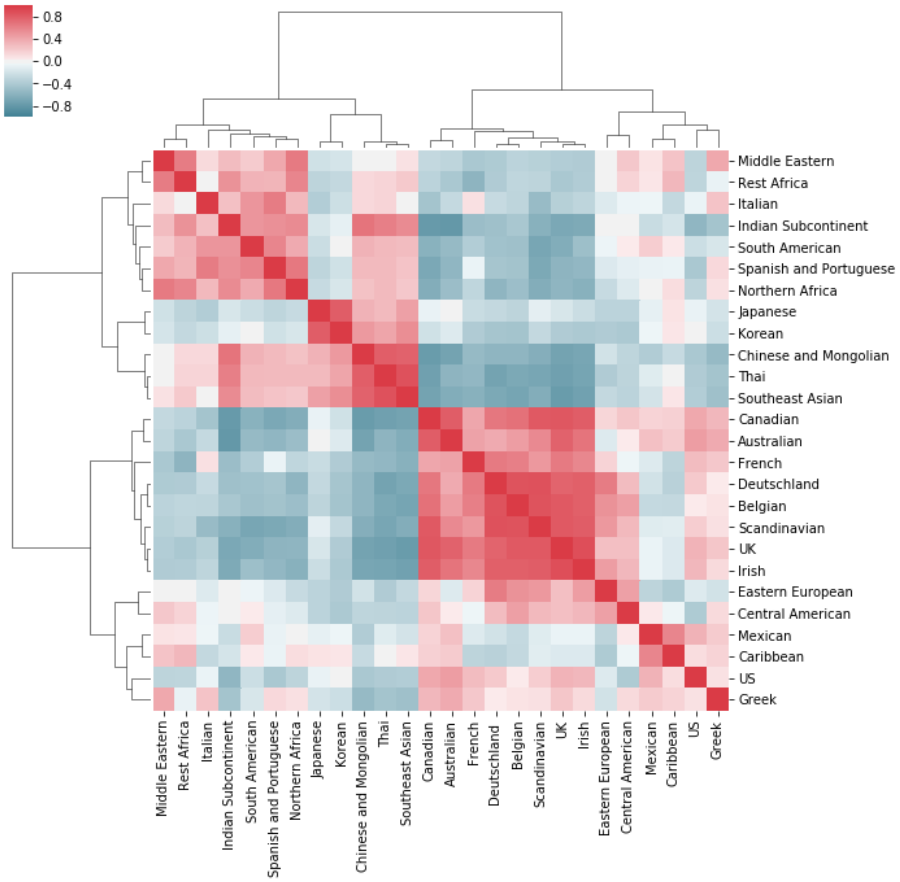


Fig. 5. Hierarchical Agglomerative Clustering based on Authenticity of Ingredients

B. *K-means Clustering*

Another popular clustering technique, K-means, was applied on our categorical data. It has been shown in [9] that hierarchical agglomerative clustering is a better approach for clustering categorical data than K-means. The elbow analysis and the subsequent WCSS score on our dataset indicates similar results. The elbow method [4] analysis fails to determine the number of appropriate clusters for our dataset. As in Figure 1, no sharp edge or elbow like structure is obtained which determines the number of clusters. While on the other hand the hierarchical agglomerative clustering technique presents with better cluster representation. Therefore, our results were predominantly determined by hierarchical agglomerative clustering.

VII. Results

To evaluate the efficiency of the proposed methodologies, the RecipeDB dataset mentioned in Section III was used. This dataset was used to identify the patterns which were then fed into the Sequential Pattern Mining based clustering algorithm while the ingredients obtained from the dataset were the input features for the Authenticity-based clustering. The corresponding code and relevant files are present in the GitHub repository (https://github.com/cosylabiiit/Hierarchical-Clustering-Ingredients).

The Hierarchical Agglomerative clustering (HAC) gives clusters for all regions based on the three approaches and presents a cluster dendrogram for each approach. Figures 2, 3 and 4 represent the clusters formed using the feature vector obtained via Euclidean, Cosine and Jaccard metrics for pairwise distance calculation respectively. Similarly, Figure 5 shows the authenticity based approach to determine the correlation of cuisine and regions, dominantly based on ingredients.

Because of the absence of a quantified validation metric for cuisine similarity, the geographical relationship among the cuisines was used to validate the accuracy in the prediction of cuisine interrelationships. It is observed that while comparing the Figures 2, 3 and 4 with Figure 6 the results received from the Euclidean distance model were most similar to the geographical distribution of the countries. On the other hand, the clusters obtained via the authenticity based clustering gave similar yet better results than Euclidean distance-based HAC when validated on geographical distance based clusters.

Authenticity-based Clustering identifies both positive and negative relationships between cuisines and items whereas pattern based techniques take only the positive relationships

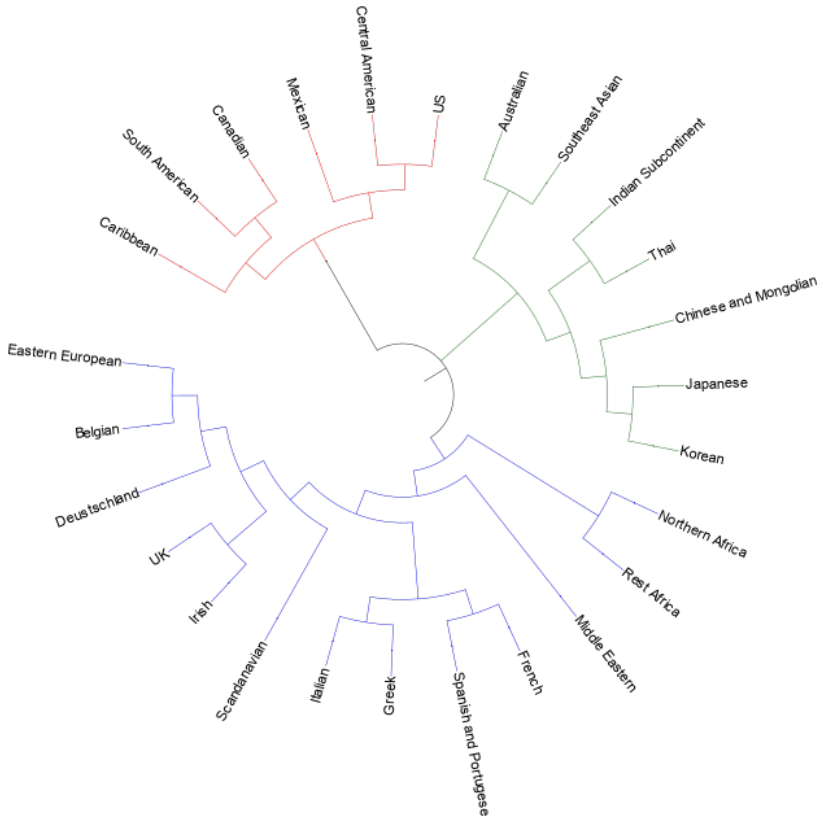


Fig. 6. Hierarchical Agglomerative Clustering based on Geographical Distance of Regions

into account. This leads to the difference in the results. Yet, despite the differences, both techniques predict a closer relationship among Canadian and French cuisines as compared to Canadian and US cuisines despite their geographical proximity. This is evident from the historical fact that Canada was a French colony. Another interesting grouping is that of Indian Subcontinent and Northern Africa. Due to prevalent use of spices in the two regions, Indian subcontinent cuisine is closer to African cuisine as compared to its geographical neighbors like Thai and Southeast Asian cuisines. Hence, the obtained clusters are also able to identify relationships deviating from the geographical similarities.

VIII. Conclusions and Future Work

In this exploratory work we proposed and analyzed two methodologies for fingerprinting cuisines and identifying their interdependence. Our clustering algorithms show how various cuisines are interrelated and show trends similar to their geographical associations. It shows how cooking practices and methods are shared by neighbouring regions. This analysis is important from a historical and cultural point of view as it helps in appreciating how cooking practices are distributed across the world. Furthermore, we also provide a verbose list of patterns identified in the cuisines. These patterns include compound patterns; combination of ingredients, processes and utensils that can be used to identify the relationship among these items.

While this article introduces new methods for investigation of cuisine correlations, it raises new research questions. How do factors such as climate, economy and genetics influence the cuisine patterns? RecipeDB is a sparse dataset in terms of utensils and processes. Hence, to what extent do they influence the relationships among cuisines is yet to be answered. Among one of the limitations of this study, it neither considers the state of ingredients nor their aliases. Therefore, future analysis need to account for the aliases along with state of ingredients and other properties like cooking time and preparation time for the task. It would also be interesting to identify more sophisticated validation metric for cuisine ontology than geographical clustering.

We believe that this study can be applied for cuisine fingerprinting, food ontology and exploration of relations between food and culture. Probing the past and present interrelatedness among cuisines can provide insight into human behavior and

cultures, and means for shaping the future of food.

IX. Acknowledgement

G.B. thanks the Indraprastha Institute of Information Technology (IIIT-Delhi) for providing computational facilities and support. T.S, U.U, J.K,S.A, S.A. and B.A. are Research Interns in Dr. Bagler's lab (Complex Systems Laboratory) at the Center for Computational Biology. All the research interns are thankful to IIIT-Delhi for the support.

References

[1] R. Agrawal, R. Srikant *et al.*, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, 1994, pp. 487–499.

[2] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási, "Flavor network and the principles of food pairing," *Scientific reports*, vol. 1, p. 196, 2011.

[3] D. Batra, N. Diwan, U. Upadhyay, J. S. Kalra, T. Sharma, A. K. Sharma, D. Khanna, J. S. Marwah, S. Kalathil, N. Singh *et al.*, "Recipedb: A resource for exploring recipes," *Available at SSRN 3482237*, 2019.

[4] M. J. Brusco and J. D. Cradit, "A variable-selection heuristic for k-means clustering," *Psychometrika*, vol. 66, no. 2, pp. 249–270, 2001.

[5] B. C. Fung, K. Wang, and M. Ester, "Hierarchical document clustering using frequent itemsets," in *Proceedings of the 2003 SIAM international conference on data mining*. SIAM, 2003, pp. 59–70.

[6] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *ACM sigmod record*, vol. 29, no. 2. ACM, 2000, pp. 1–12.

[7] S. Hanai, H. Nanba, and A. Nadamoto, "Clustering for closely similar recipes to extract spam recipes in user-generated recipe sites," in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*. ACM, 2015, p. 31.

[8] A. Jain, N. Rakhi, and G. Bagler, "Analysis of food pairing in regional cuisines of india," *PloS one*, vol. 10, no. 10, p. e0139539, 2015.

[9] A. Joshi and R. Kaur, "A review: Comparative study of various clustering techniques in data mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 3, pp. 55–57, 2013.

[10] T. Ozaki, X. Gao, and M. Mizutani, "Extraction of characteristic sets of ingredients and cooking actions on cuisine type," in *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. IEEE, 2017, pp. 509–513.

[11] Y. Shidochi, T. Takahashi, I. Ide, and H. Murase, "Finding replaceable materials in cooking recipe texts considering characteristic cooking actions," in *Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities*, 2009, pp. 9–14.

[12] N. Singh and G. Bagler, "Data-driven investigations of culinary patterns in traditional recipes across the world," *arXiv preprint arXiv:1803.04343*, 2018.

[13] R. Tuwani, N. Sahoo, N. Singh, and G. Bagler, "Computational models for the evolution of world cuisines," *arXiv preprint arXiv:1904.10138*, 2019.

[14] Y. van Pinxteren, G. Geleijnse, and P. Kamsteeg, "Deriving a recipe similarity measure for recommending healthful meals," in *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 2011, pp. 105–114.

[15] L. Wang, Q. Li, N. Li, G. Dong, and Y. Yang, "Substructure similarity measurement in chinese recipes," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 979–988.

[16] S. Yokoi, K. Doman, T. Hirayama, I. Ide, D. Deguchi, and H. Murase, "Typicality analysis of the combination of ingredients in a cooking recipe for assisting the arrangement of ingredients," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2015, pp. 1–6.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Impact of Variability on Processor Performance in Negative Capacitance FinFET Technology

Hussam Amrouch, *Member, IEEE*, Girish Pahwa, *Member, IEEE*, Amol D. Gaidhane, *Graduate Student Member, IEEE*, Chetan K. Dabhi, *Graduate Student Member, IEEE*, Florian Klemme, Om Prakash, *Member, IEEE*, and Yogesh Singh Chauhan, *Senior Member, IEEE*

***Abstract*— In this work, we investigate for the first time the impact of Negative Capacitance FinFET (NC-FinFET) technology on the performance of processors under the effects of process variations for various operating voltages. The industry compact model of FinFETs (BSIM-CMG) is fully calibrated to reproduce Intel 14 nm FinFET data of high volume manufacturing process. A physics-based negative capacitance (NC) model is integrated and solved self-consistently within the BSIM-CMG model. This allows the creation of NC-FinFET standard cell libraries, while considering the effects of various variability sources both in the ferroelectric layer as well as in the underlying constituent FinFET device. The variability-aware NC-FinFET libraries, fully compatible with the existing standard design flow of circuits, are then employed to perform simulations using commercial statistical timing analysis tools in order to study the performance of a 14 nm processor. For comprehensive analysis and comparisons, our implementation is done for both NC-FinFET and conventional (baseline) FinFET for a wide range of voltages. Our results demonstrate that process variations have a larger impact on the processor's performance in NC-FinFET – when it operates at a lower voltage compared to the baseline FinFET that still operates at the nominal high voltage – due to the additional ferroelectric-induced variability. Results also reveal that neglecting process variations leads to overestimating the benefit that NC brings to the processor's frequency improvement because of the larger timing guardband that is needed to overcome variability in NC-FinFET.**

***Index Terms*— Negative capacitance, ferroelectric, NCFET, FinFET, process variations, reliability, processor performance, steep sub-threshold slope.**

I. Introduction

NEGATIVE Capacitance Field-Effect Transistor (NCFET) is one of the promising technologies for future ultra-low power devices. The incorporated ferroelectric (FE) layer inside the transistor gate dielectric stack (see Fig. 1) exhibits, under certain conditions, a Negative Capacitance (NC) effect that allows the transistor to suppress the fundamental limit of sub-threshold swing (SS) of 60 mV/decade at room temperature. The origin of such a fundamental limit of SS stems from the Boltzmann thermal limit, which refers to the Boltzmann distribution of charge carriers at the source of transistor [1]–[3]. Boltzmann tyranny, in turn, strongly restricts the scalability of transistor threshold voltage (V_T) and as a result the ability to scale down the operating voltage (V_{DD}) of transistor with every new generation [4]. This, in practice, what had led to the discontinuation of Dennard's scaling more than a decade ago and imposed constraints on the performance of processors to avoid unsustainable power densities and elevated on-chip temperatures during run-time.

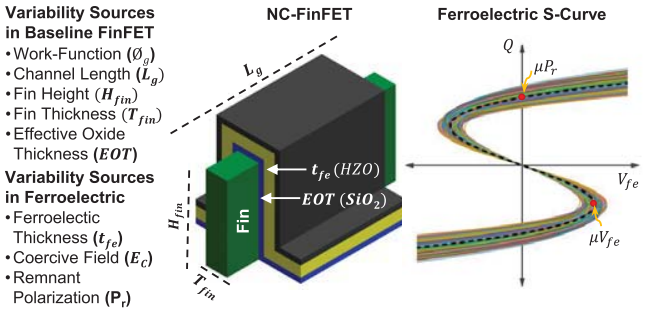


Fig. 1. Illustrating the 3-D structure of NC-FinFET and the resulting ferroelectric S-curve along with the different variability sources in the underlying baseline FinFET and the ferroelectric layer. Note that the structure of the baseline FinFET is similar to the structure of NC-FinFET with the only exception of not having a ferroelectric layer (HZO) in the gate stack.

Because NCFET can push the SS of transistor to go beyond its fundamental limit of 60mV/decade, NCFET is able to achieve the same ON-current (I_{ON}),[1] as in the conventional baseline (i.e., counterpart) FET device, but at a lower supply voltage. This makes NCFET-based circuits able to fulfill the same baseline performance, while still providing a considerable power saving [4]–[6]. The compatibility of NCFET technology, which became feasible after the discovery of ferroelectricity in HfO_2 based materials [7]–[9], with the existing CMOS process has been recently demonstrated in an industrial 14 nm FinFET technology [10]. In addition, [11]

Manuscript received December 29, 2019; revised March 23, 2020; accepted April 22, 2020. This work was supported in part by the U.S. Office of Naval Research (ONR) through KIT under Grant N00014-18-1-2672 and in part by the Swarnajayanti Fellowship Scheme of Department of Science and Technology through IITK. This article was recommended by Associate Editor A. Worapishet. *(Corresponding author: Hussam Amrouch.)*

Hussam Amrouch, Florian Klemme, and Om Prakash are with the Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany (e-mail: amrouch@kit.edu; klemme@kit.edu; prakash@kit.edu).

Girish Pahwa, Amol D. Gaidhane, Chetan K. Dabhi, and Yogesh Singh Chauhan are with IIT Kanpur (IITK), Kanpur 208016, India (e-mail: girish@iitk.ac.in; amold@iitk.ac.in; chetant@iitk.ac.in; chauhan@iitk.ac.in).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSI.2020.2990672

[1] ON-current of transistor is defined as the drain current of transistor in the saturation region, i.e. at $V_{GS} = V_{DS} = V_{DD}$.

1549-8328 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

has demonstrated that the intrinsic delay due to polarization damping in NCFET can reach $\approx$ 270 fs in doped HfO_2 ferroelectrics, which makes NCFET also suitable for digital circuits that typically operate in the GHz range like processors.

Several works have investigated the impact of NCFET on performance starting from logic gates and circuits [5] all the way to the processor level (both single-core [4], [12] and multi-core [6] processors). However, the full impact of NCFET on the performance of circuits and processors cannot be fully understood without considering the impact of process variations – this holds even more for advanced technology nodes like 14 nm and below, where the effects of metal work-function and geometrical variations become prominent. In addition to the conventional sources of variability, the incorporated ferroelectric layer within the gate stack of NC-FinFET itself results in additional sources of variability (see Fig. 1) that might seriously affect the performance of circuits, and hence, they must not be neglected. Otherwise, inaccurate estimations will be made. As depicted in Fig. 1, the variation in the ferroelectric parameters leads to the variation in its S shaped $Q - V_{fe}$ curve. This, in turn, affects the voltage amplification provided by the negative capacitance region of the ferroelectric in different NC-FinFET devices.

In this context of process variations, only a few work (e.g., [13]–[15]) aimed recently at studying how the figures-of-merit of transistors (e.g., I_{ON}, I_{OFF}, SS, V_T, etc.) can be affected by variability. However, the existing state of the art has two key limitations in which the analysis so far: 1) is merely performed at the device level, i.e. the overall impact of process variations on large circuits is still unexplored and 2) is merely performed at the nominal V_{DD} (as defined by the baseline FET technology) – even though the key idea and motivation behind the NCFET technology is to operate circuits at lower V_{DD} than the nominal voltage towards achieving power savings, while still maintaining the same baseline performance. Because the sensitivity of logic gates' and circuits' delay to process variations increases at lower voltages, it is, in fact, indispensable to compare the impact of process variations in NCFET with the corresponding impact on the baseline conventional FET at various voltages and not merely at the nominal voltage. Otherwise, the full impact of process variations on the performance of NCFET-based processors cannot be revealed, and the potential power savings, offered by NC, might be overestimated.

A. Key Focus of This Work

This work investigates the impact of process variations (both in the ferroelectric layer and in the underlying constituent transistor) on the performance of a processor in 14 nm NC-FinFET technology. The implementation and analysis are done for various voltages starting from the nominal V_{DD} (0.7 V) down to the sub-threshold V_{DD} (0.3 V). To achieve that, variability-aware standard cell libraries that incorporate the effects of different variability sources, have been created for the first time. The implementation is done for both 14 nm NC-FinFET and its corresponding counterpart baseline (conventional) 14 nm FinFET for various V_{DD}. This enables us to accurately investigate, using existing commercial statistical timing analysis tools from Synopsys, how the performance of any complex circuit like processors will be affected by the joint impact of process variations and voltage scaling. Such an in-depth analysis provides guidance on how process variations can ultimately impact the performance of complex circuits and reveals the deleterious effects of additional variability sources induced by the additional ferroelectric layer.

II. Simulation Setup and Implementation

Fig. 2 provides a general overview of our implementation to study the impact of process variations from the device level to the processor level. Individual steps are as follows.

A. BSIM-CMG Model Calibrations for the Baseline FinFETs

Calibrations for the BSIM-CMG model, which is the industy compact model for the FinFET technology [16] (see Fig. 2(a)), using SPICE simulations, are done to reproduce Intel 14 nm FinFET $I_{ds} - V_{gs}$ and $I_{ds} - V_{ds}$ measurements obtained from [17] for both nFinFET and pFinFET devices, similar to [18]. The modelcard parameters related to low field mobility and mobility degradation, sub-threshold swing (SS), drain induced barrier lowering (DIBL), velocity saturation, and series resistance etc. of both nFinFET and pFinFET devices are carefully extracted following the BSIM-CMG compact model global parameter extraction procedure [19]. Fig. 3 summarizes the calibration results for both nFinFET and pFinFET devices (L_g = 20 nm and NFin = 1). In Fig 3(a), we demonstrate how SPICE simulation results (after BSIM-CMG model calibrations) have a very good matching with measurement data w.r.t $I_{ds} - V_{gs}$ for both linear ($|V_{ds}| = 50$ mV) and saturation ($|V_{ds}| = 0.7$ V) regions. Additionally, Fig. 3(b) shows also the very good matching between the SPICE simulations and measurement data with respect to $I_{ds} - V_{ds}$ for various $|V_{gs}|$ biases, starting from 0.3 V to 0.7 V.

B. NC-FinFET Model Description

The NC-FinFET comprises of two entities (see Fig. 2(b)), the baseline constituent FinFET and the additionally incorporated ferroelectric layer inside the gate stack. The baseline part is modeled using BSIM-CMG with experimentally-calibrated modelcard parameters of Intel 14 nm FinFET technology as discussed above in Section II-A. The ferroelectric physics is modeled using the phenomenological Landau-Devonshire (L-D) theory [20], which expresses the voltage across the ferroelectric (V_{fe}) as a function of its terminal charge (Q) as shown in (1).

$$V_{fe} = t_{fe}(2\alpha Q + 4\beta Q^3) \tag{1}$$

where α and β are ferroelectric material dependent constants and t_{fe} is the thickness of ferroelectric layer. In this work, $Hf_{0.5}Zr_{0.5}O_2$ is used as the ferroelectric material with $\alpha = -1.1 \times 10^9$ m/F and $\beta = 2.5 \times 10^{10}$ $m^5/(C^2F)$ [21]. It is worth mentioning here that these parameters were extracted in [21] by fitting the Landau model to a real experimentally measured S shaped polarization-electric field curve [22]. A ferroelectric thickness (t_{fe}) of 7.7nm (as reported in [21]) is used in all the simulation results shown in this paper.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

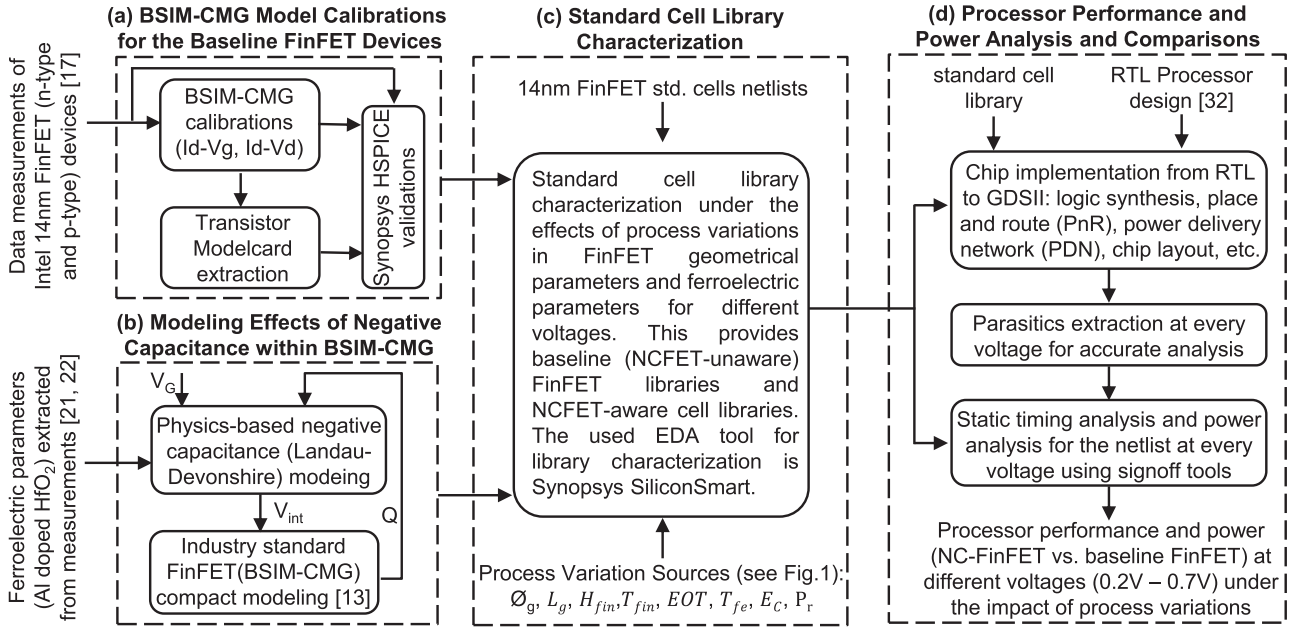


Fig. 2. General overview of our implementation. We start from Intel 14 nm FinFET measurement data that is used to calibrate the industry compact model of FinFET technology (BSIM-CMG) model. Then, variability-aware NC-FinFET standard cell libraries, that consider the effects of process variations, are created using the calibrated BSIM-CMG compact model, after incorporating a physics-based NC model inside it. Characterized cell libraries are later employed within the Synopsys commercial tool flows to accurately estimate the impact of process variations on the delay and power of processor at various voltages. Variability in the baseline FinFET has been first calibrated and benchmarked against measurement data obtained from the same used technology node (i.e., Intel 14 nm FinFET), as will be demonstrated in Section II-C and Fig. 5.

After obtaining all the model parameters, the L-D equation is solved self-consistently within the BSIM-CMG model, constrained by the equality of terminal charges of the ferroelectric and the baseline gate. Note that for simulation purpose, we have assumed the presence of a metal layer between the ferroelectric and the baseline gate, different from a more practical configuration. The metal-ferroelectric-metal-insulator-semiconductor (MFMIS) NCFET in the absence of gate leakage and domain formation shows trends very similar to the metal-ferroelectric-insulator-semiconductor (MFIS) NCFET [23], [24]. Although, using the MFMIS over the MFIS may change the absolute numbers in the circuit simulations, the conclusions and observed trends remain the same for both the device configurations. On the other hand, the MFMIS compared with the MFIS enables utilization of already existing computationally-efficient industry standard compact models such as BSIM-CMG [16], which is indispensable for standard cell library characterization to create NCFET-aware libraries. Without such libraries, the performance of complex circuits like processors cannot be studied.

Fig. 4(a, b) show the simulated $I_D - V_G$ characteristics of NC-nFinFET and NC-pFinFET along with the corresponding baseline FinFET devices. Note that gate workfunction values of NC-nFinFET and NC-pFinFET are kept different from the corresponding baseline nFinFET and pFinFET in order to ensure a hysteresis free operation of the Voltage Transfer Characteristics (VTC) of inverter, as we will explain shortly. As can be noticed, both NC-nFinFET and NC-pFinFET devices have a much steeper subthreshold slope (compared to baseline FinfET) due to the voltage amplification effect of NC. This, in turn, enables NC-FinFET to achieve the same I_{ON} (as in the baseline FinFET) but at a lower voltage (see the dashed line in Fig. 4(b)). This explains how NCFET technology results in a considerable power saving while still maintaining the same baseline performance [4]. The average amplification factor A_{avg} (see (2) and (3)) for NC-nFinFET and NC-pFinFET as a function of V_{DD} is shown in Fig. 4(c). The NC-pFinFET is found to exhibit a greater A_{avg} as compared to NC-nFinFET.

$$V_{int}(V_G) = V_{int}(0) + A_{avg} V_G, A_{avg} = \frac{1}{V_G}\int_0^{V_G} A_V \, dV_G \tag{2}$$

$$A_V = \frac{\partial V_{int}}{\partial V_G} = \frac{|C_{\text{fe}}|}{|C_{\text{fe}}| - C_{\text{int}}}, \quad |C_{\text{fe}}| > C_{\text{int}} \Rightarrow A_V > 1 \tag{3}$$

Fig. 4(d) shows the VTC of inverter built using NC-FinFETs for various voltages (from $V_{DD} = 0.7\,\text{V}$ down to $0.3\,\text{V}$). The NC-Inverter shows a hysteretic VTC for the same gate workfucntion values of NC-nFinFET and NC-pFinFET as their corresponding baseline nFinFET and pFinFET. The hysteresis originates as a result of the negative differential resistance in $I_D - V_D$ characteristics of the constituent NC-nFinFET and NC-pFinFET [25]. A hysteresis free VTC at every V_{DD} was achieved (see Fig. 4(d)) by systematically tuning the workfunction values of the constituent NC-FinFETs.

C. Considering Variability at the Device and Circuits Levels

The impact of process variations at the device level is captured by performing Monte-Carlo (MC) simulations with 2000 samples in a commercial SPICE simulator (Synopsys HSPICE) [26]. For the baseline FinFET, the considered

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

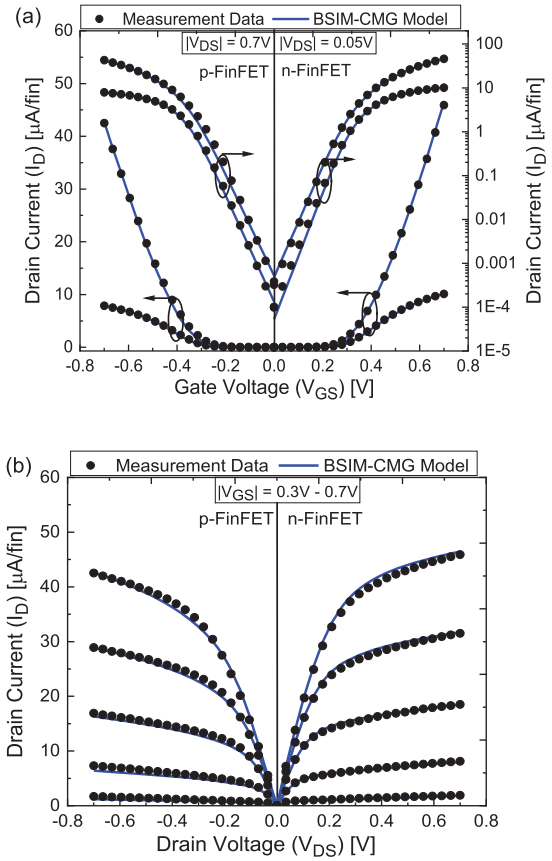


Fig. 3. (a and b) show the calibrations of BSIM-CMG (the industry compact model for FinFETs) to reproduce measurement data for Intel 14 nm nFinFET and pFinFET devices obtained from [17]. (a) demonstrates the validation of the extracted modelcard in the case of I_{ds}-V_{gs} at high and low V_{ds} biases in which a very good matching between SPICE simulations and measurement data is shown. (b) demonstrates the validation of the extracted modelcard for the case of I_{ds}-V_{ds} at various V_{gs} biases.

sources of variability are the gate work-function (ϕ_g), channel length (L_g), fin height (H_{fin}), fin thickness (T_{fin}) and effective oxide thickness (EOT). Note that our methodology and implementation are not limited to specific type of variations, and other variability sources can be analogously considered. We observed that increasing the number of MC samples from 2000 to 5000 and up to 10000 negligibly affects the overall variation data as well as the resulting regression curve ($I_{\rm ON}$ vs. $I_{\rm OFF}$), which we employ to calibrate and benchmark against measurement data (see Fig. 5). Therefore, 2000 MC samples have been selected in our analysis. All the sources of variability are assumed to have a Gaussian distribution. The procedure to assign their standard deviation (σ) values is discussed as follows. First, we assume the standard deviation in the gate work-function due to metal gate granularity ($\sigma\phi_g$) to be 15mV for the baseline nFinFET as demonstrated in [27]. This amounts to $\sigma\phi_g/\mu\phi_g\% = 0.34\%$ with the mean value of $\mu\phi_g = 4.425$V. [27] provided this information for TiN metal gate in 22 nm technology node. However, moving from 22 nm downto 14 nm is not expected to cause a significant difference because the gate surface area in advanced nodes is very small in which it can be covered by a single TiN grain as claified in [27]. In our analysis, we assume the same $\sigma\phi_g/\mu\phi_g\%$ for the pFinFET devices as well but the pFinFET has a different work-function mean value ($\mu\phi_g = 4.7$V).

Thereafter, we evaluate the amount of variability originating from all the other sources in both baseline pFinFETs and nFinFETs, i.e., from L_g, $H_{\rm fin}$, $T_{\rm fin}$ and EOT. To this end, we calibrate the variability in $I_{\rm OFF}$ vs. $I_{\rm ON}$ plot obtained from our Monte-Carlo simulations against the Intel 14 nm FinFET variability measurement data [17] of both nFinFETs and pFinFETs. Keeping the $\sigma\phi_g$ values for the nFinFETs and pFinFETs as explained above and assuming an equal contribution of rest of the variability sources (i.e., L_g, T_{fin}, H_{fin} and EOT), we iterate through MC simulations over their $\sigma/\mu\%$ values until the slope of the regression curve from our MC ($I_{\rm OFF}$ vs. $I_{\rm ON}$) data points matches with the slope of the Intel 14 nm measurement data obtained from [17]. We find that for $\sigma/\mu\% = 0.642\%$ in the remaining variability sources (i.e., L_g, T_{fin}, H_{fin} and EOT), a good matching between the regression curve of our MC data and the measurement data [17] can be obtained for both the nFinFETs and pFinFETs, as demonstrated in Fig. 5(a) and 5(b), respectively. It is noteworthy that the experimental data used for our variability calibration and benchmarking is obtained from the same Intel 14 nm FinFET technology node [17] against which we have earlier calibrated the characteristics of the baseline nFinFET and pFinFET devices (see Fig. 3).

For the NC-FinFET, in addition to the aforementioned variability sources that affect the underlying constituent FinFET, the ferroelectric layer also contributes to the overall process variations. Such additional sources are represented by ferroelectric layer thickness (t_{fe}) along with the material properties of the used ferroelectric, which are the coercive field (E_c) and remnant polarization (P_r). For the variability originating from $t_{\rm fe}$, we assume the same $\sigma/\mu\%$ value as that of the EOT since both are HfO_2 based materials. For P_r and E_c, σ/μ=5% is assumed because the resulting variation (at such a value) in I_{ON} of both the NC-pFinFET and NC-nFinFETs devices, becomes comparable to that of the baseline FinFET devices at the nominal (baseline) operating voltage $V_{\rm DD} = 0.7$ nm, as shown in Fig. 6(a) and 6(b), respectively. In other words, both baseline FinFET and NC-FinFET devices start from a similar level of variability ($\sigma/\mu\%$) w.r.t I_{ON} for more fair analysis and comparisons.

The impact of process variations at the circuit level is captured by characterizing the standard cell libraries, using a commercial cell library characterization tool (Synopsys SiliconSmart tool) [28], in which the impact of process variations is intrinsically considered (see Fig. 2(c)). For cell library characterizations, we employed our calibrated BSIM-CMG model (as described in Section II-A) along with post-layout SPICE netlists of combinational and sequential standard cells. Those netlists include the parasitics information of cells and they were obtained from the open-source 15 nm FinFET cell library released by Silvaco [29]. The delay, power and variability analysis are done for every cell under various operation conditions in which 7×7 input signal slews and output load capacitance are considered, as typically performed in commercial cell libraries. Note that in every cell analysis, both rise and fall conditions are considered for each input pin. In the created

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

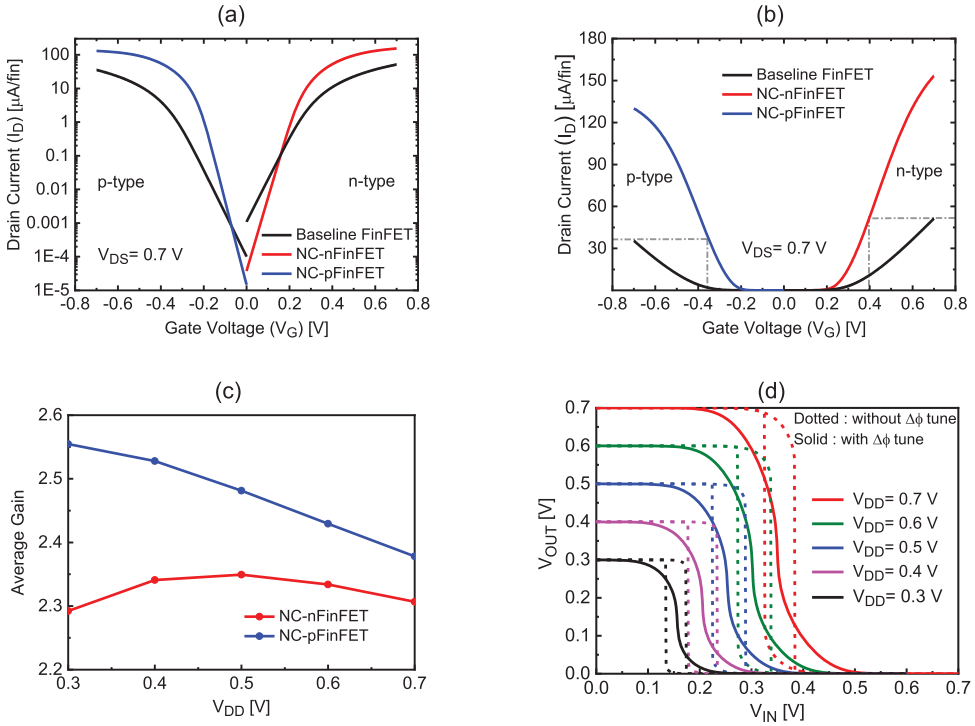


Fig. 4. (a and b) demonstrate the I_{ds}-V_{gs} comparisons between the baseline FinFET and NC-FinFET devices in the log and linear scales, receptively. NC-FinFETs exhibit a steeper sub-threshold slope (compared to the baseline FinFETs) due to the voltage amplification induced by NC. This allows NC-FinFET to achieve the same I_{ON} current (as in the baseline FinFET) but at a lower voltage (see the dashed lines in (b)). (c) shows the average gain due to NC (A_{avg}) for various V_{DD} for both NC-nFinFET and NC-pFinFET. (d) presents the VTC of inverter in NC-FinFETs. When the workfunction in NC-nFinFET and NC-pFinFET are kept as in their corresponding baseline nFinFET and pFinFET, respectively, hysteretic VTC is clearly observed (see the dotted lines). A hysteresis-free VTC was obtained (see the solid lines) by carefully tuning the workfunction ($\Delta\phi$) of NC-nFinFET and NC-pFinFET for every V_{DD} case.

variability-aware standard cell libraries (both baseline FinFET and NC-FinFET), the delay tables for every standard cell is augmented by the corresponding resulting tables of standard deviation (σ). The characterized NCFET-aware cell libraries are compatible with existing commercial EDA software such Synopsys and Cadence [30], [31]. Therefore, they can be directly plugged in and used by the standard statistical timing analysis tools. This enables designers to analyze how the performance of complex circuits like processors (as the focus of this work) can be ultimately affected by process variations (see Fig. 2(d)). The processor implementation is done using the state-of-the-art 32-bit PULP processor [32], which is based on the RISC-V Instruction Set Architecture (ISA), that was recently proposed by Berkeley [33].

The impact of process variations at the processor level is captured by characterizing the standard cell libraries in which the impact of process variations is intrinsically considered. For cell library characterizations, we employed our calibrated BSIM-CMG model (as described in Section II-A) along with post-layout SPICE netlists of combinational and sequential standard cells. Those netlists include the parasitics information of cells and they were obtained from the Silvaco's open-source 15nm FinFET cell library [29]. The delay, power and variability analysis are done for every cell under various operation conditions in which 7×7 input signal slews and output load capacitance are considered, as typically performed in commercial cell libraries. Note that in every cell analysis, both rise and fall conditions are considered for each input pin. In the created variability-aware standard cell libraries (both baseline FinFET and NC-FinFET), the delay tables for every standard cell is augmented by the corresponding tables of standard deviation (σ). The libraries are fully compatible with existing commercial EDA software like [30]. Therefore, they can be directly plugged in and used by the standard statistical timing analysis tools. This enables designers to analyze how the performance of complex circuits like processors (as the focus of this work) can be ultimately affected by process variations. The processor implementation is done using the state-of-the-art 32-bit PULPino processor [32], which is based on the RISC-V architecture [33], recently proposed by Berkeley. For accurate delay and power analysis, a full-chip design has been done and commercial signoff tools have been employed to analyze the final netlist, obtained after place and route, considering the parasitic information of the chip (see Fig. 2). The analysis is done for both baseline FinFET and NC-FinFET at every voltage (0.2 V-0.7 V).

D. Investigated Scenarios at the Device and Processor Levels

To explore the role that the ferroelectric layer plays alone w.r.t process variations as well as the overall (i.e., combined) impact of all variability sources, we investigate the following case-study scenarios at the device and processor levels:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

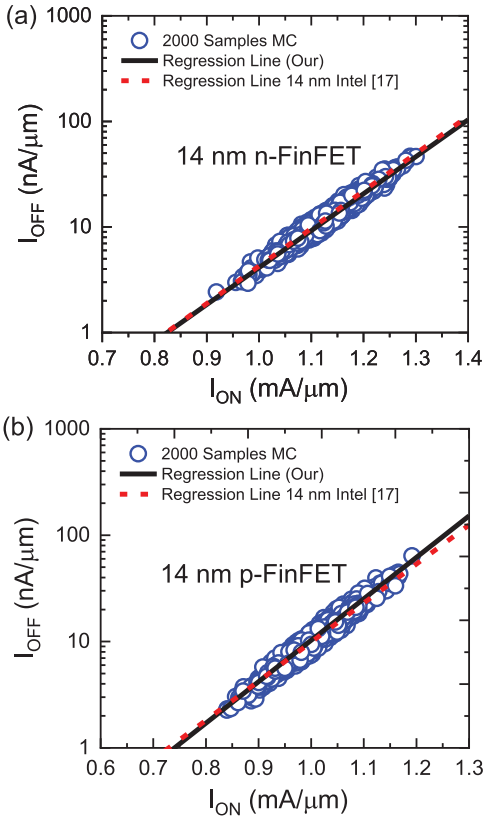


Fig. 5. Variability calibration and benchmarking against Intel 14 nm measurement data [17] for both (a) nFinFET and (b) pFinFET devices. As can be noticed, the overall variation in the I_{ON} and I_{OFF}, obtained from our MC similation data, is close to the measurement data and a good matching between the regression curve of our MC data and the measurement data [17] is obtained for both the nFinFET and pFinFET devices.

(a) Baseline FinFET: Process variations in the conventional baseline FinFET devices are studied. Hence, variations in ϕ_g, L_g, H_{fin}, T_{fin} and EOT are jointly considered.

(b) NC-FinFET (ferro-alone): Process variations in the NC-FinFET devices only in the ferroelectric layer are studied. Hence, variations in t_{fe}, E_C, and P_r are jointly considered.

(c) NC-FinFET (combined): Process variations in NC-FinFET devices both in the ferroelectric layer and in the underlying constituent FinFET. Hence, variations in ϕ_g, L_g, H_{fin}, T_{fin}, EOT, t_{fe}, E_C, and P_r are jointly considered.

III. Impact of Process Variations: Device Level

In the following, we explore the impact of variability on the device ON current (I_{ON} variability = $\sigma I_{\mathrm{ON}}/\mu I_{ON}\%$) and the device gate capacitance (C_{GG} variability = $\sigma C_{GG}/\mu C_{GG}\%$). We have selected these two device parameters as they are the primary factors that govern the propagation delay of logic gates. Including a ferroelectric layer that exhibits NC effect within the transistor gate stack results in a higher ON current but it also makes the total gate capacitance larger [12]. The former (i.e. the higher ON current) leads to a better (i.e., smaller) propagation delay, whereas the latter (i.e., larger gate capacitance) compensates the obtained gain in the propagation delay [12]. Therefore, it is crucial to analyze both device parameters together (i.e., I_{ON} and C_{GG}) to understand the delay improvements later at the processor level.

Fig. 6(a, b) show the impact of variability on I_{ON} variability for the p-type and n-type devices, respectively. As can be noticed, NC-FinFET may exhibit larger or smaller variations in I_{ON} variability in comparison to its baseline FinFET depending on the employed V_{DD}. To understand why we need to look at the two counteracting factors that simultaneously take place when the additional ferroelectric layer in NC-FinFET comes into play: (i) A superior electrostatic integrity caused by the voltage amplification effect due to NC effects, i.e., a larger gate control over the channel. This, in turn, tends to suppress device-to-device variations in the ON current. Note that I_{ON} variability in the baseline decreases with higher V_{DD} (see Fig. 6(a, b)) and hence a voltage amplification due to NC is expected to reduce the effects of variability. (ii) The additional sources of variability arising from the ferroelectric layer itself, i.e. the variation in t_{fe}, P_r and E_c play a major role. The variability in the ferroelectric layer contributes to a significant fraction of the total I_{ON} variability of NC-FinFET (i.e., the scenario of combined variation) as demonstrated in Fig. 6(a, b).

It is noteworthy that NC-pFinFET exhibits either a smaller or similar I_{ON} variability (in the combined scenario) in comparison to its baseline FinFET for all V_{DD} values as seen in Fig. 6(a). This is because the voltage amplification effect in NC-pFinFET largely compensates the additional ferroelectric-induced variability. However, due to a smaller voltage gain in NC-nFinFET (see Fig. 4(c)), the second factor (i.e., the additional ferroelectric-induced variability) dominates at high V_{DD} values causing I_{ON} variability to be larger than its baseline, as can be seen in Fig. 6(b). When reducing V_{DD}, I_{ON} variability of the baseline FinFETs grows up more rapidly compared to the corresponding NC-FinFETs (see Fig. 6(a, b)). As a result, below $V_{DD} \approx 0.5$ V, the NC-nFinFET starts to exhibit a smaller variability than the corresponding baseline FinFET (see Fig. 6(b)). Alternatively, these trends can be understood by realizing the fact that the voltage amplification effect in NC-FinFET is more effective in suppressing its constituent baseline FinFET variability for small V_{DD} values because the baseline FinFET exhibits a much large sensitivity to change in V_{DD} in this regime. Furthermore, as can be observed from Fig. 6(a) and 6(b), the lower sensitivity of the NC-FinFETs to V_{DD} variation can be attributed to their better electrostatic integrity due to the internal voltage amplification.

Finally, Fig. 6(c, d) show C_{GG} variability for the p-type and n-type devices, respectively. The baseline devices (both pFinFET and nFinFET) exhibit a negligible variation in C_{GG} variability with V_{DD} scaling. Consequently, it is expected that the voltage amplification effects obtained by the ferroelectric in NC-FinFET will have a negligible impact on suppressing C_{GG} variability. Hence, C_{GG} variability of the NC-FinFETs always becomes higher than that of the corresponding baseline FinFET due the additional variability sources induced by the ferroelectric layer itself. In other words, the inherent voltage amplification in NC-FinFET plays a negligible role when it comes to C_{GG} variability, and thus, C_{GG} variability will

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

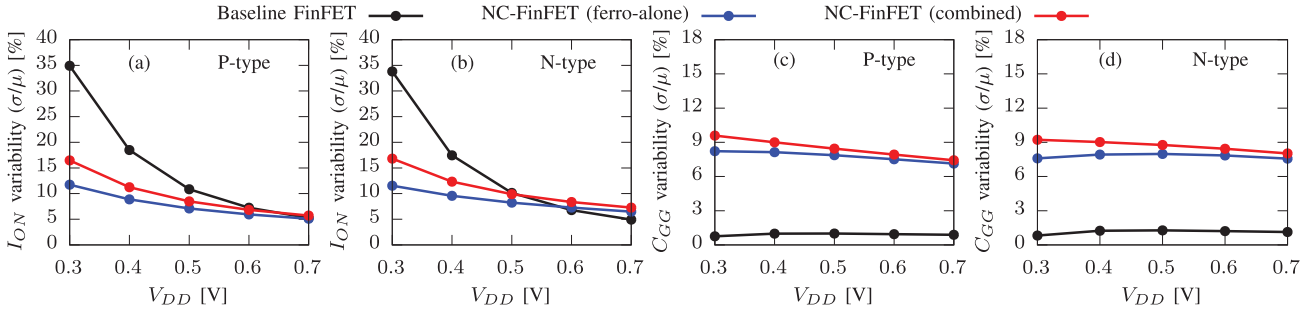


Fig. 6. Impact of process variations at the device level. Three different scenarios are investigated as described in Section II-D. (a and b) show the resulting variation in the ON current for both p-type and n-type devices, respectively. In (a), I_{ON} variability in NC-pFinFET exhibits less variation than the baseline because the impact of voltage amplification (caused by the ferroelectric layer) compensates the additional ferroelectric-induced variability. The voltage amplification in NC-nFinFET is smaller than in NC-pFinFET (see Fig. 4(c)). Therefore, the additional ferroelectric-induced variability dominates in NC-nFinFET at high V_{DD} unless 0.5 V is reached (see (b)). (c and d) demonstrate the resulting gate capacitance (C_{GG}) variation for both p-type and n-type devices, respectively. NC-FinFETs always exhibit a higher variation compared to its baseline FinFET because the additional FE-induced variability dominates.

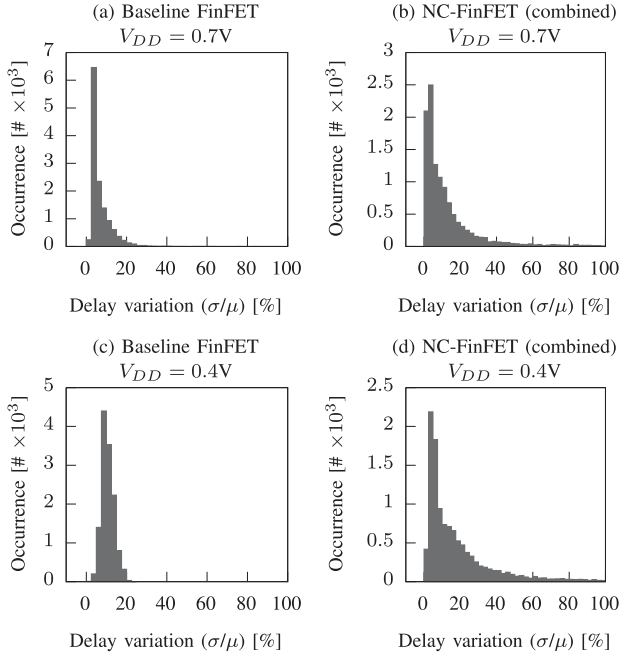


Fig. 7. The impact of process variation on the gate level. (a and b) show the distribution of delay variation across all combination and sequential gate within the standard cell library for the case of baseline FinFET and NC-FinFET, respectively, at $V_{DD} = 0.7$ V. (c and d) show the same analysis but at $V_{DD} = 0.4$ V. The variability analysis in NC-FinFET is done for the scenario of combined variability (details in Section II-D). As can be noticed, the delay of standard cells in NC-FinFET are more sensitive to process variations compared the baseline FinFET (distributions are much wider).

by dominated by the deleterious impact of the additional ferroelectric-induced variability (see Fig. 6(c, d)).

IV. Impact of Process Variations: Gate Level

As explained in Section II-C, we characterize the standard cell library for the baseline FinFET and NC-FinFET cases under the effects of process variations. Before presenting the variability analysis results at the processor level in Section V, we show in following the impact of process variations on the delay of all combinational and sequential gates across the entire standard cell library. Fig. 7(a, b) demonstrate the distribution of delay variation for the case of baseline FinFET and NC-FinFET, respectively, at $V_{DD} = 0.7$ V. Fig. 7(c, d) show the same analysis but at $V_{DD} = 0.4$ V. In the presented analysis in Fig. 7, the variability analysis for the NC-FinFET is performed for the combined variability scenario as it covers all sources of variability (both in the ferroelectric layer and the underlying constituent FinFET). As can be noticed, the delay of standard cells in the case of NC-FinFET exhibits larger variation at the same voltage. As shown earlier in Fig. 6, for the case of $V_{DD} = 0.7$ V, except the case of I_{ON} variability in p-type device, NC-FinFET suffers more from variability in both I_{ON} and C_{GG} parameters. This makes the sensitivity of the delay of standard cells to process variations in NC-FinFET higher than in baseline FinFET (i.e., the distribution of σ/μ is much wider), as Fig. 7 shows.

It is noteworthy that the key promise of NC-FinFET technology is that achieving the same I_{ON} current but at a lower V_{DD} compared to the baseline (i.e. conventional) FinFET, as explained in Section I. As shown in Fig. 3(b), NC-FinFETs (p-type and n-type devices) achieve the same I_{ON} current, as in the baseline FinFET, but at around 0.4 V (instead of 0.7 V). However, the sensitivity of NC-FinFET to process variations at $V_{DD} = 0.4$ V, compared to the sensitivity of the baseline FinFET to process variations at nominal voltage $V_{DD} = 0.7$ V, is much larger as the comparison between Fig. 7(a) and Fig. 7(d) demonstrates. Such an increase in the sensitivity to process variations in NC-FinFET results in larger timing guardbands to sustain reliability. Otherwise, errors caused by timing violations may crash the processor's operation at runtime. Employing a larger timing guardband directly leads to diminishing the efficiency gain that is expected from NC-FinFET, as will be later evaluated.

V. Impact of Process Variations at the Processor Level: Performance Analysis

In the following, we first present how process variations impact the processor delay variability. Then, we explore the ability of ring oscillator to predict the impact of process variations in comparison to the actual delay of the processor. Finally, we demonstrate how process variation results in larger timing guardbands and how it degrades the efficiency gain obtained by NCFET technology.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

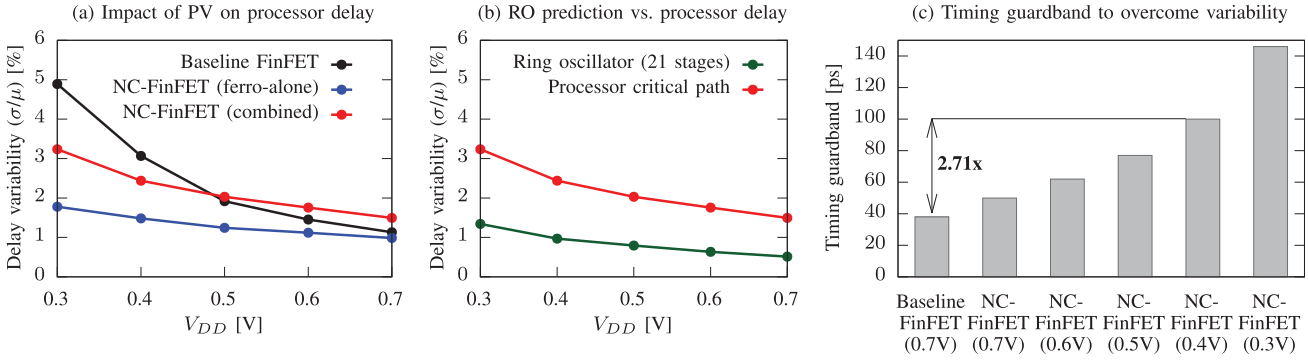


Fig. 8. (a) shows the delay variability analysis of processor under the effects process variations (PV). The included ferroelectric layer brings additional sources of variation that increases the delay variability of processor (see NC-FinFET (ferro-alone) case). Including additionally the variation sources in transistor device itself increases further the delay variability of processor (see NC-FinFET (combined) case) in high V_{DD} regime. However, in the small V_{DD} range (< 0.5 V), NC-FinFET shows a better immunity to process variations due to its less sensitivity to the V_{DD} scaling (see Fig. 6(a) and 6(b). (b) presents a comparison w.r.t the delay variability between 21-stages RO circuit and the actual critical path of processor. As shown, the RO exhibits a similar trend but it considerably underestimate the impact of process variation (delay variability of processor). (c) shows the required timing guardband (3σ) to protect the processor against process variation. At the same nominal voltage (0.7V), NC-FinFET needs 32% larger guardband than the baseline FinFET and the guardband becomes even much larger when V_{DD} scales down reaching 2.71x at 0.4V.

A. Processor Delay Variability Analysis

As explained in Section II-C, we create variability-aware standard cell libraries in which the impact of process variations is included in the delay information of every cell. The libraries have been created for the three studied scenarios: (a) baseline FinFET, (b) NC-FinFET (ferro-alone) and (c) NC-FinFET (combined), as described in Section II-D. The generated gate-level netlist after synthesizing the hardware description (i.e., Register-Transfer-Level RTL) of the processor is employed within the statistical static timing analysis (SSTA) to extract delay information of processor under the effects of process variations. This provides us with μ and σ values of the processor critical path delay. First, we present in Fig. 8(a), the sensitivity of processor to process variations represented by the delay variability (i.e., $\sigma/\mu\%$). As can be noticed, the processor delay exhibits a higher delay variability in the case of NC-FinFET for the high V_{DD} compared to the baseline FinFET. The inclusion of a ferroelectric layer results in additional sources of variation (i.e., t_{fe}, P_r and E_c) that lead to increasing the processor delay variability compared to the baseline FinFET, as shown in NC-FinFET (ferro-alone) in Fig. 8(a). The processor delay variability becomes even higher when the variation sources related to the underlying constituent FinFET itself (i.e., ϕ_g, L_g, H_{fin}, T_{fin}, EOT) are additionally included, as shown in the case of NC-FinFET (combined) in Fig. 8(a). In both baseline FinFET and NC-FinFET, the delay variability increases with lower V_{DD}. However, the sensitivity of the NC-FinFET to V_{DD} scaling is less, compared to the baseline FinFET, due to the inherent voltage amplification by the included NC. Therefore, in the small V_{DD} regime ($<$0.5V), the NC-FinFET processor delay shows a better immunity towards the process variability as compared with the FinFET processor. These trends at the processor level are in line with the trends at the device level shown in Fig. 6(a) and 6(b) where NC-FinFETs exhibit a lower variability in the low V_{DD} range due to a better electrostatic integrity achieved via the NC effect.

B. Ring Oscillator Prediction Analysis

When it comes to evaluating the overall impact of process variations on the delay variability in NCFET-based circuits, state of the art analyzed only a ring oscillator (RO) circuit [13]. In order to investigate whether such an RO-based delay analysis is able to provide a correct prediction w.r.t NC-FinFET process variations, we demonstrate in Fig. 8(b), the delay variability (σ/μ) for a 21-stages RO at different V_{DD} in comparison to the actual delay variability of the processor obtained from the previous analysis in Fig. 8(a).

The NC-FinFET (combined) scenario has been considered in this analysis for both RO and processor case studies because it covers all variability sources. As can be noticed, while RO-based analysis can still provide a proper prediction for the trend of delay variability over voltage scaling, RO-based analysis considerably underestimates the delay variability and the actual delay variability of processor is always much higher (between 2x-3x) compared to the prediction of the RO circuit. This is because the critical path of processor consists of various types of standard cells and different cells may exhibit very different (i.e., non-uniform) delay variability, as Fig. 7 demonstrates. Therefore, the ability of an RO to correctly capture the actual impact of process variations on complex circuits like processors is limited and creating variability-aware cell libraries (as we propose in this work) is necessary. Note that RO circuits can be built from other types of cells (i.e., NAND, XOR, etc.) and the delay variability of such other RO circuits might be different. However, studying different kinds of ROs is out of the scope of this work as we mainly focus on the actual delay of a processor.

C. Processor Timing Guardband Analysis

The presence of variability in the delay of processor necessitates the inclusion of a timing guardband to protect the processor against errors due to timing violations at runtime. In the context of this paper, a timing guardband is, in practice, an additional time slack that is added on top of the maximum delay of processor. As mentioned before, when

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
THE MAXIMUM FREQUENCY OF PROCESSOR WHEN NEGLECTING AND CONSIDERING EFFECTS OF VARIABILITY. NEGLECTING VARIABILITY RESULTS IN OVERESTIMATING THE PERFORMANCE IMPROVEMENT OBTAINED BY NC-FINFET

	Frequency of processor *neglecting* process variations		Frequency of processor *considering* process variations	
Operating voltage	Baseline FinFET	NC-FinFET	Baseline FinFET	NC-FinFET
0.7 V	**633 MHz**	893 MHz	612 MHz	855 MHz
0.6 V	467 MHz	847 MHz	448 MHz	805 MHz
0.5 V	299 MHz	796 MHz	283 MHz	750 MHz
0.4 V	138 MHz	**735 MHz**	127 MHz	**685 MHz**
0.3 V	33 MHz	667 MHz	29 MHz	608 MHz

the variability-aware cell libraries are used inside the SSTA, the tool provides both the mean (μ) and standard deviation (σ) of the delay of the processor's critical path. The timing guardband (GB) is then calculated as $GB = 3 \times \sigma$. A higher pessimism could be alternatively considered in which a larger timing guardband (e.g., $GB = 6 \times \sigma$) is calculated. Fig. 8(c) shows the required timing guardband for the baseline FinFET at the nominal voltage (0.7 V) compared to the NC-FinFET (combined) case for the different voltages (0.7 V down to 0.4 V). As can be observed, NC-FinFET necessitates a larger guardband at the baseline FinFET voltage i.e., $V_{DD} = 0.7$ V because it exhibits a higher delay variability (see Fig. 8(a)). The required guardband increases from 38 ps in the baseline FinFET to 50 ps in the NC-FinFET (i.e., by 32%). Scaling down V_{DD} increases further the guardband in which it reaches 2.71x at 0.4V.

D. Processor Performance and Power Analysis

The final operating frequency of processor in the presence of process variation is $freq = 1/(\mu + GB)$. Therefore, when the processor exhibits a higher delay variability (as in the case of NC-FinFET), a larger GB will be needed to sustain reliability. An increase in the GB directly results in a loss in efficiency as the processor will be then clocked at a lower frequency. To explore and compare the ultimate impact of process variations on both baseline FinFET and NC-FinFET, we present in Table I the maximum processor's frequency when neglecting and considering process variations. As shown, NC-FinFET always provides a higher frequency than baseline FinFET due to NC effect. However, neglecting the impact of process variations leads to overestimating the performance (frequency) improvement. For instance, when neglecting process variations, one will estimate that the frequency of NC-FinFET processor to be 893 MHZ at 0.7 V but, in reality, the frequency should be 855 MHZ.

One of the motivations behind NCFET technology is to operate circuits at a higher frequency without any increase in the original (baseline) power consumption [6]. In other words, under the same power budget as that of the baseline FinFET technology, the NC-FinFET technology is expected to boost the processor's performance. However, the necessity for timing guardbands to overcome variability (as shown in Fig. 8(c)) can diminish to some degree the obtained performance improvement. To evaluate that, we analyze the total power consumption in NC-FinFET processor at different V_{DD} in comparison to the baseline FinFET and summarized the results in Fig. 9. Note that NC-FinFET processor consumes at the same V_{DD} (0.7 V) larger power (> 3x) compared to the baseline FinFET due to the considerable increase in the total chip capacitance caused by NC effects [12].

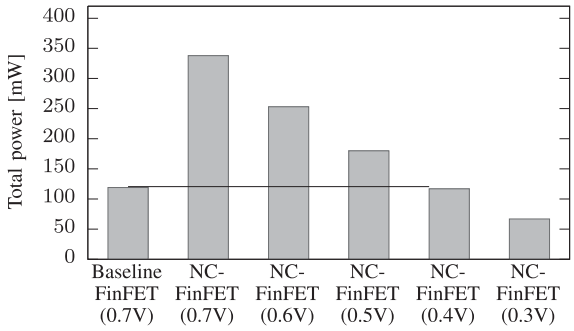


Fig. 9. Operating the NC-FinFET processor at 0.4V results in a similar power as the baseline FinFET processor while it provides 16.2% higher frequency (see Table I). However, when process variation effects are considers the improvement in the processor improvement obtained by NC-FinFET drops down to 8.25% (see Table I) due to required timing guardband (see Fig. 8(c)). Note that NC-FinFET processor consumes at 0.7V larger power (> 3x) due to the considerable increase in the total chip capacitance caused by NC effects.

As can be observed in Fig. 9, at 0.4 V, the NC-FinFET processor consumes a very similar power as in the baseline FinFET processor but it can be clocked at a higher frequency as Table I clarifies. As shown in Table I, the maximum frequency of the NC-FinFET processor at 0.4 V is 735 MHZ, which is 16.2% higher than the original processor's frequency in the baseline FinFET (633 MHZ). However, when considering the effects of variability, the maximum frequency of the NC-FinFET processor at 0.4 V drops to 685 MHZ (due to the required timing guardband as shown in Fig. 8(c)). This, in turn, reduces the frequency improvement obtained by NC-FinFET from 16.2% down to 8.2%.

In summary, neglecting process variations leads to the estimation that NC-FinFET processor can operate with a 16.5% higher frequency, while it still fulfills the same power budget as in the baseline FinFET. However, when process variations come into play, the processor's frequency becomes lower due to the larger timing guardband, and as a result, the actual frequency improvement becomes only 8.25%.

VI. SUMMARY AND CONCLUSIONS

In this work we studied the impact of process variations in NC-FinFET technology. Our implementation starts from calibrating the industry compact model BSIM-CMG to reproduce Intel 14 nm FinFET device as well as variability measurement data all the way up to the processor level. Various variability sources in both ferroelectric layer and underlying constituent FinFET are investigated at the device level as well as processor level. Using standard commercial signoff tools, after creating variability-aware standard cell libraries, we demonstrated the impact of NC-FinFET on the processor's performance. Our results revealed that the frequency improvement for an NC-FinFET processor, that is required to operate under the same power budget as in the baseline FinFET processor,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

is actually lower than what it would be estimated in the absence of process variations. All in all, considering the effects of process variations is essential to correctly estimate the actual benefit of NC-FinFET with respect to efficiency gain in processors.

ACKNOWLEDGMENT

Authors would like to thank Subrat Mishra and Souvik Mahapatra from IIT Bombay for TCAD device calibrations. Authors also thank Sami Salamin from KIT for his valuable efforts in chip design and associated delay and power analysis.

REFERENCES

[1] S. Salahuddin and S. Datta, "Use of negative capacitance to provide voltage amplification for low power nanoscale devices," *Nano Lett.*, vol. 8, no. 2, pp. 405–410, Feb. 2008, doi: 10.1021/nl071804g.

[2] V. V. Zhirnov and R. K. Cavin, "Negative capacitance to the rescue?" *Nature Nanotechnol.*, vol. 3, no. 2, pp. 77–78, Feb. 2008, doi: 10.1038/nnano.2008.18.

[3] G. Han, J. Zhou, Y. Liu, J. Li, Y. Peng, and Y. Hao, "Experimental investigation of fundamentals of negative capacitance fets," in *Proc. IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf.*, Oct. 2018, pp. 1–2, doi: 10.1109/S3S.2018.8640179.

[4] H. Amrouch, G. Pahwa, A. D. Gaidhane, J. Henkel, and Y. S. Chauhan, "Negative capacitance transistor to address the fundamental limitations in technology scaling: Processor performance," *IEEE Access*, vol. 6, pp. 52754–52765, Sep. 2018, doi: 10.1109/access.2018.2870916.

[5] S. K. Samal, S. Khandelwal, A. I. Khan, S. Salahuddin, C. Hu, and S. K. Lim, "Full chip power benefits with negative capacitance FETs," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Jul. 2017, doi: 10.1109/islped.2017.8009170.

[6] M. Rapp, S. Salamin, H. Amrouch, G. Pahwa, Y. Chauhan, and J. Henkel, "Performance, power and cooling trade-offs with NCFET-based many-cores," in *Proc. 56th Annu. Design Autom. Conf.*, Jun. 2019, p. 41, doi: 10.1145/3316781.3317880.

[7] J. Müller *et al.*, "Ferroelectricity in simple binary ZrO_2 and HfO_2," *Nano Lett.*, vol. 12, no. 8, pp. 4318–4323, 2012, doi: 10.1021/nl302049k.

[8] T. S. Boscke, J. Muller, D. Brauhaus, U. Schroder, and U. Bottger, "Ferroelectricity in hafnium oxide: CMOS compatible ferroelectric field effect transistors," in *Proc. Int. Electron Devices Meeting*, Dec. 2011, pp. 24.5.1–24.5.4.

[9] S. Mueller *et al.*, "Incipient ferroelectricity in Al-doped HfO_2 thin films," *Adv. Funct. Mater.*, vol. 22, no. 11, pp. 2412–2417, Jun. 2012, doi: 10.1002/adfm.201103119.

[10] Z. Krivokapic *et al.*, "14nm ferroelectric FinFET technology with steep subthreshold slope for ultra low power applications," in *IEDM Tech. Dig.*, Dec. 2017, pp. 1–15, doi: 10.1109/iedm.2017.8268393.

[11] K. Chatterjee, A. J. Rosner, and S. Salahuddin, "Intrinsic speed limit of negative capacitance transistors," *IEEE Electron Device Lett.*, vol. 38, no. 9, pp. 1328–1330, Sep. 2017, doi: 10.1109/led.2017.2731343.

[12] H. Amrouch, S. Salamin, G. Pahwa, A. D. Gaidhane, J. Henkel, and Y. S. Chauhan, "Unveiling the impact of IR-drop on performance gain in NCFET-based processors," *IEEE Trans. Electron Devices*, vol. 66, no. 7, pp. 3215–3223, Jul. 2019, doi: 10.1109/ted.2019.2916494.

[13] T. Dutta, G. Pahwa, A. Agarwal, and Y. S. Chauhan, "Impact of process variations on negative capacitance FinFET devices and circuits," *IEEE Electron Device Lett.*, vol. 39, no. 1, pp. 147–150, Jan. 2018, doi: 10.1109/led.2017.2770158.

[14] Y.-K. Lin *et al.*, "Effect of polycrystallinity and presence of dielectric phases on NC-FinFET variability," in *IEDM Tech. Dig.*, Dec. 2018, pp. 9.4.1–9.4.4, doi: 10.1109/iedm.2018.8614704.

[15] S.-E. Huang, C.-L. Yu, and P. Su, "Investigation of Fin-width sensitivity of threshold voltage for InGaAs and Si negative-capacitance FinFETs considering quantum-confinement effect," *IEEE Trans. Electron Devices*, vol. 66, no. 6, pp. 2538–2543, Jun. 2019, doi: 10.1109/ted.2019.2907994.

[16] Y. S. Chauhan *et al.*, *Finfet Modeling for IC Simulation and Design*. New York, NY, USA: Academic, 2015, doi: 10.1016/B978-0-12-420031-9.00012-9.

[17] S. Natarajan *et al.*, "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size," in *IEDM Tech. Dig.*, Dec. 2014, pp. 3–7, doi: 10.1109/iedm.2014.7046976.

[18] S. Mishra *et al.*, "A simulation study of NBTI impact on 14-nm node FinFET technology for logic applications: Device degradation to circuit-level interaction," *IEEE Trans. Electron Devices*, vol. 66, no. 1, pp. 271–278, Jan. 2019, doi: 10.1109/ted.2018.2875813.

[19] (Oct. 2018). *BSIM-CMG Technical Manual*. [Online]. Available: http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG

[20] A. F. Devonshire, "XCVI. Theory of barium titanate: Part I," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 40, no. 309, pp. 1040–1063, Oct. 1949, doi: 10.1080/14786444908561372.

[21] M. Hoffmann, B. Max, T. Mittmann, U. Schroeder, S. Slesazeck, and T. Mikolajick, "Demonstration of high-speed hysteresis-free negative capacitance in ferroelectric $Hf_{0.5}Zr_{0.5}O_2$," in *IEDM Tech. Dig.*, Dec. 2018, pp. 31.6.1–31.6.4, doi: 10.1109/iedm.2018.8614677.

[22] M. Hoffmann *et al.*, "Unveiling the double-well energy landscape in a ferroelectric layer," *Nature*, vol. 565, no. 7740, pp. 464–467, Jan. 2019.

[23] G. Pahwa, A. Agarwal, and Y. S. Chauhan, "Numerical investigation of short-channel effects in negative capacitance MFIS and MFMIS transistors: Subthreshold behavior," *IEEE Trans. Electron Devices*, vol. 65, no. 11, pp. 5130–5136, Nov. 2018, doi: 10.1109/ted.2018.2870519.

[24] G. Pahwa, A. Agarwal, and Y. S. Chauhan, "Numerical investigation of short-channel effects in negative capacitance MFIS and MFMIS transistors: Above-threshold behavior," *IEEE Trans. Electron Devices*, vol. 66, no. 3, pp. 1591–1598, Mar. 2019, doi: 10.1109/ted.2019.2892186.

[25] T. Dutta, G. Pahwa, A. R. Trivedi, S. Sinha, A. Agarwal, and Y. S. Chauhan, "Performance evaluation of 7-nm node negative capacitance FinFET-based SRAM," *IEEE Electron Device Lett.*, vol. 38, no. 8, pp. 1161–1164, Aug. 2017, doi: 10.1109/led.2017.2712365.

[26] (Jul. 2019). *HSPICE*. [Online]. Available: https://www.synopsys.com/verification/ams-verification/hspice.html

[27] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal-gate transistors, Part I: Modeling, analysis, and experimental validation," *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2504–2514, Oct. 2010.

[28] (Jul. 2019). *Synopsys EDA Software, SiliconSmart*. [Online]. Available: https://www.synopsys.com/implementation-and-signoff/signoff/siliconsmart.html

[29] (Jul. 2019).*Silvaco EDA Software, Open-Source Standard Cell Libraries*. [Online]. Available: https://www.silvaco.com/news/pressreleases/2019_05_30_01.html

[30] (Jul. 2019). *Synopsys EDA Software*. [Online]. Available: https://www.synopsys.com/

[31] (Jul. 2019). *Cadence EDA Software*. [Online]. Available: https://www.cadence.com/

[32] (Jul. 2019). *PULP Platform*. [Online]. Available: http://www.pulp-platform.org/

[33] (Jul. 2019). *RISC-V*. [Online]. Available: https://riscv.org/

Hussam Amrouch (Member, IEEE) received the Ph.D. degree *(summa cum laude)* from KIT in 2015. He is currently a Research Group Leader with the Chair for Embedded Systems (CES), Karlsruhe Institute of Technology (KIT), Germany. He is also leading of the Dependable Hardware Research Group. His main research interests are emerging technologies, design for reliability from physics to system level, and machine learning. He holds seven HiPEAC Paper Awards. He has three best paper nominations at DAC'16, DAC'17, and DATE'17 for his work on reliability. He has served in the technical program committees of Design Automation Conference (DAC), International Conference on Computer Aided Design (ICCAD), and Asia and South Pacific Design Automation Conference (ASP-DAC). He currently serves as an Associate Editor at *Integration, the* IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS . He also severs as a reviewer in several journals like the IEEE TRANSACTIONS ON ELECTRON DEVICES, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS: REGULAR PAPERS I, the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, and others.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Girish Pahwa (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Delhi Technological University, Delhi, India, in 2014. He is currently pursuing the Ph.D. degree with IIT Kanpur, Kanpur, India. His current research interests include physics and modeling of negative capacitance transistors.

Amol D. Gaidhane (Graduate Student Member, IEEE) received the M.Tech. degree in VLSI from the Center for VLSI and Nanotechnology, Visvesvaraya National Institute of Technology, Nagpur, India, in 2016. He is currently pursuing the Ph.D. degree with IIT Kanpur, Kanpur, India. His current research interests include simulation, modeling, and fabrication of negative capacitance transistors.

Chetan K. Dabhi (Graduate Student Member, IEEE) received the M.Tech. degree in VLSI engineering from the SV National Institute of Technology Surat (SVNIT), India, in 2015. He is currently pursuing the Ph.D. degree with the IIT Kanpur, Kanpur, India. His current research interests include compact modeling of industry-standard BSIM Models, compact model development for aging analysis at the circuit/system level, physics and the compact modeling of negative capacitance state-of-the-art FDSOI transistor.

Florian Klemme received the B.Sc. degree in system integration from the University of Applied Sciences Bremerhaven, Germany, in 2014, and the M.Sc. degree in computer science from the Karlsruhe Institute of Technology (KIT), Germany, in 2018. He is currently pursuing the Ph.D. degree with the Chair of Embedded Systems, Karlsruhe Institute of Technology. His research interests include cell library characterization and machine learning techniques in the VLSI design flow.

Om Prakash (Member, IEEE) received the Ph.D. degree in microelectronics and VLSI with the Department of Electronics and Communication Engineering, IIT Roorkee, India. He is currently a Postdoctoral Researcher with the Karlsruhe Institute of Technology, Germany. His current area of research includes nanoscale device-circuit co-design, compact modeling, and device reliability such as Bias Temperature Instability (BTI) and self-heating effects.

Yogesh Singh Chauhan (Senior Member, IEEE) was with the Semiconductor Research & Development Center, IBM, Bengaluru, from 2007 to 2010; the Tokyo Institute of Technology in 2010; the University of California at Berkeley from 2010 to 2012; and ST Microelectronics from 2003 to 2004. He is currently an Associate Professor with IIT Kanpur (IITK), India. He is also the developer of industry standard BSIM-BULK (formerly BSIM6) model for bulk MOSFETs and ASM-HEMT model for GaN HEMTs. His group is also involved in developing compact models for FinFET, Nanosheet/Gate-All-Around FET, FDSOI transistors, Negative Capacitance FETs, and 2D FETs. He has published more than 200 articles in international journals and conferences. His research interests are characterization, modeling, and simulation of semiconductor devices. He is the member of IEEE-EDS Compact Modeling Committee and a fellow of the Indian National Young Academy of Science (INYAS). He received Ramanujan Fellowship in 2012, the IBM Faculty Award in 2013, and P. K. Kelkar Fellowship in 2015, the CNR Rao Faculty Award, and Humboldt Fellowship in 2018. He has served in the technical program committees of the IEEE International Electron Devices Meeting (IEDM), the IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), the IEEE European Solid-State Device Research Conference (ESSDERC), the IEEE Electron Devices Technology and Manufacturing (EDTM), and the IEEE International Conference on VLSI Design and International Conference on Embedded Systems. He is the Founding Chairperson of the IEEE Electron Devices Society U.P. chapter and a Vice Chairperson of IEEE U.P. Section. He is an Editor of the IEEE TRANSACTIONS ON ELECTRON DEVICES and a Distinguished Lecturer of the IEEE Electron Devices Society.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 02,2020 at 07:21:08 UTC from IEEE Xplore. Restrictions apply.

International Journal of Electrical Engineering and Technology (IJEET)
Volume 11, Issue 2, March-April 2020, pp. 178-189, Article ID: IJEET_11_02_022
Available online at http://www.iaeme.com/IJEET/issues.asp?JType=IJEET&VType=11&IType=2
ISSN Print: 0976-6545 and ISSN Online: 0976-6553
Journal Impact Factor (2020): 10.1935 (Calculated by GISI) www.jifactor.com
© **IAEME** Publication

IMPLEMENTATION OF ELLIPTIC - CURVE CRYPTOGRAPHY

Monika
Department of Applied Mathematics,
Delhi Technological University, Delhi, India

Tanya Tomar
Department of Applied Mathematics,
Delhi Technological University, Delhi, India

Vineet Kumar
Department of Applied Mathematics,
Delhi Technological University, Delhi, India

Yogesh Kumar
Department of Applied Mathematics,
Delhi Technological University, Delhi, India

ABSTRACT

Elliptic curve cryptography (ECC) is a procedure to generate public key between two distant partners namely, Alice and Bob used in public key cryptography. This method is based on the algebraic structure of elliptic curves over finite fields. ECC is important in the sense that it involve keys of smaller length in comparison to other non-Elliptic curve cryptography to provide equivalent security. During this work, we implement an algorithm in python programming language to generate public key using the method of ECC.

Keywords: Diophantine Equation, Elliptic - Curve Cryptography, Elliptic Curve Diffie – Hellman Key Exchange, Elliptic Curve Integrated Encryption Scheme.

Cite this Article: Monika, Tanya Tomar, Vineet Kumar and Yogesh Kumar, Implementation of Elliptic - Curve Cryptography, *International Journal of Electrical Engineering and Technology*, 11(2), 2020, pp. 178-189.
http://www.iaeme.com/IJEET/issues.asp?JType=IJEET&VType=11&IType=2

1. INTRODUCTION

A finite field is defined as a field with finite number of elements. It is also known as Galois field. The number of elements in a finite field is always prime p or a power of a prime p n. For each prime power q where q = p n, there exists one finite field Fq (up to an isomorphism). These finite fields form an abelian group with respect to the binary operation ∗, and therefore,

every finite field has a generator G. It is a very well known fact that computer stored every number in binary form and they are the components in finite field whose order is two. Also data can be stored in a computer as a vector in a finite field that allows mathematical operations to remove data easily and effectively. Finite Field has widespread application in Cryptography. Since each byte of data can be represented as a vector in a finite field so encryption and decryption schemes using basic mathematical operation is easily manipulable. Data Encryption Standard (DES) developed by IBM uses only 56-bit key. Since the advancement of technology break the 56-bit key in less than 24 hours so revolutionary algorithm is needed [1– 4]. In 2001, Vincent Rijmen and John Daemon discovered a more complicated encryption algorithm to cater the need of advanced encryption algorithm, now known as Advanced Encryption Standard (AES). Elliptic-curve cryptography (ECC) is a public-key cryptography whose working is based on elliptic curves over the finite fields. The algorithm of ECC mainly works on Elliptic Curve Discrete Logarithm Problem (ECDLP) that is very hard to solve quickly. If some mathematical problem is posed in terms of elliptic curve then there is no straight method that gives the solution of the posed problem. So the only way to find the solution is to try with random numbers. Since each bit size provides more options than RSA so there is more room for growth. The amount of data which is necessary to exchange or transfer for validation is less than RSA. ECC requires less computing power and battery resource. Despite of many advantages of ECC, there are only a few implementation of ECC. This motivate us to work on the implementation of Elliptic Curve cryptography. In this work, we have studied few cryptographical protocols based on elliptic curve and then we have generated code in python programming language to implement ECC.

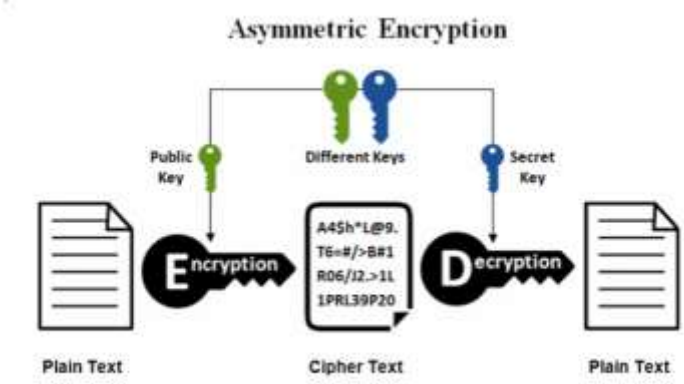


Figure 1 Asymmetric Encryption

This is often however most hybrid encryption schemes works (the encryption process)

Figure 2 Encryption process

This is often however most hybrid decryption schemes works (the decryption process)

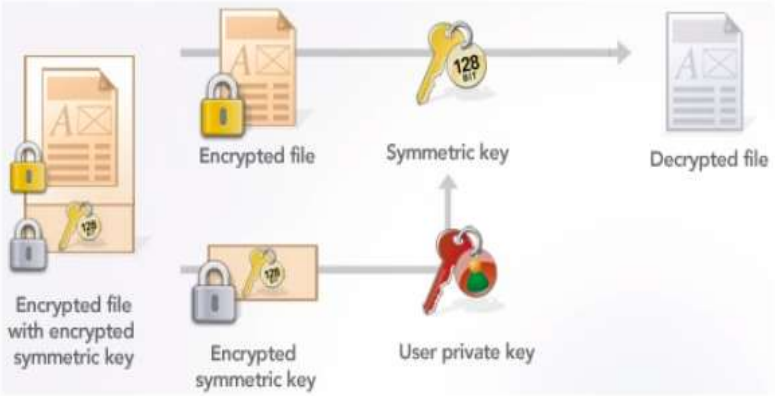


Figure 3 Decryption process

1.1. Elliptic - curve

An elliptic curve E over finite field Fq is the set of all points (x, y) lying on the curve

$$y^2 = x^3 + ax + b \quad (1)$$

where a ,b ∈ Fq and $4a^3 + 27b^2 \neq 0$, together with a point at infinity (∞).

E forms an abelian group with the identity element ∞.The rule for the addition of two points on the curve is given below:

- If $P = (x_1, y_1) \in E$ then $-P = x_1, -y_1$
- If $Q = (x_2, y_2) \in E$, $Q = -P$, then $P + Q = (x_3, y_3)$,

where $x_3 = \lambda^2 - x_1 - x_2$, $y_3 = \lambda(x_1 - x_3) - y_1$ (2)

and

$$\lambda = (y_2 - y_1) / (x_2 - x_1), \quad \text{if } P \neq Q$$
$$\lambda = (3x_1^2 + a) / 2y_1, \quad \text{if } P=Q \quad (3)$$

Example: Consider the elliptic curve E : $y^2 = x^3 + x + 1$ defined over Z_{13}. Generator of $E(Z_{13})$ is P = (0,1). The points on the curve E can be expressed in terms of P as:

$$P = (0,1) = (x_1, y_1), 2P = (x_3, y_3) \quad (4)$$

Considering equation 3:

$$\lambda = (3x_1^2 + a) / 2y_1, \quad \text{if } P = Q$$
$$= \frac{3(0)^2 + 1}{2(1)} \bmod 13$$
$$= \frac{1}{2} \bmod 13$$
$$x_3 = \lambda^2 - 2x_1$$
$$= (3)^2 - 2(0)$$
$$= 9$$

Considering equation 2:

$$y_3 = \lambda(x_1 - x_3) - y_1$$
$$= ((3(0-0)-1) \bmod 11$$
$$= -1 \bmod 11$$
$$= 10$$

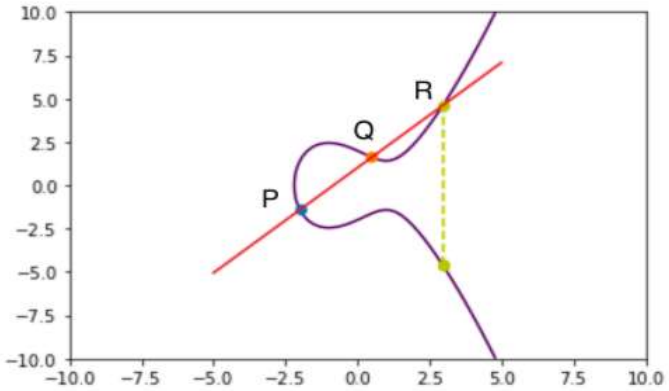


Figure 4 Elliptic curve

It is known that every elliptic curve is not useful in cryptography. So it is important to know what type of elliptic curve is useful in Cryptography. The elliptic curves that form cyclic groups and the elliptic curve point that generates all the points of the cyclic subgroup are useful in cryptography. Cryptography based on Elliptic curve is known as Elliptic curve cryptography (ECC). A 160-bit ECC key is roughly equivalent to a 1024-bit RSA key.

2. ELLIPTIC CURVE DIFFIE-HELLMAN KEY EXCHANGE PROTOCOL

In this section, we will discuss the Diffie–Hellmann key ex- change protocol on elliptic curves. The Elliptic Curve Diffie- Hellmann (ECDH) protocol is a secret key generation protocol between two distant parties Alice and Bob and the secret key is one of the point on the elliptic curve. There are also some practical implementations of ECDH that can be found in [6– 9]. The ECDH protocol can be described as follows: Suppose E denote the elliptic curve over F_q. Before starting the protocol, let us assume that the two distant partners Alice and Bob mutually agreed upon a point Q on the elliptic curve. Alice secretly chooses a randomly generated integer K_A and calculate the point $X = K_AQ$. Then Al- ice send X to Bob. Likewise, Bob also generate an integer K_B randomly and compute the point $Y = K_BQ$. Bob then send Y to Alice. After obtaining X from Alice, Bob compute $K_BX = K_BK_AQ$ and in a similar fashion, Alice compute $K_AY = K_AK_BQ$. Since K_A and K_B are integers so $K_BK_AQ = K_AK_BQ$. Thus the common shared key between Alice and Bob is $K_S = K_AK_BQ$. If there is any eavesdrop- per in the middle who wanted to steal the information contained in the encoded text then the eavesdropper should have knowledge about the secret key K_A and K_B of Alice and Bob. But since the secret key K_A and K_B is known only to Alice and Bob respectively so it is not possible for any eavesdropper to gain information about the encoded message.

2.1. Implementation of ECDH in Python Programming Language

We are now in a position to implement ECDH protocol in python programming language. We use the tinyec library for ECC in Python.

```
pip install tinyec
 from tinyec import registry
 import secrets
 def compress(publicKey):
 return hex(publicKey.x) + hex(publicKey.y % 2)[2:]
 curve = registry.get_curve('brainpoolP256r1')
```

```
Ka = secrets.randbelow(curve.field.n)
X = Ka ∗ curve.g print("X:", compress(X))
Kb = secrets.randbelow(curve.field.n)
Y = Kb ∗ curve.g print("Y:", compress(Y))
print("Currently exchange the publickey (e.g. through Internet)")
aliceSharedKey = Ka ∗ Y
print("Alice shared key :",compress(aliceSharedKey))
bobSharedKey = Kb ∗ X
print("Bob shared key :",compress(bobSharedKey))
print("Equal shared keys:", aliceSharedKey == bobSharedKey)
```

2.2. Example

We consider here that the secret keys K_A and K_B are generated randomly by sender Alice and receiver Bob using the code explained in the previous section. Therefore, the randomly generated keys K_A and K_B are given by

K_A = 0*x*66*c*808*e*6*b*5*be*6*d*6620934*bc*6*ffa*2*b*8*b*47*f* 9786*c*00 2*bfb*06*d*53*a*0*c*27535641*a*5*d*1 (5)

K_B = 0*x*7*d*15195432*d*1*ac*7*f*38*aeb*054*d*07*d*9*b*2*e*1*faa*913*b*7
*d*08*a*4*d*5*efdd*4*a*1*ee*8*d*9*a*31916*d*53*a*0*c*27535641*a*5*d*0*d*1 (6)

Let us assume that Alice and Bob pre-agreed with the point Q given by

Q = (0*xd*458*e*7*d*127*ae*671*b*0*c*330266*d*246769353*a*012073*e*97*acf*8,
0*x*325930500*d*851*f* 336*bddc*050*cf* 7*fb*11*b* 5673*a*1645086*df* 3*b*) (7)

When Alice send the point $X = K_A Q$ to Bob and Bob shared the point $Y = K_B Q$ with Alice then the generated secret key is shared between Alice and Bob . This secret key is common for both the users and is given by

K_S=0*x*90*f*5*a*1*cf*2*ed*1*dbb*0322178*df*6*bb*0*dd*72*c*541884618*b*2989*a*3*e*5*e*663198 667*a*621 (8)

The elliptic curve which is being used for the ECDH calculations is 256-bit named curve brainpoolP256r1 [14] (uses Diophantine equation for the generation of points). The private keys are randomly 256-bit (64 hexadecimal digits) . The public keys and shared keys are 257 bits (65 hexadecimal digits , 256 bit due to key compression). Due to randomization the secret keys K_A and K_B are different but the calculated shared secret key between Alice and Bob will always be same.

3. ECC-BASED SECRET KEY DERIVATION (USING ECDH)

In this section we will derive the shared secret key which is used for symmetric encoding with the help of same rule as in ECDH mentioned above with some modification. The discussion over secret key generation protocol is as follows: Assume we have a cryptographic elliptic curve E over finite field F_q and its generator point Q on the elliptic curve .Con- template we have a EC private-public key pair. We would like to encrypt and decrypt data using these keys. Working of Asymmetric encryption can be given as : if we encrypt the plain text by a private key K_P and calculate $X = K_P\ Q$ which is the public key known in the medium for transfer of data, then likewise we will decrypt the cipher text later by the as- sociated ciphertext public key $Y = K_C\ Q$ which is calculated by considering the private key of cipher text K_C. We will use the subsequent two functions to calculate a shared ECC secret key S :

- Calculate Encryption Key($K_P Q$) –> (S, $K_C Q$)
  1. Generate new random private key = K_C.
  2. Calculate $Y = K_C Q$.
  3. Calculate $S = (K_P Q) * K_C$.
  4. Return each the $S + (K_C Q)$. Use the S for symmetric encryption. Use the (K_C Q) to calculate the decryption key
- Calculate Decryption Key($K_P , K_C Q$) –> S
  1. Calculate $S = (K_C Q) * K_P$.
  2. Return the S and further use it for the decoding.

The equation takes the subsequent form:

$(K_P * Q) * K_C = (K_C * Q) * K_P = S$

This is what these two functions calculate, following the ECDH key derivation scheme. Within the hybrid encryption schemes the encapsulated (K_C Q) is referred to as ”ephemeral key” , as it is employed temporary, to derive the symmetric encryption key, using the ECDH key agreement scheme.

3.1. Implementation of ECC-Based Secret Key in Python Programming Language

The below Python programming language code uses the tinyec library to urge a ECC private-public key pair for the message recipient (based on the brainpoolP256r1 curve) so derive a S (for encryption) and (K_C Q) (for ECDH) from the recipient’s public key and later derive same S (for decryption) from the recipient’s private key and thus the generated earlier ($K_C * Q$) :

```
from tinyec import registry
import secrets
curve =registry.get_curve('brainpoolP256r1')
def compress_point(point):
return hex(point.x) + hex(point.y % 2)[2:]
def ecc_calc_encryption_keys(X):
    Kc = secrets.randbelow(curve.field.n) Y = Kc * curve.g
    S = X * Kc
    return (S, Y)
def ecc_calc_decryption_key(Kp, Y):
    S = Y * Kp
    return S
Kp = secrets.randbelow(curve.field.n)
  X = Kp * curve.g
  print("Kp:", hex(Kp))
  print("X:", compress_point(X))
(encryptKey, Y) = ecc_calc_encryption_keys(X)
```

```
print("Y:",compress_point(Y))
print("encryption key:",compress_point(encryptKey))
decryptKey = ecc_calc_decryption_key (Kp, Y)
    print("decryption key:",compress_point(decryptKey))
```

3.2. Example

We consider here that the private secret key K_P is generated randomly by medium for plain and cipher text in Asymmetric encryption using the code explained in the previous section. Therefore, the randomly generated keys K_P is given by:

$$K_P = 0x2e2921b4cde59cdf01e7a014a322abd530b30\ 15085c31cb6e59502da761d29e9 \quad (9)$$

Let us assume that the generator point Q in the protocol is given by

$$Q = (0xd458e7d127ae671b0c330266d246769353a012073e97acf8,\ 0x325930500d851f336bddc050cf7fb11b5673a1645086df3b) \quad (10)$$

The public keys which are known in the medium for the transfer of data are denoted by X and Y for plain text and cipher text respectively

$$X = 0x850d3873cf4ac50ddb54ddbd27f8225fc43bd3f4c2cc0a4f9d1f\ 9ce15fc4eb711 \quad (11)$$

$$Y = 0x71586f9999d3ee050005054bc681c1d96c5eb054ca\ 15b080ba245e495627003b0 \quad (12)$$

The generated common key shared for encryption and decryption in symmetric encryption

$$Encryptionkey = 0x9d13d3f8f9747669432f575731926b5ed99a6883f\ 00146cbd3203ffa\ 7ff\ 8b1ae1 \quad (13)$$

$$Decryptionkey = 0x9d13d3f8f9747669432f575731926b5ed99a6883f\ 00146cbd3203ffa\ 7ff\ 8b1ae1 \quad (14)$$

$$Encryptionkey = Decryptionkey = S \quad (15)$$

The code demonstrates that we are able to generate a pair

$\{S + (K_C * Q)\}$ from given EC public key and later we are able to recover the similar secret key from the pair $\{ (K_C * Q) + K_P \}$.

It is clear that encryption key (derived from the public key) and the decryption key (derived from the corresponding private key) are the equivalent. The above mentioned keys are going to be used for encoding and decoding in an integrated encryption scheme.

4. ELLIPTIC CURVE INTEGRATED ENCRYPTION SCHEME

The expanded version of encryption and decryption scheme is the Elliptic Curve Integrated Encryption Scheme (ECIES). This scheme may be considered as another form of the ElGamal scheme proposed by Abdalla, Bellare, and Rogaway in [10] and can be considered as a hybrid encrytion scheme. It is using Diffie-Hellmann protocol followed by symmetric encryption scheme. A standard symmetric key which is encoded with a 1024 bits RSA key mainly produces an output of 128 bytes as compared with the output of 84 bytes if the encryption is performed with ECIES [11].The scheme ECIES can be implemented using following three algorithms:

- Key generation algorithm,
- Encryption algorithm and
- Decryption algorithm.

ECIES can be shown by the following diagram:

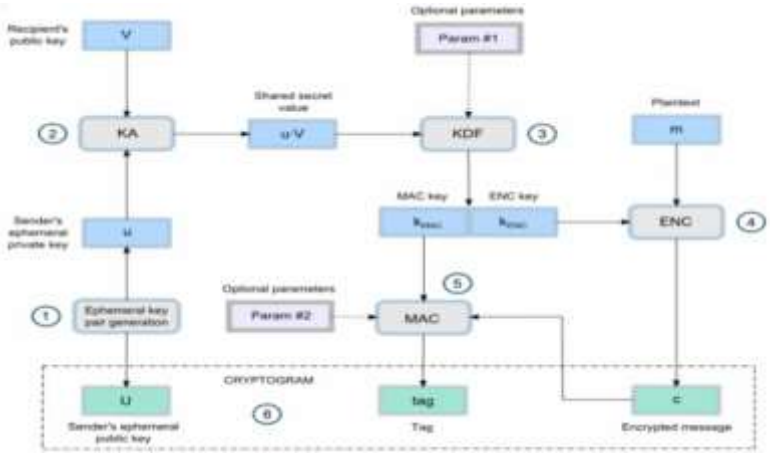


Figure 5 ECIES

1. Algorithm for key generation is as follows:

- Choose a shared secret $Q \epsilon [1, p]$ as a private key.
- Generate a public key $K_B = [Q]P$, where P is a point on the elliptic curve.
- Return key pair (K_B, Q).

2. Encryption scheme algorithm can be described as follows:

In this encryption scheme, let us assume that Alice wants to send a message to Bob. Bob possesses public key K_B. Alice and Bob both knew the private key Q.

- Alice generates a random number $k \epsilon [1, p]$.
- Alice then calculates $U = [k]P$ and $V = [k]K_B$.
- Alice uses a key derivation function (KDF) to compute two keys k_1 and k_2 from V .
  - uses SHA-256 to hash 192-bit x coordinate of Point V into 256 bit code.
  - breaks it into 128-bit key $k1$ and 128-bit key k_2.
- Alice uses 128-bit AES encryption algorithm for encoding her message m with key k_1, and further obtain cipher text c.
- Alice then chooses HMAC-SHA 256 to calculate a MAC (message authentication) code denoted by r with key k_2.
- Alice sends (U, c, r) as the cipher text to Bob.

3. Decryption scheme algorithm can be explained in the following way:

- Bob split the cipher text e into three components (U, c, r).
- Bob uses his secret private key Q, so he can computes V= [Q]U.
- Bob uses the same key derivation function to obtain $k1$, and $k2$.
- Bob computes MAC using $k2$ and compares the result with r.
- If they are different, then he declare that the message is fake.
- Lastly, Bob uses the 128-bit AES decryption algorithm to decrypt c.

In this way, Bob is able to decode the original message m.

The ECIES encoding scheme is a framework, not a concrete formula. It may be enforced by plugging totally different algorithms, e.g. the secp256k1 elliptic curve for the public-key calculations + Scrypt (Password based KDF) + AES(Advanced Encryption Standard) for symmetric cipher and authentication tag + HMAC-SHA512 for MAC (unauthenticated encryption).

4.1. Implementation of ECIES in Python Programming Language

In this section, we have presented a code in python programming language to implement ECIES protocol. The code is given below.

```
pip install eciespy
from ecies.utils import generate_eth_key
from ecies import encrypt, decrypt
import binascii
Kp = generate_eth_key()
privateKeyHexadecimal = privKey.to_hex()
publicKeyHexadecimal =privKey.public_key.to_hex()
print("Encryption publickey:",publicKeyHexadecimal)
print("Decryption privatekey:",privateKeyHexadecimal)
m = b'Some plaintext for encryption'
print("Message:", m)
encrypted = encrypt(publicKeyHexadecimal,m)
print("Encrypted:",binascii.hexlify(encrypted))
decrypted = decrypt(privateKeyHexadecimal,encrypted)
print("Original Message:", decrypted)
```

The python code above generate ECC public + private key pair using ecies.utils.generate eth key() and call the ecies.encrypt((K_B Q), msg) and decrypt(K_B, encryptedMsg) functions from the eciespy library.

4.2. Example

As discussed earlier the input which is needed to proceed with ECIES are public key which is used in encryption and the message *m* which is encrypted and then transferred to the receiver so that it is not eavesdropped in between the secured channel. Finally at the senders end the encrypted text is con- verted back to original message using private key. So the pair of public-private key generated by using code above are :

Encryptionpublickey=0*x*0*dc*8*e*06*c*055*b*45*ecf*110258*ed*5*c*0261*ce*2019*b*1*bd*0*f*8*f*
226*dcd*010*dade*448*b*8*f*304*a*0915*c*68*cdf*7*ddded*8*e*4021*d*28*fb*92*e*27*d*08*df*695*f*
48*a*0*d*2*c*41*ddee*750*fc*7 (16)

Decryptionprivatekey=0*x*487*fd*8*b*53*c*471*e*3*c*38484*a*0*fbe*4751*ace*67*a*9*ed*28*e*60*ea*6*b*0*b*4
4*c*445*b*881*f* 99*d* (17)

Consider the message *m* over which the ECIES protocol is applied for further understanding:

Message: b'Some plaintext for encryption'

Before the message is sent to the sender this is the encrypted text formed using the code above :

Encrypted=b'045699078*bbd*101*e*270572*d*0*d*68*e*87*a*8*f*7*b*6*cc*377*ebeeffb*60*d*2*fcac*5*dc*7*bdd*86*a*26*d*7*f*79*d*13*b*92*e*923*a*0*e*2*cdbe*418*a*7856*b*27157*ef*150*d*5*c*72*f*4*f*8*f*312467*d*13221*ebe*7049*b*7*ed*2*f*0*ed*253*bce*13117129*a*7*b*01*bb*881*b*8*dfbf*004*ff*11*f*3*ebed*4*c*732744*bc*49*ea*03230*c*2*d*1*b*2*ec*80774*e*79*c*075431*d*2019464*d*3*de*97*ceb*96' (18)

At the sender's side the encrypted text is decrypted to its original message:

Original Message: b'Some plaintext for encryption'

The Python eciespy library uses ECC cryptography over the curve + AES authenticated encoding. Note that the above encrypted message holds these values: {K_C Q, AES nonce, authentication Tag, Advanced Encryption Standard-ciphertext} , packed in binary form.

5. DISCUSSION

Some browsers don't support EC certificates. Microsoft de- signed support for EC into Windows Vista, however earlier versions together with Windows XP do not support ECC. Reliable information for supporting ECC in mobile platforms is- not presently offered. ECC signature verification may be a in- tensive task in carrying computation and it may be slower than RSA on the devices which contain slower processors. Still has some patent issues, particularly for binary curves.

6. CONCLUSION

Encryption is defined as the process of converting plain- text into randomly generated meaningless text referred to as cipher text. Decryption is defined as converting the cipher text to its original form. The aim of every encryption and decryption algorithm is to make it difficult to attack on the data and hence the transfer of data from sender to receiver is un-interrupted because we come across many incidents involving cybercrime (computer -oriented crime) which attacks data or network. We have implemented ECC, and its modified algorithm (ECIES) and (ECDH) in python explaining each step of every algorithm via computer code. Initially considering brain poolP256r1 elliptic curve and applying the EC multiplication property to the points (public and private key) , which returned encrypted data, decrypted data and shared keys. Security benefit of ECC is much more than RSA and DSA. A serious weakness in RSA is sharp increase in key size is remunerated by smaller keys in EC. ECC is additionally quicker for a variety of reasons. Firstly, smaller keys means that there is less data which needs to be transmitted from the server to the client during an SSL handshake [17]. Moreover, EC needs less process power (CPU) and memory, leading to consider- ably quicker response times and throughput on Web servers once it is in use. A essential benefit of using ECC is Perfect Forward Secrecy (PFS) [18]. The cipher suites which is being supported by modern Web servers and browsers which implement PFS also implement ECC. Web servers which prefer Ephemeral ECDH (ECDHE) using cipher suites gain the advantage of each EC and PFS. ECIES uses the most secure way to calculate a key is by employing a secure "key derivation function" (KDF) [19]; the output of KDF is a "derived key." [20] Derived keys are even as secure as random keys.

Particularly, they create it less expensive to buy, use, and maintain systems that uses them.

REFERENCES

[1] ”Fact Sheet NSA Suite B Cryptography”. U.S. National Security Agency. Archived from the original on 2009.

[2] “Commercial National Security Algorithm Suite “.www.nasa.gov 19 August 2015. Archieved from the original on 2091-06-04. Retrived 2020.

[3] RSA Laboratories. ”6.3.4 Are elliptic curve cryptosystems patented?”. Archived from the original on 2016.

[4] Bernstein, D. J. ”Irrelevant patents on elliptic-curve cryptography”.

[5] https://eprint.iacr.org/2011/283.pdf.

[6] Certicom Research, Standards for efficent cryptography, SEC 1: Elliptic Curve Cryptography, Version 2.0, May 21, 2009.

[7] American National Standards Institute (ANSI), Public Key Cryptography for the Financial Services Industry: Key Agreement and Key Transport Using Elliptic Curve Cryptography, X9.63, 2001.

[8] National Institute of Standards and Technology (NIST), Recommendation for Pair-wise Key Establishment Schemes Using Discrete Logarithm Cryptography, SP 800-56A, 2005.

[9] Institute of Electrical and Electronics Engineers (IEEE), Standard Specifications for Public Key Cryptography, Std. 1363, 2000.

[10] M. Abdalla, M. Bellare, P. Rogaway, DHAES: An encryption scheme based on the DiffieHellman problem. Submission to IEEE P1363, November 1998. http://grouper.ieee.org/groups/1363/StudyGroup/; M. Abdalla, M. Bellare, P. Rogaway, DHIES: An encryption scheme based on the DiffieHellman problem, in IEEE P1363a, 1998.

[11] V.G. MAtnez, L.H. Encians, and C.S. Avila, *Journal of computer science and engineering* 2,7, 2010.

[12] Commercial National Security Algorithm Suite and Quantum Computing FAQ U.S. National Security Agency, January 2016.

[13] NSA Suite B Cryptography, Sutie B Implementers’ Guide to NIST SP 800-56A Archived 2016-03-06 at the wayback Machine, July 28, 2009.

[14] ECCurveRegistry. https://github.com/alexmgr/tinyec/blob/master/tinyec/registry.py

[15] HybridSystem. https://www.cryptopp.com/wiki/Elliptic_Curve_Integrated_Encryption_Scheme.

[16] Library. https://pypi.org/project/eciespy/

[17] https://casesecurity.org/2014/06/10/benefits-of-elliptic-curve-cryptography/

[18] https://whatis.techtarget.com/defination/perfect-forward-secracy

[19] “Salted Password Hashing – Doing it Right”. CrackStation.net. Retrieved 29 January 2015.

[20] “Password Hashing Competition”

[21] Akshatha M M, Lokesh B, Nuthan A C, A Novel Technique of Visual Cryptography for Maintaining the Security of Visual Information Transaction, *International Journal of Electronics and Communication Engineering & Technology*, 5(8), pp. 194–199. 2014.

[22] Arun Kumar Gandhi, Dr. S.S. Tyagi , A Survey of Elliptic Curvecryptography Using Character Conversion Method, *Journal of Computer Engineering & Technology*, 1(1), pp. 10–17. 2013.

[23] Dr. R. Mala and K. Karthikeyan. Artificial Neural Cryptography Datagram Hiding Techniques For Computer Security Objects Register, *International Journal of Computer Engineering and Technology*, 7(2), pp. 36–43. 2016.

[24] Ahmed Tariq Sadiq, Amjed Abbas Ahmed, Sura Mazin Ali, Attacking Classical Cryptography Method Using Pso Based On Variable Neighborhood Search, *International Journal of Computer Engineering and Technology*, 5(3), pp. 34–49. 2014.

[25] T. Devasena and Dr. P. Nageswara Rao, Collaboration Pattern in Cryptography Research Output (1976 –2015): A Scientometric Study. *International Journal of Library & Information Science*, 7(2), pp.18–29. 2018.

Journal of ELECTRONIC MATERIALS
https://doi.org/10.1007/s11664-020-08178-8
© 2020 The Minerals, Metals & Materials Society

In-Depth Analysis of Structures, Materials, Models, Parameters, and Applications of Organic Light-Emitting Diodes

SHUBHAM NEGI,[1] POORNIMA MITTAL ,[2,4] and BRIJESH KUMAR[3]

1.—Department of Electronics and Communication Engineering, Graphic Era Deemed to be University, Dehradun, India. 2.—Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi, India. 3.—Department of Electronics and Communication Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, India. 4.—e-mail: poornimamittal@dtu.ac.in

Organic electronics and specifically organic light-emitting diode (OLED) devices based on organic semiconductor materials enable economically viable large-area flexible applications. This review presents a comprehensive and detailed assessment of different aspects related to OLEDs, focusing on the impact of multilayered architectures on device performance, in particular the impact of different layers and architectures to enhance the output from such devices. Furthermore, characteristic parameters, materials, and fabrication methodologies are reviewed in depth to highlight the major advancements related to OLEDs over the years. Mathematical models are important for predicting the internal operation and characteristics of such devices, thus these are also discussed, focusing on methods to improve the performance parameters of OLEDs. Applications of OLEDs are also discussed, with a primary focus on research related to their improvement and enhancement for use in displays, sensors, and visual light communications. Although OLEDs show great promise for a bright future, several challenges such as the development of blue light-emitting materials, lifetime improvement, and application-specific architectures must be addressed to achieve more dynamic devices for emerging commercial applications.

Key words: Organic semiconductor (OSC), organic light-emitting diode (OLED), biosensor, visual light communication (VLC)

INTRODUCTION

Organic electronics is currently a nascent but rapidly growing field,[1–4] being the part of electronic science that deals with organic semiconductors (OSCs).[5] This technology provides an alternative platform for the design of low-cost, large-area, flexible, and economically viable electronic devices[6] and thus opens up new avenues of research.[7–12] It provides the complete new dimension of novel OSC-based electronic devices[5] for further realization into potential applications such as memory,[6] displays,[13] microsensors,[14–19] biosensors,[20–22] wearable fabrics,[23–27] radiofrequency identification (RFID),[28] visual light communication (VLC), and the design of novel circuits for power generation applications.[29–36] The concept is to combine the advantages of both fields, viz. organic and inorganic, to achieve far-reaching results.

OSC-based technology offers numerous advantages, including the flexible nature of such devices,[37–41] the use of unconventional substrate materials (plastics, paper, etc.),[42,43] lower fabrication costs,[44–47] etc. Flexibility[39–41] is one of the most promising benefits of this technology. Furthermore, organic technology involves low-temperature solution-based fabrication processes[48] including dip coating, screen printing,[49] inkjet printing,[50,51] etc.,

(Received December 15, 2019; accepted April 29, 2020)

Published online: 21 May 2020

and the utilization of unconventional substrates (fiber glass, polythene, and cloth). As a consequence, the device size is not restricted to the wafer size (a maximum of 8 inches for conventional technology), and they can be fabricated over a large area such as 10 × roll to roll.[49–51] Additionally, a huge inventory is not required[52] owing to the use of low-temperature processes.[53,54] Therefore, the overall cost of fabrication is greatly reduced.[55] Organic materials are not greatly affected by impurities,[1] hence avoiding stringent cleanroom requirements. Thus, OSC-based devices can play a vital role in future improvements to technology, providing an alternative and motivating platform to achieve low-cost consumer devices.

In comparison with conventional silicon-based technology, organic devices are superior in some regards but are lagging in others. Such devices will have far-reaching outcomes in terms of large-area fabrication,[1] flexibility,[38–41] easier manufacturing processes, robust design,[56–59] and lower cost.[50] On the other hand, in terms of characteristics such as mobility, response time, power consumption, and scaling to nanolevels, conventional silicon-based technology[1] outperforms organic devices.[51] Therefore, these two technologies can be amalgamated in the form of hybrid devices to utilize the advantages of both, as reported by various researchers.[6] Figure 1 shows a detailed comparison of organic- and inorganic-based devices (on a scale from 1 to 10), clearly revealing the advantages and disadvantages of each.

From the first report of a conducting OSC by Shirakawa et al.[60] in 1976 to the present day, substantial breakthroughs have been made. This started with the advent of thin-film technology, which revealed that materials with low mobility could also be used for the design of transistors. This led to the invention of organic thin-film transistors (OTFTs) in 1986 by Tsumuru et al.[61] However, mobility remains a significant bottleneck in the development of organic devices. Nevertheless, with rigorous research efforts, a remarkable augmentation in mobility values has been achieved. At the beginning of the 21st century, the maximum reported mobility was around 3.2 cm^2/V-s[62] for organic thin films, with an average mobility on the order of 10^{-2} cm^2/V-s or poorer. However, a value of 10 cm^2/V-s was reported by Liu et al.[63] in 2018, and average mobility values in the range from 2 cm^2/V-s to 4 cm^2/V-s were obtained. As these values remain low compared with those in conventional devices, architectural features are continuously updated to compensate for these shortcomings.

Another landmark achieved by organic devices was the successful testing of organic circuits based on OTFTs. Subsequently, the organic inverter,[6,55] as well as NAND and NOR logic gates were realized,[6,28] leading to the development of a fully organic six-transistor static random-access memory (6T-SRAM) cell[6] and other complex circuits. Other OSC-based devices such as OLEDs and solar cells have also been explored. Tang et al.[64] reported the first OLED in 1987, and thereafter, tremendous developments were reported with the passage of time. The most important practical use of OLED is in display applications for mobile phones (Apple Co.), and televisions (LG). Moreover, OLED-based sensors and VLC[65] devices have been also reported. The present review focuses on different aspects of OLEDs, focusing on the main points in their development over the years. First, the basic operating principle of an OLED is discussed, then different characteristic parameters related to OLEDs are analyzed. This is followed by a description of different methods utilized by researchers to enhance their performance.

These different methods are related to structural and material aspects, both of which are thus also reviewed comprehensively. The review of structural aspects focuses on novel structures and layers of different materials applied for performance

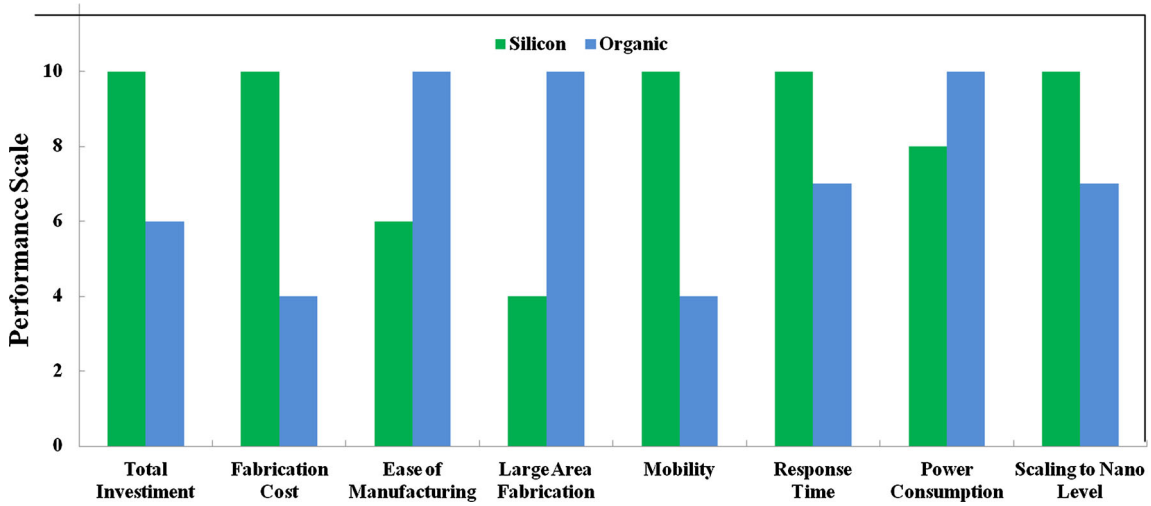


Fig. 1. Comparison of organic and conventional silicon-based technology.

enhancement. The materials summary describes the development of novel emitting materials targeting in particular the color of the emitted light. The performance of electronic devices also depends on the fabrication technology applied, thus this perspective is also given some consideration. Thereafter, models applied to provide detailed knowledge on the operation of OLED devices are considered. Different OLED applications related to the display, biomedical, and communication fields are finally explored.

This review is divided into nine sections, including the present "Introduction" section. "Organic Light-Emitting Diodes" section presents a brief summary of the operation of an OLED. Thereafter, "Characteristic Parameters" section reviews different performance parameters of OLEDs that are imperative to comprehend their specific applications. "OLED Structures" section focuses on different architectural aspects to understand the impact of the structure of an OLED on its performance. The materials facet is explored in "OLED Materials" section, whereas novel fabrication approaches are reviewed in "Fabrication Methodologies" section. "-OLED Models" *section then discusses various models developed for internal investigation of OLED structures, followed by consideration of the use of OLEDs in different novel applications in "OLED Applications" section. Finally, "Conclusions" section concludes the article with important remarks.

ORGANIC LIGHT-EMITTING DIODES

OLED device technology is growing rapidly, based on the continual evolution and development in the field of organic electronics. These devices use OSCs to generate light. Hong Kong–American physical chemist Ching W. Tang and American chemist Steven Van Slyke, at Eastman Kodak, built the first practical OLED device in 1987.[64] Since then, these devices have found wide use in display applications. The technology giants of the display arena (LG and Sony) are working on various technologies for incorporating OLEDs into their displays, with huge success. Hence, analysis of OLEDs and related techniques to enhance their performance becomes essential to stay up to date with this ever-changing technology and ensure their ongoing use.

Operating Principle

The structure of an OLED consists of an OSC material encased on two sides by electrodes. The OLED differs greatly from conventional devices, where *p*- and *n*-type sections directly connected to a supply through metal contacts are responsible for the light emission (Fig. 2a). Unlike conventional light-emitting devices, the OLED has separate electrodes for carrier injection and extraction. The structure of an OLED is illustrated in Fig. 2b. On

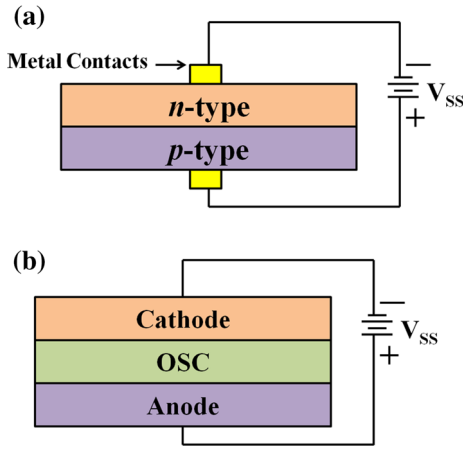


Fig. 2. Basic structure of (a) conventional LED and (b) OLED.

applying a proper bias voltage, charge carriers are injected into the device. Electrons are injected from the cathode, while holes are injected from the anode.[66] Unlike in a transistor, the electrodes are generally not interchangeable in an OLED due to its bipolar nature with different requirements for the injection of electrons and holes.

For the injection of electrons, a low-work-function electrode material such as calcium or magnesium is preferred, offering a work function that coincides with the lowest unoccupied molecular orbital (LUMO) of the OSC material. On the other hand, a high-work-function electrode material such as indium tin oxide (ITO) is required for the injection of holes. This eases the flow of holes into the highest occupied molecular orbital (HOMO) of the OSC material.[67,68] Once the charge carriers have been injected into the OSC, they start to move within the organic material, and on their recombination, energy is released in the form of light. The whole operational process of light emission from an OLED is shown in Fig. 3 and can be broadly categorized into five major steps:

(a) Injection of charge carriers
(b) Transportation of charge carriers
(c) Recombination of charge carriers
(d) Formation of excitons (i.e., excited state)
(e) Radiative decay

A brief discussion on each of these physical processes is provided below:

(a) *Injection of charge carriers:* On the application of a proper bias voltage, electrons and holes are injected into the OSC. Electrons are injected from the cathode into the LUMO, and holes from the anode into the HOMO level. The important parameter for the injection of these charge carriers is the difference in work function between the electrode and OSC the respective energy orbital[66] of the OSC. These

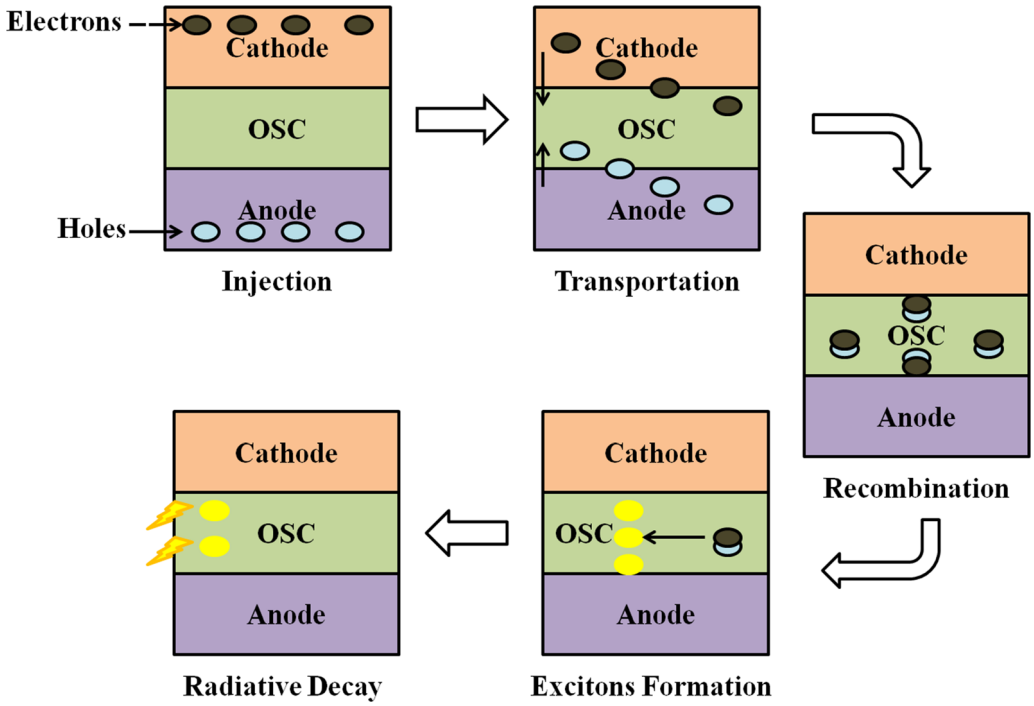


Fig. 3. Operation of OLED, depicting the series of processes for light emission.

orbitals, i.e., the HOMO and LUMO levels, are equivalent to the valence and conduction band, respectively. Generally, one of the electrodes is transparent so that light can be emitted from the device.[13] The injection can be enhanced by using specifically dedicated layers (charge injection layers) to decrease the injection barrier.

(b) *Transportation of charge carriers:* Once injected, carriers travel in the OSC to recombine with each other. However, a significant difference is observed between the mobilities of the different types of charge carrier, often being greater than two orders of magnitude.[5,28] The mobility of holes dominates over its counterpart[28,42] for all materials reported to date. This mobility difference between electrons and holes affects the recombination rate in the device. Ideally, it is desired that the recombination take place at the center of the OSC, but it actually occurs in the vicinity of one electrode due to this mobility difference.[13,66] Therefore, various architectural modifications have been suggested, mainly including the utilization of carrier transport or carrier blocking layers to address this mobility issue.

(c) *Recombination of charge carriers:* The electrons and holes move in their respective LUMO and HOMO energy levels inside the OSC. Some of these carriers come into contact with each other and recombine to form excitons, which further release energy (light) on returning to their ground state.[13] Owing to their low mobility, this recombination process in an organic semiconductor is not governed by conventional physics but can be described by Langevin theory.[69] According to this theory, the recombination rate is governed by the probability that charge carriers will find each other in such lower-mobility devices.[69] On the other hand, the recombination in conventional devices can be described by the Shockley–Read–Hall (SRH) model, specifically with faster-moving charged particles. Recombination in OLEDs may also occur by filling of trap sites. Some free carriers can also lead to the formation of excited states, thus architectural modifications also include carrier-confinement structures.

(d) *Formation of excitons:* An exciton is a state in which an electron and hole are combined because of their mutual Coulomb attraction. After recombination, the electron releases energy to return to the ground state. The energy lost by the electron is carried away by a photon in accordance with the law of conservation of energy.[66] This released energy is greatly affected by the type of semiconductor material, similar to conventional LEDs. Excitons can be of two types, viz. singlet or triplet, depending on their spin state.[66,69] Light is generated when excitons decay radiatively.

(e) *Radiative decay:* Each excited particle returns to its ground state after decaying. The decay process can be radiative or nonradiative.[66] Radiative decay results in spontaneous emission, whereas nonradiative decay leads to contact quenching and heat loss. The efficiency of fluorescent material-based devices is lower

Table I. Luminescence performance of OLEDs over the last 15 years

Year	Applied bias (V)	Luminescence (cd/m^2)	Color/Wavelength (nm)	Ref.
2004	18	10,460	Blue (492)	72
2006	20	22,350	490	73
2006	13	4067	Yellow	74
		2234	White	
2006	18	23,570	540	75
2007	25	33,000	Green	76
2007	9	16,300	525	77
2008	11.5	100	NR*	78
2010	NR	14,110	NR	79
2010	27.5	6000	NR	80
2010	NR	3000	Blue (470)	81
2011	11.5	7068	Green (tandem OLED)	82
	7.2	3133	Green (single-unit OLED)	
2012	10	63,800	Green (fluorescent)	70
	8	1,33,500	Green (phosphorescent)	
2014	16	100	408	71
2014	20	49,999	509	83
2014	10	50,000	NR	83
2018	18	45,869	Red/green (598)	85
	18	11,018	White (495–577)	
2018	NR	22,344	NR	86

than that of phosphorescent devices due to the weakness of the spin–orbital interaction that controls the intersystem crossing between the triplet and singlet states, resulting in light emission from both states. This light emission from singlet and triplet states further increases the device efficiency. As compared with a conventional LED, an OLED exhibits some unique characteristics such as (a) generation of a diverse range of colors,[66] (b) wide viewing angle,[68] (c) minimal shift in the spectral characteristics,[69] etc. Their other exclusive advantages include flexibility, low-temperature fabrication, economically viable realization, and usage of unconventional substrates. Despite all these features, advancements in OLED technology were steady, and it took almost a decade to realize potential OLED-based displays for use in practical applications.

CHARACTERISTIC PARAMETERS

There are various characteristic parameters that describe the performance of an OLED and help to ascertain its viability for a given application. The different characteristic parameters that are helpful in determining the applicability of an OLED in numerous fields include the luminescence, current density, external quantum efficiency, power efficiency, and current efficiency. Various factors affect these parameters, including the electron and hole mobility, emission material used (phosphorescence or fluorescence), device architecture, electrodes, different supporting layers, etc. Generally for an OLED, it is essentially required that these parameters should attain fairly high values.

Luminescence

The luminescence is defined as the intensity of light emitted per unit area measured in candelas per square meter (cd/m^2). It can also be described as a photometric measurement of the light traveling in a particular direction and arriving on a unit area.[69] The luminescence depends greatly on the color of light emitted, the type of material used (phosphorescent or fluorescent), the formation of excitons, the applied voltage, etc. Since good luminescence performance is a basic requirement for any OLED, a lot of effort has been investigated by the scientific community to enhance it over the last two decades. The maximum luminescence recorded for an OLED was 133,500 cd/m^2 when using 4,4′-bis(*N*-carbazolyl)-1,1′-biphenyl) doped with Ir(pyy)$_3$(factris(2-phenylpyridine)iridium) dye (CBP,a phosphorescent materials) in a multilayered device structure.[70] Table I summarizes some of the best luminescence results obtained for OLEDs over the last 15 years.

The results presented in this table reveal that the luminescence has continuously improved over time, with average values in the range from 3000 cd/m^2 to 6000 cd/m^2. Different colors of light show different luminescence values, but the trend is for an increase in the luminescence for each color. The improvement in luminescence values over a period

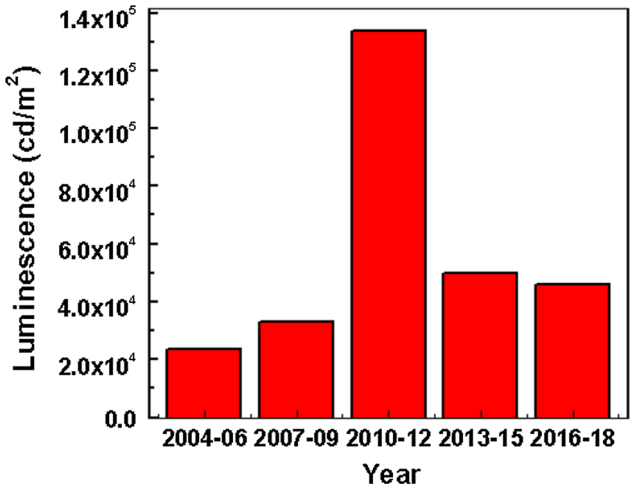


Fig. 4. Trend of luminescence over 15 years.

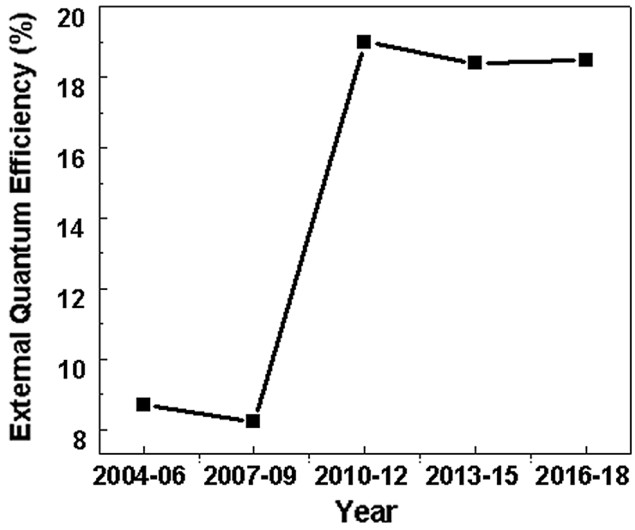


Fig. 5. Improvement in external quantum efficiency over time.

of 15 years is also illustrated in Fig. 4. Initially, Guo et al.[71] reported a low luminescence of 100 cd/m^2 for a new star-shaped material that emitted blue and ultraviolet light. Thereafter, various new approaches were applied in OLED development to enhance the luminescence characteristics. One of these included the use of double hole-blocking layers (HBLs) in the architecture.

Subsequently, Yang et al.[75] focused on the multilayered OLED architecture and improved the luminescence characteristics by 9.5% through the addition of a hole-blocking layer (HBL). Thereafter, a variety of structures such as a multilayered OLED without a HBL, and OLEDs with single and double HBLs were proposed and analyzed. Rigorous device analysis revealed that HBLs indeed help in improving the luminescence of OLEDs by directly affecting the recombination rate in the device. Through this analysis, it was observed that the luminescence could be improved from 17,000 cd/m^2 for a multilayered OLED without HBL to 23,750 cd/m^2[75] for the same OLED with the inclusion of HBLs. Similarly, Yun et al.[70] reported an improved performance for an inverted top-emitting structure with both fluorescent and phosphorescent materials. Those researchers emphasized that the thickness and proper matching of the supporting materials for the electron transport layer (ETL), hole transport layer (HTL), and HBL significantly affect the performance of the OLED. By modifying these layers, maximum luminescence values of 54,500 cd/m^2 and 133,500 cd/m^2 were achieved for fluorescent and phosphorescent materials, respectively.

Further, Ohmori et al.[76] proposed a high-performance OLED with doping of highly emissive phosphorescent dyes as host materials for emission of green and red colors. Subsequently, a device with [Ir(ppy)$_3$] showed a high luminescence of 33,000 cd/m^2, whereas a value of 8800 cd/m^2 was reported for the red-emitting dye [Ir(piq)$_3$].[76] The luminance is not high for red color, but reasonable compared with previous results.[76] Additionally, the luminescence performance of the OLED was enhanced by using a proper microcavity effect. Different researchers have shown that a Fabry–Pérot microcavity effect occurs in a multilayered OLED,[87] thereby enhancing its luminescence. Researchers including Huang et al.,[87] Zhang et al.,[88] and Chen et al.[82] worked on this and found a narrower and much brighter emission spectral width.

External Quantum Efficiency (EQE, η_{ext})

The external quantum efficiency is defined as the ratio of the number of photons emitted by the OLED to the total number of electrons passing through the device. Alternatively, it can be stated as the ratio of the number of photons emitted to the number of electrons injected into the OLED. It can be deduced from the expression

$$\eta_{ext} = n_r \times \phi_f \times \chi \times \eta_{out} = \eta_{int} \times \eta_{out}, \qquad (1)$$

where η_r is the probability of electron–hole recombination to form excitons and ϕ_f is the fluorescent quantum efficiency or the fraction of excitons that decay radiatively. The probability that radiative decay occurs is represented by χ, while η_{int} is the internal quantum efficiency of the device. η_{out} represents the fraction of photons that will exit the device. According to Langevin theory for OSC materials, the EQE depends on the probability that electrons and holes will find each other. Therefore, the EQE will be higher when the concentration of electrons and holes is maximum. Figure 5 shows the improvement in the external quantum efficiency achieved over time, while Table II presents the different EQE values reported for various devices with different architectures and emission colors. This data summary reveals an increasing trend in the EQE over the years, from 2.3% in 2005 to a high value of 18.5% in 2018.[89] It should also be noted that this parameter depends greatly on the emission

Table II. External quantum efficiency (EQE) for different emission colors

Year	Applied bias (V)	Color/Wavelength (nm)	EQE (%)	Ref.
2004	18	Blue (492)	3.9	72
2005	6.1	Deep blue	2.3	91
		Sky blue	8.7	
		White		
2006	NR	NR	3.3	92
2006	13	Yellow	2.6	74
		White	1.9	
2007	9	525	1.72	77
2007	25	Green	8.2	76
2011	NR	434	1.48	93
2012	10	Green (fluorescent OLED)	2.57	70
	8	Green (phosphorescent OLED)	12.81	
2012	NR	NR	19	94
2013	NR	White (single-layered OLED)	6.7	90
		White (fluorescent OLED)	18.4	
		White (tandem OLED)	27	
		White (multilayer OLED)	14.4	
2014	16	408	2.56	71
2014	20	509	5.12	83
2014	10	NR	2	84
2018	18	Red/green (598)	5.10	85
	18	White light (495–577)	0.63	
2018	NR	Green	18.5	89

Table III. Current efficiency as function of applied bias

Year	Applied bias (V)	Current Efficiency (cd/A)	Ref.
2004	18	8.5	72
2005	6.1	12.8,16.2	91
2006	20	22.4	73
2006	18	9.16	75
2007	28	29	76
2008	11.5	2.7	78
2010	27.5	16.4	80
2011	NR	36.5	82
2012	NR	4.97	95
2013	NR	58.5	96
2014	16	10.37	71
2014	20	16.2	83
2018	15	28	97
2018	16	9.14, 10.97	85
2018	NR	59.6	89

color.[90] Green-emitting materials are most well developed and hence exhibit very high EQE values, whereas blue-emitting materials do not show such good results .

Various researchers, including Xie et al.,[89] Chang et al.,[90] and Yun et al.,[70] have reported a reasonable improvement in EQE values, including a maximum EQE of 27%, 18.5%, and 18.4% for tandem white, green, and fluorescent white light, respectively. Xie et al.[89] utilized the dip-coating fabrication technique instead of evaporation and spin coating for the realization of OLED devices. The dip-coated devices exhibited a higher EQE (18.5%) in comparison with the fully evaporated device (17.1%). Similarly, Chang et al.[90] analyzed different OLED structures based on materials emitting white light. These included a single-layer device, a fluorescent white OLED, a tandem OLED, a multiple-emissive-layer white OLED, etc. Each structure showed different external quantum efficiency. In addition, Yun et al.[70] utilized different phosphorescent and fluorescent materials to achieve high device performance. Note that the EQE is greatly affected by the choice of the material, even for the same device architecture.

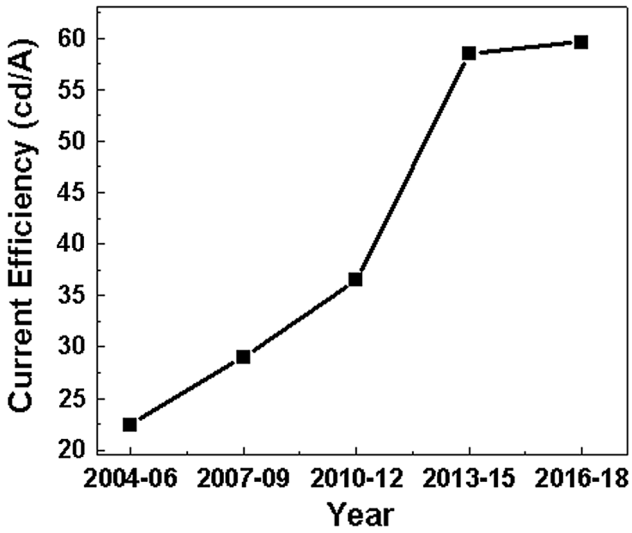


Fig. 6. Increasing trend in current density over time.

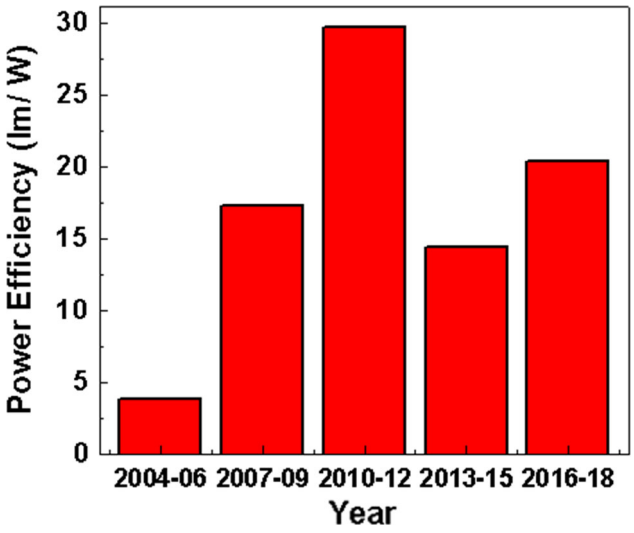


Fig. 7. Improvement in power efficiency.

Table IV. Power efficiency results for different colors

Year	Applied bias (V)	Color/Wavelength (nm)	Power efficiency (lm/W)	Ref.
2004	18	Blue (492)	3.8	72
2006	13	Yellow	2.5	74
		White	1.3	
2006	18	540	3.39	75
2007	25	Green	17.3	76
2010	NR	Blue (470)	6.53	81
2011	11.5	Green (tandem OLED)	21	82
	7.2	Green (single-unit OLED)	10.1	
2012	10	Green (fluorescent OLED)	4.6	70
	8	Green (phosphorescent OLED)	29.7	
2014	16	408	8.69	71
2014	20	460	12.9	83
		460	14.4	
2018	15	NR	20.3	97
2018	18	Red/green (598)	6.11	85
	18	White (495–577)	0.54	
2018	NR	NR	5.9	86

Current Efficiency (η_L)

Usually, the current efficiency is described as the charge transferred in a system. However, for an OLED, it is defined as the ratio of the luminescence (L) to the current density (J). This is yet another important parameter to analyze OLED performance. It is also a measure of how well the human eye can perceive a color, i.e., the photopic response of the human eye. Therefore, it also depends on the color of the light, being maximum for green light as compared with red or blue. The trend in the current efficiency also shows a continuous improvement over time, from the value of 8.5 cd/A reported by Chen et al. in 2004[72] to the maximum value of 59.6 cd/A reported by Xie et al. in 2018.[89] Table III summarizes important results for the current efficiency over the years. The increasing trend in the current efficiency over the years is plotted in Fig. 6 as well.

The most notable works reporting very high current efficiency values are as follows: Ohmori et al.[76] utilized phosphorescent material complexes of [Ir(ppy)$_3$] and [Ir(piq)$_3$] with poly(*N*-vinylcarbazole) (PVCz) and 1,3,5-tris[4-(diphenylamino)phenyl]benzene (TDAPB) as host materials for green and red light emission,[76] respectively. Using these materials, they were able to achieve a luminescence of 3800 cd/m^2 with an EQE of 8.25% and a high current efficiency of 29 cd/A. Chen et al.[82] also reported a good current efficiency of around 36.5 cd/A and a luminescence of 7068 cd/m^2[82] by using an OLED with a tandem structure. Additionally, a bulk heterojunction formed of a ZnPc and C_{60} fullerene blend was incorporated in the middle of the device. This charge generation layer

led to the formation of electrons and holes on applying an appropriate bias voltage, and transferred them to the emission layer.

Similarly, Gao et al.[96] achieved a high current efficiency of 58.5 cd/A for an OLED emitting white light via a combination of blue and yellow. They used a mixed interlayer (MI) in the device architecture. The MI provides an intermediate energy level between the two materials, thus balancing the charge carriers among them. Recently, Xie et al.[89] reported a quite high luminescence of 59.6 cd/A for a green OLED with an equally high EQE of 18.5%[89] by improving the fabrication process. OLEDs processed by dip coating show superior performance over devices fabricated by spin coating and thermal evaporation processes. Furthermore, the possibility of making large-size displays using the dip-coating process is also highlighted.

Power Efficiency (η_P)

The power efficiency is defined as the ratio of the total luminous flux (i.e., the luminous energy per unit time) to the power consumption of the OLED. It is also known as the luminous efficacy. Table IV presents the most important results for OLED power efficiencies reported in recent years, showing a steady improvement, although some of the recent data have been very low. Figure 7 illustrates the improvement in power efficiency over time. The most notable power efficiency was reported by Yun et al.[70] in 2012 for phosphorescent OLEDs. Additionally, Ohmori et al.,[76] Chen et al.,[82] Chen et al.,[83] and Li et al.[97] have also reported good power efficiency results. Ohmori et al.[76] and Chen et al.[82] worked on intermediate layers for enhancing the device performance, achieving good power efficiency values of 17.3 lm/W and 21 lm/W, respectively.

Yun et al.[70] reported an OLED based on phosphorescent materials with an inverted top architecture, including electron transportation and injection layers to enhance the device performance. The results for the device indicated a power efficiency of 29.7 lm/W,[70] in addition to a quite high luminescence of 133,500 cd/m^2. Furthermore, in 2014, Chen et al.[83] worked on novel materials such as luminogens with tetraphenylethene (TPE) to make 4-{2-[4-(Dimesitylboranyl)phenyl]-1,2-diphenylvinyl}-N,N-diphenylaniline (TPE-NB) and 4′-{2-[4′-(Dimesitylboranyl)-(1,1′-biphenyl)-4-yl]-1,2-diphenylvinyl}-N,N-diphenyl-(1,1′-biphenyl)-4-amine (TPE-PNPB) combinations. Their electron donor–acceptor interaction was found to be very strong, resulting in a decrease in the emission efficiency along with a red-shift. However, both luminogen materials showed good power efficiency of 12.9 lm/W and 14.4 lm/W for TPE-NB and TPE-PNPB, respectively.[98]

Li et al.[97] also reported a power efficiency of 20.3 lm/W for a phosphorescent blue OLED (among the highest values for blue OLEDs). In their research, two novel host materials based on C-Si and N-Si were developed. Both the silicon-containing carbazole and arysilane showed good properties, with a power efficiency of 20.3 lm/W along with an equally good current efficiency of 28.7 cd/A and EQE of 15.8% for the OLED.

Lifetime

The lifetime is another major performance parameter for OLEDs and other display devices. Even though there is no standard method for its calibration, it is measured as the mean time taken by the device to reach half its brightness.[66] The performance of OLEDs is not very good in terms of lifetime as compared with conventional LEDs. In fact, this aspect did not receive much attention from the research community, thereby hampering its improvement, but there is great scope to identify techniques that may limit the decay of organic materials over the course of time. Some of the first significant efforts aimed at increase the lifetime of OLEDs were reported by Heck et al.[99] Lifetime of less than 2000 h was reported for blue fluorescent OLEDs, which is very low compared with that of green fluorescent material, for which Weaver et al. reported a lifetime of 20,000 h.[100]

Furthermore, Wen et al.[91] analyzed blue OLED material in 2005 and reported a longest lifetime of 7000 h. The lifetime of OLEDs is affected by a number of factors, including the high device luminescence, operating specifications, and the conditions in which the devices are stored. Geffory et al.[66] also reported the lifetime of different devices depending on their emission color. Red OLEDs showed the longest lifetime of 22,000 h, followed by green at 20,000 h and blue at only 1000 h.[66] These results highlight the importance of the emission color in determining the lifetime of OLEDs. Furthermore, Tsain et al.[79] reported a short lifetime for white OLEDs. Cognizance of the heating process that occurs in OLEDs, it has been demonstrated that, with the help of heat-transmitting ultraviolet (UV)-cured glue, the device lifetime can be significantly improved.[79] Additionally, Lu et al.[81] worked on a *p*-type-doped blue OLED material, i.e., tetrafluorotetracyanoquinodimethane (F4-TCNQ) doped with {4,4′,4″-tris(*N*-(2-naphthyl)-*N*-phenylamino) triphenylamine} (2T-NATA), for the hole injection layer. Overall, the use of these materials has improved the device performance and enhanced the lifetime, reaching values as long as 3800 h.[81]

OLED STRUCTURES

OLEDs are greatly affected by the materials used in the their fabrication as well as the structure/architecture of each device. Unlike transistors, an OLED is a two-terminal diode, and architectural modifications can only be applied somewhere between the anode and cathode. However, some novel architectural modifications have resulted in

substantial performance enhancements.[101] In 1987, Tang et al.[64] first proposed a bilayer OLED consisting of two OSCs (one *p*-type and one *n*-type) between the two electrodes. However, due to the difference between the electron and hole mobility (the latter being higher in OSCs), recombination did not take place at the center where the two OSCs meet but rather inside the OSC with slower-moving charge carriers. Moreover, the high energy barrier between the electrode work function and the OSC orbital[102] is another disadvantage of this structure.

In 1995, Ammermann et al.[102] proposed a multilayered OLED architecture with each OSC layer used for a specific purpose. Charge injection was a major problem with previous structures, hence in multilayered architectures, dedicated layers are added to enhance the injection of electrons and holes. However, according to the application requirements, charge blocking layers for electrons and holes are also added to create a carrier-confinement structure so that recombination can be enhanced within the emission layer. Additionally, charge generation layers and spacer layers[73] can be incorporated into the device architecture to enhance the charge carrier concentration. To further improve the charge injection, a mixed interlayer[96] with different HOMO and LUMO levels can also be utilized. The inclusion of emission layers is also a very effective approach, where a dedicated layer is used to harness a specific luminescence.[85] Over time, novel top-emitting OLED (TOLED)[88,103] and inverted OLED architectures have also been proposed from the circuit point of view. A brief discussion of these aspects is presented below.

Additional Layers in OLED Structure

As discussed above, layers can be added to the OLED structure to enhance the injection of charge carriers. The addition of a hole transport layer (HTL)[104,105] and electron transport layer (ETL)[106] aids recombination, but the energy barrier between the electrodes and these layers makes it difficult to inject sufficient charge carriers into the device.[102] Therefore, a separate hole injection layer (HIL) and electron injection layer (EIL) are chosen for injection and transportation of charge carriers. The purpose of these layers is to reduce the barrier to injection of charge carriers by introducing an intermediate energy level between the highest occupied or lowest unoccupied molecular orbital and the work function of the electrode. Thus, a staircase-type architecture[107–110] is created, directly increasing the charge carrier injection into the OLED.[102]

Furthermore, charge blocking layers[108,109] can be introduced to restrict the movement of charge carriers. Since the hole mobility is higher than the electron mobility in organic materials, holes cross the emission layer at a higher velocity even before the electrons arrive there.[109] This results in (a) less recombination in the emission layer, (b) recombination in a layer other than the desired one, which reduces the device efficiency, and finally (c) carrier quenching at the electrodes.[110] Therefore, to restrict the movement of holes, hole blocking layers can be utilized. Once the holes are thus restricted to near the emission layer, electrons arriving at the emission layer recombine with these holes, thereby improving the device efficiency.[75] Furthermore, the assimilation of holes increases the bias of the device, which helps to inject a greater number of electrons. Table V presents the results of some of the notable work on OLED architectures.

The results presented in this table illustrate the importance of different layers in the architecture of OLEDs. The injection layer, blocking layer, along with the transportation layers[111] all have a significant impact on the performance of OLEDs; for example, adding an HIL of MoO_x improved the power efficiency by 83.7%,[86] and a similarly improvement (22%) was observed by Lu et al.[81] on using the combination of F4-TCNQ and 2T-NATA for the HIL and HTL, respectively. A spacer layer also has a considerable impact on the device performance, as observed when using LiF to separate the ETL and EML.[73] As a consequence, the luminescence efficiency of the device can be improved by sevenfold. In a similar manner, addition of other layers can enhance the characteristic performance of OLEDs. Such layers must be judiciously applied in device architectures, as they also have a direct impact on the dimensions.

Use of Novel Layers

Apart from the use of charge injection, transportation, and blocking layers to improve the performance, novel architectural modifications have also emerged for multilayered OLEDs. Some of the most promising approaches in this regard include the charge generation layer (CGL),[82] spacer layers,[73] mixed interlayer (MI),[96] microcavity structure, etc. Park et al.[112] first performed a numerical analysis on devices including a charge transport control layer (CTCL) and EML layer. It was deduced that, in the HTL/CTCL/ETL layer architecture, high HOMO and LUMO levels are preferable at the HTL/CTCL and CTCL/ETL interfaces,[112] respectively.

Chen et al.[82] made use of a charge generation layer (CGL) based on a heterojunction of ZnPc and C_{60} materials. The inclusion of these layers enhanced the charge generation and charge injection properties. The device showed enhanced performance mainly in terms of the power efficiency. In a similar manner, Gao et al.[96] used a mixed interlayer (MI) between two different emission materials. The MI can be hole or electron predominant, thus helping to balance the charge carrier transport in the device. With the help of an MI, the charge carriers could be balanced between the fluorescent blue and phosphorescent yellow materials, resulting

Table V. Layered architectures and improvement in parameters of OLEDs

Year (research group)	Device architecture	Studied layers	Analysis	Performance improvement
2004 (Chen et al.)[72]	Al (Cathode) LiF (Cathode) Alq_3 (EIL) PPOS (EML & ETL) TPD (HTL) CuPc (Buffer Layer) ITO (Anode)	1,2,3,4,5-Pentaphenyl-1-(8-phenly-1,7-octadiynyl)silole (PPOS) as EML and ETL Alq_3 as EIL and N,N′-Bis(3-methylphenyl)-N,N′-diphenylbenzidine (TPD) as HTL	Increase in thickness of TPD, Alq_3, and PPOS layer enhanced the current efficiency	1. On scaling up TPD thickness from 17 nm to 50 nm, the current efficiency increased from 1.4 cd/A to 7.7 cd/A 2. Maximum current efficiency was 8.5 cd/A when the Alq_3 thickness was 7 nm, but it decreases with decreasing Alq_3 thickness
2005 (Mori et al.)[108]	Al (Cathode) LiF (Cathode) BCP & BAlq (HBL) NPD (HTL & EML) CuPc (HIL) ITO (Anode)	Bathocuproine (BCP) Aluminum(III) bis(2-methyl-8-quinolinate)4-phenylphenolate (BAlq), both as HBL	Comparison between BCP and BAlq for better HBLs. OLED with NPD and BCP showed blue emission	1. With reduction of BAlq thickness and introduction of Alq_3, recombination zone shifted from NPD/BAlq to BAlq/Alq_3 interface 2. Electroluminescence due to BCP not observed in NPD/BCP architecture 3. Energy transfer from BCP to NPD when recombination occurs in BCP 4. BCP is a better HBL as compared with BAlq
2006 (Yang et al.)[75]	Al (Cathode) LiF (Cathode) BAlq & BPhen (HBL) Alq_3 QAD ALQ_3 (EML) NPB (HTL) m-MTDATA (HIL) ITO (Anode)	Aluminum(III) bis(2-methyl-8-quinolinate)4-phenylphenolate (BAlq) 4,7-Diphenyl-1,10-phenanthroline (BPhen), both as HBL	Compared different multilayered devices without and with single/double HBL	1. Device with double HBL showed the highest current efficiency and power efficiency 2. Current efficiency improved by 43% and power efficiency reached 3.39 lm/W, being 1.9 times higher than multilayered OLED 3. Double-HBL device also showed maximum luminescence of 23,750 cd/m^2
2006 (Kim et al.)[73]	Al (Cathode) LiF (Cathode) LiF & Alq_3 (ETL) Alq_3 (EML) NPD (HTL) CuPc (HIL) ITO (Anode)	Lithium fluoride (LiF)	Compared structures without and with LiF to separate the Alq_3 layer as EML and ETL	1. Using the LiF layer (to create a separation between the ETL and EML), the device luminescence efficiency increased by 7-fold 2. Device with LiF separation also showed highest power efficiency

Table V. continued

Year (research group)	Device architecture	Studied layers	Analysis	Performance improvement
2010 (Lu et al.)[81]	Al (Cathode) LiF (Cathode) Alq_3 & PyPySPyPy (1:1200 ratio) (ETL) TBADN & DSA-ph (EML) NPB (HTL) 3% F4-TCNQ doped 2T-NATA & 2T-NATA (HIL) ITO (Anode)	Tetrafluorotetracyanoquinodimethane (F4-TCNQ) 4,4′,4″-tris(*N*-(2-naphthyl)-*N*-phenylamino) triphenylamine (2T-NATA) TBADN{2-t-butyl-9,10-di-2-naphthylanthracene} *p*-Bis(*p*-*N*,*N*-diphenylaminostyryl) benzene (DSA-ph)	Used different combinations of F4-TCNQ and 2T-NATA as HIL and HTL	1. Bilayer of T4-TCNQ (doped with 2T-NATA) and 2T-NATA resulted in the best performance and lifetime of 3800 h, better than for undoped devices 2. Device with HIL of T4-TCNQ doped with 2T-NATA showed improved power efficiency and long operation time; it diffuses into HTL and EML and produces annihilation zones 3. Power efficiency improved by 22% in doped device
2010 (Lim et al.)[111]	Metal (Cathode) (ETL) OSC (EML) NPB (HTL) ITO (Anode)	*N*,*N*′-bis(1-naphthyl)-*N*,*N*′-diphenyl-1,1′-biphenyl-4,4′-diamine (NPB)	Variation in thickness of NPB analyzed	1. The resistance of the device and forbidden energy gap decreased with an increase in NPB thickness
2018 (Zheng et al.)[86]	Al (Cathode) LiF (Cathode) BPhen (HBL) Alq_3 (EML) NPB (HTL) MoO_x (HIL) ITO (Anode)	Molybdenum oxide (MoO_x)	Used MoO_x in solution phase as hole injection material	1. Improved device luminescence and power efficiency by 83.7% and 1.8 times, respectively, as compared with device without HIL of MoO_x 2. Compared with MoO_x deposited by other methods, solution-processed MoO_x showed best performance

in the production of white light. Previously, Park et al.[113] successfully used a spacer layer in their device architecture to separate the two different light-producing entities of the OLED.

Additional Changes in Basic OLED Structure

The basic OLED architecture suffers from many structural disadvantages,[114,115] including emission from the anode side. Owing to this, the driving circuitry can restrict the light or produce a shadow. Furthermore, the emitted light is also absorbed internally by the device. To overcome these shortcomings, some novel architectural modifications have been suggested, including an inverted OLED structure, the use of a microcavity structure, and top-emitting OLEDs. Dobbertin et al.[116] were among the first to demonstrate the fabrication of top-emitting OLEDs. The advantage of this structure is the ease of integration of the device with the driving circuitry.[66] Moreover, the circuitry does not block the emitted light, which is a common problem in OLED-based displays. The only requirement for that particular device was a transparent top electrode. Afterwards, Wu et al.,[77] Huang et al.,[114] Wang et al.,[115] and Zhang et al.,[88] also focused on top-emitting OLED structures. Table VI summarizes the work carried out by those researchers.

Table VI also highlights the importance of the microcavity effect for the luminescence characteristics of the device. Reflective anodes[88] and the ITOLED architecture[115] can be utilized to enhance the microcavity effect to improve the light output from the device. In a similar manner, the DBR effect was also used by Huang et al.[114] to improve the light output from the device; this is achieved by using multilayered electrodes. These additional architectural modifications considerably improved the device characteristics.

Microcavity Effect in OLED Structures

The OSC and electrodes used in OLEDs reflect the light. This reflected light interferes with the light emitted from the OLED, resulting in a Fabry–Pérot resonator[114] that can enhance the light output. The light output depends on the wavelength of the light and the length of the device forming the microcavity.[117,118] Choosing the correct length for the device can result in positive interference, thereby enhancing the luminescence. Researchers have successfully applied this microcavity effect to enhance the device performance. The effect is easier to use in the top-emitting OLED structure, where the bottom electrode is reflective.[119] Wu et al.[77] utilized the microcavity structure with a metallic electrode,[120] leading to a redistribution of the photon density of states (DOS), thus resulting in a certain cavity mode.[77] This led to the formation of narrow-band emission and color selection over a wide wavelength range.

Similarly, Huang et al.[114] investigated the nonzero phase shift at the interface between metallic and dielectric materials. This phase shift affects the top-emitting OLED microcavity and results in a color shift. Lee et al.[121] fabricated a resonant-cavity OLED (RCOLED) with a microcavity structure. Their work resulted in the formation of a distributed Bragg reflector (DBR) with SiO_2 and TiO_2, along with highly reflective Al metal. The results showed four emission peaks at 428 nm, 472 nm, 540 nm, and 612 nm, corresponding to deep-blue, blue, green, and red light, respectively, for white-light generation. The RCOLED exhibited enhanced luminescence, efficiency, color rendering index (CRI), and Commission internationale de l'éclairage (CIE) coordinates due to the better microcavity effect when using the optimal microcavity length. Furthermore, Huh et al.[122] used capping layers (CL) to enhance the microcavity effect in the OLED. In a bottom-emitting OLED, the CL layers helped to create a constructive interference effect, whereas in a top-emitting OLED, the CL affected the transmittance of the device.

Chen et al.[82] included a charge generation layer in a tandem OLED structure, resulting in the microcavity effect. Owing to this effect, the spectral deviation in the viewing angle was observed to be negligible. Zhang et al.[88] used Cu as a reflective anode in a top-emitting OLED, which caused a poor microcavity effect due to the lower reflectance of Cu. This resulted in lower device efficiency. Following a similar methodology, both Huang et al.[114] and Wang et al.[115] tried to modify the microcavity structure in their respective OLEDs. Wang et al.[115] utilized an inverted top-emitting OLED to include the microcavity effect, and Huang et al.[114] used stepped doping in the emission layer to fix the wavelength of the emitted light and narrow the spectral width. Both structures exhibited enhanced light output.

OLED MATERIALS

The performance of organic devices depends on the OSC material employed. In the preceding section, it was observed that different layers can be used to enhance the performance of OLEDs. In addition, different materials show different properties when applied in a particular architecture. Thus, different materials have been investigated for the emission layers. The use of different materials affects the performance of OLEDs due to their inherent properties. The organic materials along with the architecture can be combined to achieve superior devices. This section covers different materials that are used as the emission layers to output different colors of light.

Blue-Emitting Materials

OLEDs based on OSC materials have attracted much attention from researchers due to their

Table VI. Novel OLED architectures and performance improvements

Year (research group)	Structure	Analysis	Performance
2007 (Wu et al.)[77]	Ag (Cathode) Al (Cathode) LiF (Cathode) Alq_3 (EML & ETL) NPB (HTL) m-MTDATA (HIL) Ag (Anode) SiO_2 (Substrate)	Demonstrated optical and electrical properties of top-emitting OLED while varying the thickness of the Ag anode	1. Electroluminescence and reflectivity of the device with thicker anode (device 1) was higher than that with thinner anode (device 2) 2. At 7 V, the luminescence of device 1 was reported as 4078 cd/m^2 compared with 2356 cd/m^2 for device 2 3. Current efficiency of device 1 was also higher as compared with device 2: 5.2 cd/A and 2.7 cd/A respectively
2008 (Huang et al.)[114]	Ag (Cathode) Al (Cathode) LiF (Cathode) Alq_3 (EML & ETL) NPB (HTL) m-MTDATA (HIL) Au (Anode) Ni (Anode) Al (Anode) Glass (Substrate)	Analyzed the effect of multilayer anode (Ni–Au) phase shift on the performance of OLED	1. Anode exhibited periodic Ni-Au behavior similar to distributed Bragg reflector (DBR) 2. Due to presence of multicavity, there was a phase shift at the interface of metals 3. Phase shift effect altered the optical thickness of the OLED, which increased with an increase in Ni-Au pairs
2010 (Wang et al.)[115]	Alq_3 (Capping layer) Ag (Anode) MoO_3 (HIL) NPB (HTL) Alq_3 (EML) BCP (Spacer) Alq_3 (ETL) Ag (Bottom Cathode) LiF (Bottom Cathode) Al (Bottom Cathode)	Illustrated the inverted top-emitting OLED (ITOLED) structure. Anode used as the emitting electrode, and cathode as reflective electrode	1. ITOLED structure has the advantage of strong microcavity effect 2. BCP used as hole and exciton confining layer, thus improving device efficiency 3. Alq_3 capping layer on Ag anode improved the outcoupling efficiency and emission enhancement factor G_{cav} 4. Improved G_{cav} also resulted in high current efficiency and light emission intensity
2011 (Zhang et al.)[88]	Ag (Cathode) Al (Cathode) LiF (Cathode) Alq_3 (EML & ETL) NPB (HTL) MoO_x (HIL) Cu (Reflective Anode) Glass (Substrate)	Cu and MoO_x utilized as reflective anode and HIL, respectively, for top-emitting OLED	1. Device with Cu/MoO_x had enhanced luminous efficiency of 3.6 cd/A compared with 0.11 cd/A for devices without Cu/MoO_x 2. Cu/MoO_x enhanced the work function, which reduced the injection barrier for holes 3. Cu/MoO_x devices showed wide emission spectra with full-width at half-maximum (FWHM) of about 100 nm
2012 (Yun et al.)[70]	ZnS (Anode) Ag (Anode) WO_3 (Anode) NPB (HTL) Alq_3 (for Phosphorescence) & $Ir(ppy)_3$ doped CBP (for Fluorescence) (EML) Alq_3 & BPhen (ETL) CS_2Co_3 (EIL) Al (Cathode) Glass (Substrate)	Examined the effect of ETL on performance of inverted top-emitting OLED Effect of dielectric–metal–dielectric anode structure also investigated	1. Charge injection is dependent on the mobility of the ETL (even for materials with same LUMO levels) 2. The dielectric–metal–dielectric electrode resulted in low reflectance at organic–electrode interface 3. This structure subdued the strong cavity effect, but it was high enough to yield resonance-induced enhancement in luminescence efficiency

capability to produce a wide range of colors. Primarily red, green, and blue light are required for display applications.[123,124] Among these, red- and green-emitting materials are readily available and show good luminescence characteristics; however, blue-emitting materials are not on a par with these.[125] Table VII presents a comprehensive literature survey with regard to the development of blue-emitting OLED materials.

Researchers including Shi et al.,[126] Culligan et al.,[127] Hosokawa et al.,[128] and Shen et al.[129] have experimented with diarylanthracenes, but none have achieved a high-performance blue-emitting material comparable to MADN. Hosokawa et al.[128] reported on the di(styryl)arylene group based on DPVBi, which showed satisfactory performance. Wen et al.[91] further improved the structure of DPVBi to DPVPA to achieve a good-performing blue-emitting material. Wen et al.[91] utilized material engineering techniques to shorten the conjugate length of di(styryl)benzene-based dopant to a monostyrylbenzene core. Moreover, their group demonstrated that hole-blocking layers can be utilized to achieve deep-blue emission, as these layers reduce the recombination zone, consequently resulting in enhanced device efficiency.[130] Triboni et al.,[124] Suzuki et al.,[131] Luo et al.,[132] and Li et al.[97] are other researchers who have contributed significantly to the development of blue-emitting materials.

Red/Orange-Emitting Materials

The preceding section covered different research and development efforts to identify better blue-emitting materials. There are two main reasons for focusing on blue-emitting materials: their poor luminescence characteristics, and the lack of technology for their development. However, the same is not true for materials emitting green and red colors. Furthermore, a complete color display requires good red, green, and blue (RGB) color emissions. The material technology for red and green colors is quite mature, thus materials emitting in the red have not received much attention in recent times. The typical lifetime reported for red-color phosphorescent materials is 22,000 h at a luminescence of 500 cd/m^2.[99] A few details regarding red- and orange-emitting materials are presented in Table VIII.

Yellow-Emitting Materials

The basic colors that the human eye can perceive are red, green, and blue (RGB). It is the combination of these colors that creates other colors. This is also true for display devices, although there are some instances where white light is produced by some other method. The first approach in this regard is the mixing of wide-band yellow- and blue-emitting materials to produce white light.[137–139] Choukri et al.[74] produced white light by varying the position of the yellow emission of rubrene. Consequently, yellow light with CIE coordinates of (0.51, 0.48) was achieved with an external quantum efficiency of 1.3%.[74] By varying the position and thickness of the rubrene layer[140] in NPB and DPVBi, blue–yellow emission occurred, resulting in the formation of bright white light. These types of structure were previously used by other researchers including D'Andrade et al.,[141] Li et al.,[140] and Thompson et al.[142] for the generation of white light.

Akimoto et al.[137] suggested the use of benzodithiophene and triphenylamine copolymers, i.e., P-PBTx, for wide-band yellow and blue emission. Blue emission from the same material was also reported by Nishida et al.[143] According to Akimoto et al.,[137] the monomer of PhBTx exhibited white-light emission consisting of wide yellow and blue bands. They attributed the emission of yellow light to a modified molecular group such as a dimer or excimer formed between BDT moieties. Gao et al.[96] focused on the fabrication of white-light OLEDs by mixing blue and yellow light. A mixed interlayer (MI) was utilized to separate the blue-emitting material 4,4′-bis(9-ethyl-3-carbazovinylene)-1,1′-biphenyl (BCzVBi) and the phosphorescent yellow-emitting material iridium(III) bis(4-phenylthiophene[3,2-*c*]pyridibato-*N*,$C^{2'}$) acetylacetonate (PO-01). This F–I–P–I–F structure, i.e., fluorescent and phosphorescent materials separated by interlayers, was first suggested by Sun et al.[144] Other researchers including Seo et al.[145,146] and Chen et al.[135] also advocated the use of these types of structure.

Green-Emitting Materials

The human eye is most sensitive to green colors, which is why green LEDs were among the first to be developed. Furthermore, the best luminescence performance (133,500 cd/m^2) has also been achieved for green color by utilizing phosphorescent materials. The development of organic green-emitting materials started early and is now in a mature phase. Presently, most research is focused either on circuit applications,[147] or structural methods or the use of supporting materials to improve their performance. However, some novel research related to the development of green OSC materials is continuously progressing, as shown by the brief review presented in Table IX.

Alq_3 and dimethylsulfoxide (DMSO) are among the materials that show the highest luminescence performance for green light. Therefore, researchers have utilized these materials to enhance device performance. Among these, Li et al.[149] and Yim et al.[80] reported good OLED performance. Yim et al. used a poly(3,4-ethylenedioxythiophene):polystyrene sulfonate (PEDOT:PSS)[150–152] anode along with DMSO to obtain a high luminescence value of 6000 cd/m^2. Researchers including Tong et al.[153] and Krujatz et al.[125] have shown that a high EQE of 86% can be achieved with the help of fluorescent dyes. Using these fluorescent dyes, Ohomori et al.[76] reported a high luminescence of 33,000 cd/m^2.

Table VII. Different materials for development of blue-emitting OLEDs

Year (research group)	Material used	Device architecture	Performance parameters	Ref.
2004 (Chen et al.)	1,2,3,4,5-Pentaphenyl-1-(8-phenyl-1,7-octadiynyl) silole (PPOS)	Al (Cathode) LiF (Cathode) Alq_3 (EIL) PPOS (EML & ETL) TPD (HTL) CuPc (Buffer Layer) ITO (Anode)	1. PPOS emitted blue light at 492 nm 2. Highest values of parameters Luminescence—10,460 cd/m^2 Current efficiency—8.5 cd/A External quantum efficiency—3.9% Power efficiency—3.8 lm/W	72
2005 (Wen et al.)	Diarylanthracene group: 2-Methyl-9,10-di(2-napthyl)anthraxcene (MADN) 2-(*t*-Butyl)-9,10-di(2-napthyl)anthraxcene (TBADN)	Al (Cathode) LiF (Cathode) Alq (ETL) MADN (EML) NPB (HTL) CF_x (HIL) ITO (Anode)	1. Both materials show potential for deep-blue OLED application 2. MADN exhibits current efficiency of 1.4 cd/A with CIE of (0.15, 0.10) for blue light 3. TBADN has CIE of (0.13, 0.19), but its efficiency was poor	91
	Di(styryl)arylene group: DPVA (replacing central biphenyl nucleus of DPVBi with diphenyl anthracene).	Al (Cathode) LiF (Cathode) Alq (ETL) TBD doped MADN (EML) NPB (HTL) CF_x (HIL) ITO (Anode)	1. DPVPA has higher quantum efficiency than DPVBi 2. It shows fluorescence of 448 nm with 20 nm green shift 3. External quantum efficiency is 3% more than MADN with CIE of (0.14, 0.17), i.e., slightly greener	
	Tetra(*t*-butyl) perylene (TBP) group: TBD doped in MADN (0.5% v/v)	Al (Cathode) LiF (Cathode) Alq (ETL) TBD doped MADN (EML) NPB (HTL) CF_x (HIL) ITO (Anode)	1. Current efficiency of 3.4 cd/A with CIE of (0.13, 0.20), representing sky-blue color 2. The material did not have functional group, therefore its emission properties cannot be altered	
	Diphenylamino-di(s-tyryl)arylene (DSA-ph) group: Bis(diphenyl)aminostyryl benzene (DSA-ph)	Al (Cathode) LiF (Cathode) Alq (ETL) DSA-Ph doped MADN (or DPVPA) (EML) NPB (HTL) CF_x (HIL) ITO (Anode)	1. Absorption at 410 nm with fluorescence at 458 nm 2. When 3% DSA-ph is doped with MADN, the observed current and power efficiency are 9.7 cd/A and 5.5 lm/W, respectively 3. Expected lifetime of 46,000 h	
2006 (Fischer et al.)	*N*,*N*′-Diethyl-3,3′-bicarbazyl (DEC) 4,4′-bis(2,2′-diphenylvinyl)-1,1′-biphenyl (DPVBi)	Al (Cathode) LiF (Cathode) Alq_3 (ETL) BCP (HBL) DEC & DPVBi (EML) NPB (HTL) CuPc (HIL) ITO (Anode)	1. 2% DEC doped in DPVBi 2. The device shows EQE, current efficiency, and power efficiency of 3.3%, 4.7 cd/A, and 1.3 lm/W with luminescence of 2825 cd/m^2 3. Good deep-blue emission	92

Table VII. continued

Year (research group)	Material used	Device architecture	Performance parameters	Ref.
2011 (Lee et al.)	4-Hydroxy-8-methyl-1,5-naphthyrdine aluminum chelate ($AlmND_3$)	Al (Cathode) LiF (Cathode) $AlmND_3$ (EML) mCP (Buffer Layer) NPB (HTL) ITO (Anode)	1. Material showed ambipolar nature with electron and hole mobility of 5×10^{-5} cm^2/V-s and 1×10^{-5} cm^2/V-s, respectively 2. CIE of material is (0.16, 0.08), i.e., deep-blue emission 3. EQE of device is 1.59% with luminescence of 3730 cd/m^2	93
2014 (Guo et al.)	Star-shaped molecule (SSM) with carbazole and arylamine backbone	Al (Cathode) LiF (Cathode) TPBi (ETL) SSM doped DSA-Ph (EML) TCTA doped NPB (HTL) 2T-NATA (HIL) ITO (Anode)	1. Emission of SSM in wide range of 375 nm to 525 nm with peak at 408 nm 2. Current efficiency of 10.37 cd/A with power efficiency of 8.6 lm/W	71
2015 (Wahyuningrum et al.)	Nitrated *N*,*N*-diphenylamine (DPA)	NR	1. After nitration process, material showed absorption peaks at 337 nm and 406 nm, corresponding to peak emission of normal DPA, i.e., 330 nm 2. Emission peaks of new material at 413 nm and 498 nm, corresponding to blue emission	123
2015 (Triboni et al.)	4-(2-Fenoxi-*p*-xileno) *N*-methyl-1,8-naphthalimide (NPOX)	Al (Cathode) NPOX (EML) PEDOT:PSS ITO (Anode)	1. Emission spectrum of NPOX centered at 435 nm with mobility of 5×10^{-5} cm^2/V-s 2. Reported CIE is (0.211, 0.313), corresponding to blue emission	124
2016 (Krujatz et al.)	2′-7′-Di-*tert*-butyl *N*,*N*-diphenyl-7-(4-(1-phenyl-1*H*-benzo[*d*]imidazole-2-yl)phenyl)-9,9′-spirobi[fluorene]-2-amine (spiro-bifluorene derivatives)	Al (Cathode) Ba (Cathode) TPBi (ETL) Spirobifluorene Derivative PEDOT ITO (Anode)	1. Material emitted blue light at 457 nm 2. EQE of 2.9% with maximum luminescence of 1717 cd/m^2 3. Reported CIE is (0.17, 0.10)	125
2018 (Li et al.)	3,9-Bis(triphenylsilyl)-9*H*-carbazole (SiCzSi) 3,6,9-Tris(triphenylsilyl)-9*H*-carbazole (DSiCzSi)	Al (Cathode) MoO_X/ LiF HOST MoO_X/ LiF ITO (Anode)	1. DSiCzSi has current and power efficiency of 25 cd/A and 16.4 lm/W, respectively, with EQE of 13.7% 2. SiCzSi showed current efficiency of 28.7 cd/A along with power efficiency of 20.3 lm/W. EQE reported to be 15.8%	97

Table VIII. Device architecture and performance for red/orange-emitting materials

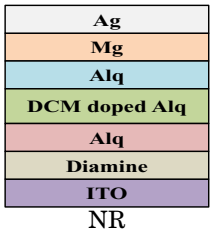
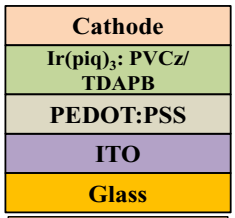
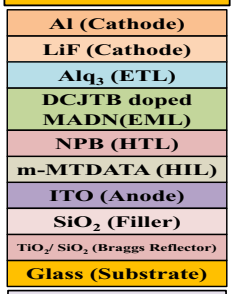
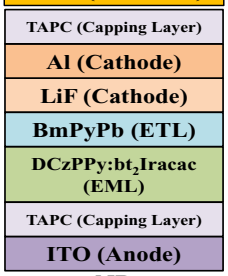
Year (research group)	Material	Device Architecture	Reported Results	Ref.
1989 (Tang et al.)	4-(Dicyanomethylene)-2-methyl-6-(*p*-dimethylaminostyryl)-4*H*-pyran (DCM)-doped Alq_3		1. Emission wavelength observed between 610 nm and 650 nm, corresponding to orange and red color	133
2006 (Ko et al.)	4-(Dicyanomethylene)-2-*t*-butyl-6(1,1,7,7-tetramethyljulolidyl-9-enyl)-4*H*-pyran (DCJTB) and rubrene-doped Alq_3	NR	1. CIE for both materials was (0.64, 0.36) 2. Current efficiencies—2.5 cd/A and 3.2 cd/A 3. Power efficiencies—0.9 lm/W and 1.2 lm/W	134
2007 (Ohmori et al.)	Tris-(1-phenylisoquinoline)iridium [$Ir(piq)_3$]-doped PVCz–TDAPB		1. Reported external quantum efficiency 6.3% 2. Power efficiency 3 lm/W 3. Luminescence 6000 cd/m^2 and 820 cd/m^2 for PVCz and TDAPB devices, respectively	76
2010 (Chen et al.)	4-(Dicyanomethylene)-2-*t*-butyl-6(1,1,7,7-tetramethyljulolidyl-9-enyl)-4*H*-pyran (DCJTB)-doped MADN		1. DCJTB emission for conventional and resonant-cavity OLED at 560 nm and 612 nm, respectively 2. Emission of red/orange light observed Luminescence efficiency—6.32 cd/A Carrier lifetime—1.53 ns	135
2012 (Huh et al.)	2,6-Bis(3-(carbazol-9-yl)phenyl) pyridine (DCzPPy) Bis(2-phenylbenzothiozolato-*N*,$C^{2'}$) iridium(III) (acetylacetonate) (Bt2Iracac)		1. On the top-emitting side, the EQE increased from 1.90% to 3.06% with the use of capping layer 2. In contrast, on the bottom side, the use of capping layer decreased the EQE from 11.55% to 10.10% 3. The intensity of radiance for red light at 565 nm and 605 nm decreased till capping layer thickness of 60 nm, increasing thereafter	122
2015 (Kim et al.)	Quantum dots (QDs)	NR	1. Quantum dots used to enhance the emission intensity for the red color along with absorption of unwanted green and blue light	136
2018 (Zhou et al.)	4-(Dicyanomethylene)-2-*t*-butyl-6(1,1,7,7-tetramethyljulolidyl-9-enyl)-4*H*-pyran (DCJTB)-doped Alq_3	Mixed device architecture to emit white light	1. Device showed Maximum luminescence of 45,869 cd/m^2 Current power of 9.14 cd/A Power efficiency of 6.11 lm/W EQE of 5.10% 2. Electroluminescence varied from 644 nm to 598 nm, resulting in red to red–green emission	85

Table IX. Reported work on development of green-emitting materials for OLEDs

Year (research group)	Material	Device structure	Performance	Ref.
1997 (Shi et al.)	5,12-Dihydro-5,12-dimethylquino[2,3-*b*]acridine-7,14-dione (DMQA)-doped Alq_3	Ag (Cathode) Mg (Cathode) Alq (ETL) Alq_3: DMQA (EML) NPB (HTL) CuPc (HIL) ITO (Anode)	1. Reported CIE of device is (0.39, 0.59) 2. Current efficiency is reported as 7.3 cd/A	148
1997 (Jabbour et al.)	*N*,*N*′-diphenylquinacridone (QAD)-doped Alq_3	AL (Cathode) LiF (Cathode) Alq_3: QAD (EML) TPD (HTL) ITO (Anode) Glass (Substrate)	1. Insertion of Mg into the Al electrode to improve the characteristics 2. EQE is observed to increase from 3% to 3.2% with use of Mg	5
2006 (Li et al.)	Tris(8-hydroxyquinoline) aluminum (Alq_3)	Al (Cathode) LiF (Cathode) BCP (HBL) Alq_3 (EML & ETL) TPD/ NPB (HTL) PVK (CBL) PEDOT:PSS (HIL) ITO (Anode) Glass (Substrate)	1. Alq_3 device emission is in green range with peak at 520 nm and full-width at half-maximum (FWHM) at 90 nm 2. Introduction of the hole blocking layer increased the device luminescence twofold	149
2007 (Ohmori et al.)	Fac-tris(2-phenylpyridine) iridium ($Ir(ppy)_3$)-doped 1,3,5-tris[4-(diphenylamino)phenyl] benzene (TDAPB)	Cathode $Ir(ppy)_3$ doped TDAPB (EML) PEDOT:PSS (HTL) ITO (Anode) Glass (Substrate)	1. The device exhibited External quantum efficiency—8.2% Power efficiency—17.3 lm/W Maximum luminescence—33,000 cd/m^2	76
2010 (Yim et al.)	Dimethylsulfoxide (DMSO)-doped PEDOT:PSS as anode and $Ir(ppy)_3$-doped TPD as emission layer	Al (Cathode) LiF (Cathode) $Ir(ppy)_3$:TPD(EM) DMSO:PEDOT:PSS (Anode) PES (Substrate)	1. Device showed low turn-on voltage of 4.5 V 2. Highest luminescence—6000 cd/m^2 (Bias voltage—27.5 V) 3. Maximum current efficiency—16.4 cd/A (At 8.5 V)	80
2011 (Chen et al.)	Zinc phthalocyanine (ZnPc):fullerene (C_{60}) as CGL Alq_3 as EML	Al (Cathode) LiF (Cathode) NPB MoO_3 $ZnPc:C_{60}$ LiF Alq_3 LiF ITO	1. The modified tandem device yielded power efficiency twice (21 lm/W) that of the conventional OLED 2. They also reported twofold improvement in luminescence and current efficiency	82

Table IX. continued

Year (research group)	Material	Device structure	Performance	Ref.
2016 (Krujatz et al.)	Bis(2,6-diphenylpyrimidinato)iridium(III) picolinate [$Ir(ppm)_2(pic)$] Bis(2,6-diphenylpyrimidinato)iridium(III) 5-(2′-pyridyl)-3-trifluoromethyl-1,2,4-triazolate [$Ir(ppm)_2(taz)$]	Ag (Cathode) Ca (Cathode) TPBi $Ir(ppm)_2(pic)$ or $Ir(ppm)_2(taz)$ PEDOT:PSS ITO	1. $Ir(ppm)_2(pic)$: The emission wavelength was 549 nm, with EQE of 83% 2. $Ir(ppm)_2(taz)$: The emission wavelength was 523 nm, with reported EQE of 86%	125

FABRICATION METHODOLOGIES

The various fabrication processes used for realizing a device play an important role in determining its performance. Even a minute error can result in complete alteration of the device performance. However, at the present time, fabrication processes are standardized to such a high precision level that yields are very high. Fabrication of a conventional transistor follows a very complex flow which includes photolithography,[154] masking,[155] thermal evaporation,[156] ion implantation and diffusion. In contrast, solution processes such as spin coating, the Langmuir–Blodgett (LB) process,[157,158] dip coating,[159] inkjet printing,[98] polymer inking, stamping,[147] transfer printing,[160] etc. are used to fabricate OLEDs. These techniques are usually employed at room temperature and thus do not require huge inventory. High-temperature processes may destroy the atomic structure of OSCs and are thus avoided.

The basic fabrication process for an OLED starts with selecting a substrate or electrode. For a conventional OLED, generally an ITO electrode is preferred, but novel electrode materials such as nontoxic aluminum-doped zinc oxide (AZO)/Ag/AZO[95] or Al have also been utilized. In case of a flexible electrode (PEDOT:PSS), glass or plastic is utilized to provide the mechanical support. The first fabrication process includes cleaning of the substrate/electrode to remove any contamination from the surface, as illustrated in the flowchart shown in Fig. 8. This includes a set of cleaning processes such as dipping in acetone and isopropanol, treatment in an ultrasonic bath, followed by oxygen plasma treatment[161] and drying using a nitrogen gun. Afterwards, the electrode is deposited by sputtering[149] (ITO), magnetic sputtering (AZO/Ag/AZO),[95,162,163] and thermal evaporation (metallic electrodes, Al or Cu). Thermal deposition techniques are carried out in a thermal evaporation chamber at high pressure.[70,164]

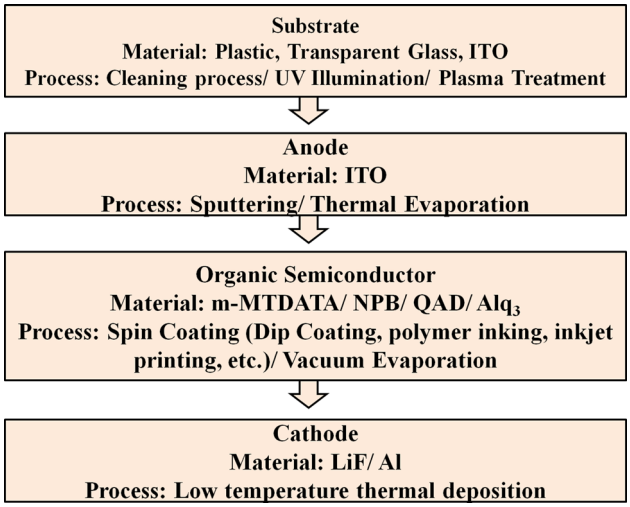


Fig. 8. Basic fabrication flow of bottom-emitting OLED.

The OSC is deposited using either thermal deposition or solution processes.[1] Thermal evaporation techniques are utilized to deposit small-molecule materials, as their structure is not affected by high temperatures[165–167] and solvents for these materials are not readily available.[1] Solution processes are employed for conducting polymers, as they dissolve easily in organic solvents. Solution processes are carried out at low temperature and are very cost effective, thereby invariably being utilized for fabrication of large-scale flexible devices and displays. Finally, the top electrode is deposited with the help of a low-temperature high-pressure thermal evaporation process that does not damage the underlying OSC layer.

To achieve OLEDs with better performance, researchers have tried to enhance and develop new fabrication processes over the last decade. Jenfeng et al.[168] deposited PEDOT nanoparticles as a hole injection layer by using the LB technique.[157] Their analysis revealed that the LB process deposited PEDOT particles with much smaller

size (20 nm to 40 nm),[168,169] thus their device showed a higher current density than that fabricated by conventional methods. Keum et al.[170] developed a blading technique[171,172] specifically for liquid-crystalline OSC[171,173,174] to adjust the molecular orientation. Along with this, an annealing process was also utilized.[170,175] When adopting these techniques, the efficiency of the device was significantly increased owing to the proper alignment of the liquid-crystal OSC molecules.

Recently, Zheng et al.[86] compared the performance of two different OLEDs based on HIL deposition using a low-cost solution process[176–178] and an evaporation process.[86] In the former case, the deposition of MoO_x (as the HIL) was performed by spin coating (at 4000 rpm for 60 s) followed by an annealing process. This device exhibited current and power efficiencies of 7.9 cd/A and 5.9 lm/W, respectively, being 43.6% and 73.5% higher than for the device fabricated using the vacuum evaporation process. Furthermore, Xie et al.[89] utilized dip coating[179–181] to deposit the transport and emission layers. By controlling the dipping time, speed, and concentration of solution, they achieved a current efficiency of 4.6 cd/A and power efficiency of 2.6 lm/W with dip-coated layers. All these results illustrate the impact of fabrication methodologies on the device performance.

OLED MODELS

Mathematical models are often used for predicting the internal device physics and understanding device operation.[182,183] This helps to analyze the device performance under different conditions.[184] Various researchers have developed the mathematical models that are essentially required to understand the physics behind carrier injection and transportation within different layers, thus leading to the innovation of novel device architectures.

Numerical Analysis Based on Poisson's Equation

Poisson's equation is among the most important equations used by mathematicians and physicists.[184] Malliaras et al.[185] applied this equation to calculate the spatial charge distribution inside an OLED and investigated the role of injecting the charge carriers along with their mobility on the device performance. The results illustrated a requirement for equal mobility for electrons and holes, which results in an enhanced recombination rate. Chan et al.[186] also adopted Poisson's equation with an additional optical model to analyze the microcavity effect and determine the optical thickness.[186]

Park et al.[112,113,187] utilized Poisson's equation along with drift–diffusion and singlet exciton rate equations to analyze the internal physics of the device.[184] Their group performed numerous numerical analyses on OLEDs based on this set of equations and investigated the temperature dependence[188] of OLEDs and their different energy loss mechanisms.[113] This analysis revealed that the loss phenomenon occurs based on the absorption of photons emitted from thin quantum wells in conventional LEDs, whereas in organic devices, this loss occurs due to energy transfer from the layer with a higher transition energy.

OLED Model Based on Atomistic Simulation

Work based on atomistic simulation was carried out by Nagata et al.[189] in 2008, wherein the electrostatic potential charges are used to calculate the site energies. Additionally, the Miller–Abraham equation is employed to calculate the charge mobility. This analysis was undertaken to better understand the Poole–Frenkel (PF) mobility model. The analysis began with Gill's law[189] followed by the Miller–Abraham-type hopping rate. Thereafter, the calculation of the mobility μ is completed with the help of Marcus theory. The results of their analysis confirmed the Poole–Frenkel-type behavior of the charge carrier mobility because of the electric field. Four key issues are addressed in their analysis:

1. The site energy distribution is much wider for an amorphous material, and it governs the PF behavior of the charge mobility;
2. The spatially correlated site energy is not the dominant mechanism underlying the PF behavior;
3. To explain the PF behavior, it is necessary to have random mesh connectivity in a three-dimensional amorphous structure;
4. The charge carrier density is dependent on the charge carrier mobility and it results in an enhancement of the PF dependence of the mobility versus the electric field in the space-charge-limited current.

Modeling of Thermal Properties of OLEDs

Bergemann et al.[94] used the transmission matrix approach to model the effect of different thermal effects such as convection, conduction, and radiation on the OLED temperature. This approach uses the Laplace transform for the heat transfer equations. The layers are treated as the product of the transmission matrices for several films (in series), whereas the incident heat flux is assumed to be fragmented between the two isolated independent channels with no flow between them (in parallel). These two channels, parallel and series, are then combined to model the heat transfer.

Numerical Analysis of OLEDs Based on Maxwell's Equations

Pflumm et al.[190,191] developed a model driven by the current for OLEDs in the context of the drift–

diffusion model. The process governing the formation of excitons is assumed to be Langevin. A method for the calculation of the electric fields based on Maxwell's equations is described. This method is helpful, as it differentiates between the current and the voltage excitation of the OLED. The current and voltage behavior of the OLEDs can be modeled using both methods, although the method is much quicker when the current is specified.

The above-mentioned models completely predict the thermal profile of the OLED accurately, thus providing an understanding of the various factors that determine the device operating temperature. Such models can also be utilized for numerical analysis of OLEDs consisting of different layers, viz. emission layers and charge transport control layers (CTCL), similar to the work of Park et al.[112] Moreover, these affect the analysis of the energy loss mechanisms in different device structures, thereby enhancing the architectural aspects.

OLED APPLICATIONS

Currently, OLEDs are at the apogee of OSC-based display technology. These devices offer numerous advantages which can be used in a range of applications. The most well-known applications of OLEDs include display and sensor technology. There are some reported cases of use of OLEDs in VLC links as well. These main applications of OLEDs are reviewed in this section.

OLED-Based Displays

Displays are one of the major applications based on OLEDs and have seen considerable improvement over the past decade. Technology giants such as LG and Apple are incorporating OLED-based displays into their devices. The main reason for this shifting trend from inorganic LEDs to OLED is the impeccable color properties and contrast that OLED-based displays offer. OLED displays are also better in that they do not require any backlighting,[1] resulting in better contrast. Further, they can generate a wide variety of colors, resulting in enhanced image output. Present research is focused on enhancing their efficiency, as the OLED is the component that consumes most power in any circuit. Shin et al.[192] employed dynamic voltage scaling, as there is no particular power saving methodology which incorporates a minimum change in OLED color. The proposed method suggests scaling down the voltage supply, which results in a reduced power requirement due to the lower voltage drop across the driver transistor. The overall power saving is reported to be 52.5% in comparison with other methods.[192]

Similarly, Wee et al.[193] adopted a method known as saliency to reduce the power consumption of OLEDs. The method depends upon the amount that a particular image stands out from a particular scene. The power is reduced in areas which are not in direct eye contact. The overall approach, first used by El-Nasr et al.,[194] is to differentiate between areas of direct and indirect focus. The approach saved 11% of the display power and 4% of the device power.[193] Further, Lin et al.[195] enhanced the technique for scaling of image pixels. By scaling down the pixel values, it gives the user flexibility to scale even irregular shapes. In their method (CURA), saliency was also incorporated. Any type of display device is swept continuously by refresh current signals numerous times during operation. These signals cause a lot of stress on the OLED and result in shorter lifetime and poor performance. In this regard, Liu et al.[196] analyzed the impact of different types of applied voltages on OLED performance.

This is a follow-up of the work proposed by Jacobs et al.[197] and Buso et al.[198] In their work, it was observed that the amplitude modulation technique for direct current (DC) can help to dim an OLED effectively, as long as multiple emitters are not involved. Lin et al.[199] proposed a new buck driver for OLEDs, as the dimming response of the device is strongly influenced by its own heavy parasitic capacitance. The buck driver circuitry also includes a variable-frequency control mechanism. This resulted in a reduction of the output voltage rise time from 25 μs to 10 μs with a corresponding reduction in the switching frequency from 100 kHz to 48 kHz. In a more recent article, Fan et al.,[200] presented an active-matrix (AM)OLED display driven by polysilicon TFTs. A high display operating speed can be achieved using the presented scheme. It was observed that the OLED display has a high frame rate of 240 kHz for an emission driving scheme and the input data period must be 3.5 μs.

OLEDs as Sensors

Sensing is another potential application of OLED-based devices. Various researchers are working on light-based detection, as almost every element in the Periodic Table interacts with light. These particles absorb a light of a particular wavelength to reach a higher excited state, then release this energy when returning to the ground state.[201] The wavelength of the light absorbed by a particular element is fixed (based on its bandgap). Furthermore, OLEDs are known to emit a wide range of colors, implying that light with two colors of similar wavelengths can easily be differentiated. This property can be precisely controlled and is therefore used to differentiate between various elements more accurately.

Miyamoto et al.[202] utilized OLED-based displays for chemical image sensing. The sensor is based on the principle of a light addressable potentiometric sensor (LAPS). Werner et al.[203] further worked on an OLED-based LAPS sensor with a much better response time. The driving methodology suggested by them resulted in modulation frequencies in the

range from 1 kHz to 16 kHz, which directly enhanced the sensing speed by 40 times compared with previously reported devices. Thereafter, many research articles reported the use of OLEDs as photoluminescence sensors, biosensors, and absorption and transmission sensors.[99] Shinar et al.[204] described gas sensors based on the photoluminescence properties of OLEDs. Their method used Ru complexes that are sensitive to oxygen, thus the luminescence intensity of the OLEDs becomes dependent upon the oxygen concentration.

The advantage of using OLEDs in sensor applications is that they provide a noninvasive method.[125,205] Sensors based on the emission from OLEDs with organic photodiodes (OPDs) for detection have also been reported.[206] These sensors can be tuned to become sensitive to different elements or compounds such as oxygen, ammonia, or carbon dioxide. A novel approach in this regard suggests the replacement of the OPD with an OLED for the detection of the light[201] as well. The OLED can thus be utilized as both the light source and detector for primary screening of ovarian cancer.[201] Temperature can also be sensed[205] utilizing OLEDs.

Krujats et al.[125] used the DBR effect to control the light emission from a top-emitting OLED. Based on the excitons from the OLED and the emission of light from the particles to be detected, OLED-based sensors can be utilized in many different ways. Only a small number of biosensors using the transmission and scattering effect have been reported. Prabowo et al.[207] detected anti-mouse immunoglobulin G with the help of an OLED integrated with a surface plasmon resonance device. Biocatalysis is another approach in which the photoluminescence of OLEDs can act on co-embedded enzymes or catalysts. Choudhury et al.[208] presented this approach for a glucose biosensor based on oxygen-sensitive dyes such as Ru or Pt along with glucose oxidase (GOx) in a solution gel sensor film.

OLEDs in VLC Devices

VLC relies on sending information in the form of light due to its highest speed. OLEDs can be used as sources for such devices. Unfortunately, work regarding the use of OLEDs as light sources for VLC is not widely reported. To date, the VLC field has been dominated by conventional LEDs, as they offer high optical power and wide bandwidth.[209] However, at the same time, being a point source, these devices may be harmful and can cause damage to human eyes.[65] Therefore, OLEDs are much more suitable for this particular application owing to their low cost and flexible nature.

Haigh et al.[210,211] first reported an OLED-based VLC device, wherein OLEDs and photodetectors were utilized to create a 1.4-Mbps link using discrete multitone modulation. Their group also analyzed pulse modulation schemes including pulse position modulation (PPM) and on–off keying (OOK). Their analysis revealed that, when using the OOK modulation scheme, OLED-based VLC was limited to a data rate of 550 kbps. Thereafter, their work supported the use of the PPM scheme with the inclusion of an improved driving scheme. After this, equalization of the signals was also performed, resulting in a VLC link with speed of 2.7 Mbps,[65] representing a twofold improvement in speed compared with another defined scheme.

CONCLUSIONS

This exhaustive review covers various aspects of OLEDs, such as performance parameters, device architectures, different materials, fabrication techniques, and applications. Scientific and research advancements related to these aspects over the past two decades are covered in depth, and their impact on OLED performance discussed. A number of challenges have been addressed, and researchers have made continuous progress towards successful commercialization of OLEDs. These include enhancement of the device architecture, which is instrumental for improving their performance characteristics. The development of novel materials has also played an important role in achieving excellent luminescent characteristics. Display-based applications have already been developed and show huge commercial potential.

There are still numerous aspects that could be improved and should become research focuses in the near future. These include the development of application-specific OLED materials with a focus on novel sensors and VLC devices. Material properties should also be improved, bearing in mind the domain-specific requirements. Further, the lifetime of the developed materials and devices should become a center of focus. In this regard, device packaging methodologies should be improved to prevent degradation of OLEDs under various atmospheric conditions. This will lead to greatly improved and enhanced lifetimes. OLEDs are multifaceted devices, so additional novel, contemporary, state-of-the-art applications should be investigated. These may lie in the fields of sensors and VLC devices. Novel sensors could be developed to act as both the light source and detector, as well. Additionally, organic devices should be complemented with silicon-based devices. This will lead to a boost in the present technology with a focus on combining the advantages of both approaches, which may result in highly superior devices.

REFERENCES

1. B. Kumar, B.K. Kaushik, and Y.S. Negi, *Polym. Rev.* 54, 33 (2014).
2. X. Fan, J. Chen, J. Yang, P. Bai, Z. Li, and Z.L. Wang, *ACS Nano* 9, 4236 (2015).
3. J. Yang, J. Chen, Y. Liu, W. Yang, Y. Su, and Z.L. Wang, *ACS Nano* 8, 2649 (2014).

4. Z. Lin, J. Chen, and J. Yang, *J. Nanomater.* 2016, 5651613 (2016).
5. G.E. Jabbour, Y. Kawabe, S.E. Shaheen, J.F. Wang, M.M. Morrell, B. Kippelen, and N. Peyghambarian, *Appl. Phys. Lett.* 71, 1762 (1997).
6. B. Kumar, B.K. Kaushik, and Y.S. Negi, *Microelectron. Reliab.* 54, 2801 (2014).
7. J. Chen, G. Zhu, J. Yang, Q. Jing, P. Bai, W. Yang, X. Qi, Y. Su, and Z.L. Wang, *ACS Nano* 9, 105 (2015).
8. W. Yang, J. Chen, X. Wen, Q. Jing, J. Yang, Y. Su, G. Zhu, W. Wu, and Z.L. Wang, *ACS Appl. Mater. Interfaces* 6, 7479 (2014).
9. X. Zheng, B. Chen, C. Wu, and S. Priya, *Nano Energy* 17, 269 (2015).
10. Y. Wu, Q. Jing, J. Chen, P. Bai, J. Bai, G. Zhu, Y. Su, and Z.L. Wang, *Adv. Funct. Mater.* 25, 2166 (2015).
11. C. Yan, Y. Gao, S. Zhao, S. Zhang, Y. Zhou, W. Deng, Z. Li, G. Jiang, L. Jin, G. Tian, and T. Yang, *Nano Energy* 67, 104235 (2020).
12. G. Zhu, B. Peng, J. Chen, Q. Jing, and Z.L. Wang, *Nano Energy* 14, 126 (2015).
13. A. Dodabalapur, *Solid State Commun.* 102, 259 (1997).
14. J. Chen and Z.L. Wang, *Joule* 1, 480 (2017).
15. G. Zhu, P. Bai, J. Chen, and Z.L. Wang, *Nano Energy* 2, 688 (2013).
16. T.C. Hou, Y. Yang, H. Zhang, J. Chen, L.J. Chen, and Z.L. Wang, *Nano Energy* 2, 856 (2013).
17. N. Zhang, C. Tao, X. Fan, and J. Chen, *J. Mater. Res.* 32, 1628 (2017).
18. J. Chen, G. Zhu, W. Yang, Q. Jing, P. Bai, Y. Yang, and Z.L. Wang, *Adv. Mater.* 25, 6094 (2013).
19. F. Yi, L. Lin, S. Niu, P.K. Yang, Z. Wang, J. Chen, Y. Zhou, Y. Zi, J. Wang, Q. Liao, and Y. Zhang, *Adv. Funct. Mater.* 25, 3688 (2015).
20. Z. Lin, J. Yang, X. Li, Y. Wu, W. Wei, J. Liu, J. Chen, and J. Yang, *Adv. Funct. Mater.* 28, 1704112 (2018).
21. C. Yan, W. Deng, L. Jin, T. Yang, Z. Wang, X. Chu, H. Su, J. Chen, and W. Yang, *ACS Appl. Mater. Interfaces* 10, 41070 (2018).
22. K. Meng, J. Chen, X. Li, Y. Wu, W. Fan, Z. Zhou, Q. He, X. Wang, X. Fan, Y. Zhang, and J. Yang, *Adv. Funct. Mater.* 29, 1806388 (2019).
23. N. Zhang, J. Chen, Y. Huang, W. Guo, J. Yang, J. Du, X. Fan, and C. Tao, *Adv. Mater.* 28, 263 (2016).
24. J. Chen, Y. Huang, N. Zhang, H. Zou, R. Liu, C. Tao, X. Fan, and Z.L. Wang, *Nat. Energy* 1, 1 (2016).
25. Y. Peng, J. Chen, A.Y. Song, P.B. Catrysse, P.C. Hsu, L. Cai, B. Liu, Y. Zhu, G. Zhou, D.S. Wu, and H.R. Lee, *Nat. Sustain.* 1, 105 (2018).
26. L. Cai, A.Y. Song, W. Li, P.C. Hsu, D. Lin, P.B. Catrysse, Y. Liu, Y. Peng, J. Chen, H. Wang, and J. Xu, *Adv. Mater.* 30, 1802152 (2018).
27. L. Cai, A.Y. Song, P. Wu, P.C. Hsu, Y. Peng, J. Chen, C. Liu, P.B. Catrysse, Y. Liu, A. Yang, and C. Zhou, *Nat. Commun.* 8, 1 (2017).
28. B. Kumar, B.K. Kaushik, and Y.S. Negi, *J. Mater. Sci. Mater. Electron.* 25, 1 (2014).
29. W. Yang, J. Chen, Q. Jing, J. Yang, X. Wen, Y. Su, G. Zhu, P. Bai, and Z.L. Wang, *Adv. Funct. Mater.* 24, 4090 (2014).
30. Q. Jing, G. Zhu, P. Bai, Y. Xie, J. Chen, R.P. Han, and Z.L. Wang, *ACS Nano* 8, 3836 (2014).
31. L. Jin, J. Chen, B. Zhang, W. Deng, L. Zhang, H. Zhang, X. Huang, M. Zhu, W. Yang, and Z.L. Wang, *ACS Nano* 10, 7874 (2016).
32. S.Y. Kuang, J. Chen, X.B. Cheng, G. Zhu, and Z.L. Wang, *Nano Energy* 17, 10 (2015).
33. B. Zhang, J. Chen, L. Jin, W. Deng, L. Zhang, H. Zhang, M. Zhu, W. Yang, and Z.L. Wang, *ACS Nano* 10, 6241 (2016).
34. J. Chen, J. Yang, Z. Li, X. Fan, Y. Zi, Q. Jing, H. Guo, Z. Wen, K.C. Pardel, S. Niu, and Z.L. Wang, *ACS Nano* 9, 3324 (2015).
35. W. Yang, J. Chen, G. Zhu, X. Wen, P. Bai, Y. Su, Y. Lin, and Z. Wang, *Nano Res.* 6, 880 (2013).
36. H. Zhang, Y. Yang, Y. Su, J. Chen, K. Adams, S. Lee, C. Hu, and Z.L. Wang, *Adv. Funct. Mater.* 24, 1401 (2014).
37. X. Feng, V. Marcon, W. Pisula, M.R. Hansen, J. Kirkpatrick, F. Grozema, and K. Müllen, *Nat. Mater.* 8, 421 (2009).
38. Z.L. Wang, J. Chen, and L. Lin, *Energy Environ. Sci.* 8, 2250 (2015).
39. W. Liu, J. Chen, Z. Chen, K. Liu, G. Zhou, Y. Sun, M.S. Song, Z. Bao, and Y. Cui, *Adv. Energy Mater.* 7, 1701076 (2017).
40. K. Liu, B. Kong, W. Liu, Y. Sun, M.S. Song, J. Chen, Y. Liu, D. Lin, A. Pei, and Y. Cui, *Joule* 2, 1857 (2018).
41. J. Wan, J. Xie, X. Kong, Z. Liu, K. Liu, F. Shi, A. Pei, H. Chen, W. Chen, J. Chen, and X. Zhang, *Nat. Nanotechnol.* 14, 705 (2019).
42. D. Gupta, M. Katiyar, and D. Gupta, *Org. Electron.* 10, 775 (2009).
43. B. Kumar, B.K. Kaushik, and Y.S. Negi, *J. Vac. Sci. Technol. B Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* 13, 012401 (2013).
44. S.M. Xu, X. Liang, X.Y. Wu, S.L. Zhao, J. Chen, K.X. Wang, and J.S. Chen, *Nat. Commun.* 10, 1 (2019).
45. J. Yang, J. Chen, Y. Su, Q. Jing, Z. Li, F. Yi, X. Wen, Z. Wang, and Z.L. Wang, *Adv. Mater.* 2, 1316 (2015).
46. P. Bai, G. Zhu, Q. Jing, J. Yang, J. Chen, Y. Su, J. Ma, G. Zhang, and Z.L. Wang, *Adv. Funct. Mater.* 24, 5807 (2014).
47. Z. Lin, J. Chen, X. Li, Z. Zhou, K. Meng, W. Wei, J. Yang, and Z.L. Wang, *ACS Nano* 11, 8830 (2017).
48. D. Li and L.J. Guo, *J. Phys. D Appl. Phys.* 41, 105115-1 (2008).
49. D. Threm, J.L. Gugat, A. Pradana, M. Rädler, J. Mikat, and M. Gerken, *IEEE Photonics Technol. Lett.* 24, 912 (2012).
50. P.D. Yang, *Nature* 425, 243 (2013).
51. B.K. Kaushik, B. Kumar, S. Prajapati, and P. Mittal, Organic Thin-Film Transistor Applications: Materials to Circuits (CRC Press, 2016).
52. B. Ruhstaller, T. Beierlein, H. Riel, S. Karg, J.C. Scott, and W. Riess, *IEEE J. Sel. Top. Quantum Electron.* 9, 723 (2003).
53. B. Ruhstaller, S.A. Carter, S. Barth, H. Riel, W. Riess, and J.C. Scott, *J. Appl. Phys.* 89, 4575 (2001).
54. B. Lamprecht, T. Abel, E. Kraker, A. Haase, C. Konrad, M. Tscherner, and T. Mayr, *Phys. Status Solid-R* 4, 157 (2010).
55. P. Mittal, Y.S. Negi, and R.K. Singh, *J. Semicond.* 35, 124002 (2014).
56. P. Bai, G. Zhu, Z.H. Lin, Q. Jing, J. Chen, G. Zhang, J. Ma, and Z.L. Wang, *ACS Nano* 7, 3713 (2013).
57. W. Yang, J. Chen, G. Zhu, J. Yang, P. Bai, Y. Su, Q. Jing, X. Cao, and Z.L. Wang, *ACS Nano* 7, 11317 (2013).
58. J. Yang, J. Chen, Y. Yang, H. Zhang, W. Yang, P. Bai, Y. Su, and Z.L. Wang, *Adv. Energy Mater.* 4, 1301322 (2014).
59. K. Meng, S. Zhao, Y. Zhou, Y. Wu, S. Zhang, Q. He, X. Wang, Z. Zhou, W. Fan, X. Tan, and J. Yang, *Matter* (2020). https://doi.org/10.1016/j.matt.2019.12.025.
60. H. Shirakawa, E.J. Louis, A.G. MacDiarmid, C.K. Chiang, and A.J. Heeger, *J. Chem. Soc. Chem. Commun.* 16, 578 (1977).
61. A. Tsumura, H. Koezuka, and T. Ando, *Appl. Phys. Lett.* 49, 1210 (1986).
62. J.H. Schon and B. Batlogg, *J. Appl. Phys.* 89, 336 (2001).
63. C. Liu, X. Liu, T. Minari, M. Kanehara, and Y.Y. Noh, *J. Inf. Display* 19, 71 (2018).
64. C.W. Tang and S.A. Vanslyke, *Appl. Phys. Lett.* 51, 913 (1987).
65. P.A. Haigh, Z. Ghassemlooy, I. Papakonstantinou, and H. Le Minh, *IEEE Photonics Technol. Lett.* 25, 1687 (2013).
66. B. Geffroy, P. Le Roy, and C. Prat, *Polym. Int.* 55, 572 (2006).
67. B. Kumar, B.K. Kaushik, Y.S. Negi, and V. Goswami, *Microelectron. Reliab.* 54, 100 (2014).

68. G.G. Malliaras and J.C. Scott, *J. Appl. Phys.* 85, 7426 (1999).
69. K. Müllen and U. Scherf, John Wiley & Sons (2006).
70. C. Yun, H. Cho, T.W. Koh, J.H. Kim, J.W. Kim, Y. Park, and S. Yoo, *IEEE Trans. Electron Devices* 59, 159 (2012).
71. K. Guo, J. Zhang, T. Xu, X. Gao, and B. Wei, *J. Display Technol.* 10, 642 (2014).
72. H. Chen, J. Chen, C. Qiu, B.Z. Tang, M. Wong, and H.S. Kwok, *IEEE J. Sel. Top. Quantum Electron.* 10, 10 (2004).
73. Y.M. Kim, J.W. Lee, J.H. Jung, K.K. Paek, M.Y. Sung, J.K. Kim, and B.K. Ju, *IEEE Electron Device Lett.* 27, 558 (2006).
74. H. Choukri, A. Fischer, S. Forget, S. Chenais, T. Billeton, M.C. Castex, and P. Maisse, in *IECON 2006-32nd IEEE Annual Conference on Industrial Electronics* (2006), p. 4865.
75. H. Yang, Y. Zhao, J. Hou, and S. Liu, *Microelectron. J.* 37, 1271 (2006).
76. Y. Ohmori, H. Kajii, and Y. Hino, *J. Display Technol.* 3, 238 (2007).
77. Z. Wu, H. Guo, and J. Wang, *Microelectron. J.* 38, 686 (2007).
78. H. Tetsuka, T. Ebina, T. Tsunoda, H. Nanjo, and F. Mizukami, *Jpn. J. Appl. Phys.* 47, 1894 (2008).
79. Y.S. Tsai, S.H. Wang, F.S. Juang, S.W. Chang, C.H. Chen, M.H. Chung, and T.C. Liao, *J. Display Technol.* 6, 279 (2010).
80. Y. Yim, J. Park, and B. Park, *J. Display Technol.* 6, 252 (2010).
81. L. Lu, J. Yu, L. Long, F. Yu, J. Zhang, H. Zhang, and B. Wei, *Phys. Status Solidi (a)* 208, 2321 (2011).
82. Y. Chen, J. Chen, D. Ma, D. Yan, L. Wang, and F. Zhu, *Appl. Phys. Lett.* 98, 114 (2011).
83. L. Chen, Y. Jiang, H. Nie, R. Hu, H.S. Kwok, F. Huang, and B.Z. Tang, *ACS Appl. Mater. Interfaces* 6, 17215 (2014).
84. C.A. Mills, F.L.M. Sam, A.S. Alshammari, L.J. Rozanski, N.G. Emerson, and S.R.P. Silva, *J. Display Technol.* 10, 125 (2014).
85. N. Zhou, S. Wang, Y. Xiao, and X. Li, *J. Lumin.* 196, 40 (2018).
86. Q. Zheng, D. Qu, Y. Zhang, W. Li, J. Xiong, P. Cai, and X. Zhang, *Opt. Laser Technol.* 101, 85 (2018).
87. Y.H. Huang, B.T. Liou, J.D. Chen, and Y.K. Kuo, *IEEE Photonics Technol. Lett.* 23, 480 (2011).
88. X.W. Zhang, L.M. Liu, J. Li, L. Zhang, X.Y. Jiang, Z.L. Zhang, and X.Y. Liu, *J. Display Technol.* 7, 515 (2011).
89. Y.M. Xie, Q. Sun, T. Zhu, L.S. Cui, F. Liang, S.W. Tsang, and L.S. Liao, *Org. Electron.* 55, 1 (2018).
90. Y.L. Chang and Z.H. Lu, *J. Display Technol.* 9, 459 (2013).
91. S.W. Wen, M.T. Lee, and C.H. Chen, *J. Display Technol.* 1, 90 (2005).
92. A. Fischer, S. Chénais, S. Forget, M.C. Castex, D. Ades, A. Siove, and B. Geffroy, *J. Phys. D Appl. Phys.* 39, 917 (2006).
93. C.C. Lee, C.H. Yuan, S.W. Liu, and Y.S. Shih, *J. Display Technol.* 7, 454 (2011).
94. K.J. Bergemann, R. Krasny, and S.R. Forrest, *Org. Electron.* 13, 1565 (2012).
95. F. Li, Y. Zhang, C. Wu, Z. Lin, B. Zhang, and T. Guo, *Vacuum* 86, 1895 (2012).
96. C.H. Gao, X.B. Shi, D.Y. Zhou, L. Zhang, Z.K. Wang, and L.S. Liao, *Int. J. Photoenergy* 2013, 1 (2013).
97. H. Li, Y. Tao, R. Chen, H. Li, L. Chen, Y. Jiang, and W. Huang, *Dyes Pigm.* 151, 187 (2018).
98. H.C. Chen, C.P. Kung, W.G. Houng, Y.R. Peng, Y.M. Hsien, C.C. Chou, C.J. Kao, T.H. Yang, and J. Hou, *IEEE/OSA J. Display Technol.* 5, 216 (2009).
99. M. Hack and J.J. Brown, *Inf. Display* 20, 12 (2004).
100. M.S. Weaver, V. Adamovich, M. Hack, R. Kwong, and J.J. Beown, in *Proc. Int. Conf. Electroluminescence of Molecular Materials and Related Phenomena*, Korea, O-35 (2003).
101. P. Mittal, B. Kumar, Y.S. Negi, B.K. Kaushik, and R.K. Singh, *Microelectron. J.* 43, 985 (2012).
102. D. Ammermann, A. Böhler, S. Dirr, H.H. Johannes, and W. Kowalsky, *Int. J. Electron. Commun.* 50, 327 (1996).
103. H. Riel, S. Karg, T. Beierlein, W. Rieß, and K. Neyts, *J. Appl. Phys.* 94, 5290 (2003).
104. H. Fujikawa, M. Ishii, S. Tokito, Y. Taga, in *Proc. Mater. Res. Soc. Symp.* 621, Q3.4.1. (2000).
105. D.F. O'Brien, P.E. Burrows, S.R. Forrest, B.E. Koene, D.E. Loy, and M.E. Thompson, *Adv. Mater.* 10, 1108 (1998).
106. B.J. Chen, W.Y. Lai, Z.Q. Gao, C.S. Lee, S.T. Lee, and W.A. Gambling, *Appl. Phys. Lett.* 75, 4010 (1999).
107. B. Kumar, B.K. Kaushik, and Y.S. Negi, *J. Comput. Electron.* 13, 627 (2014).
108. T. Mori, in *Lasers and Electro-Optics Society*, 2005. LEOS 2005. The 18th Annual Meeting of the IEEE (2005), p. 585.
109. S. Negi, P. Mittal, and B. Kumar, *Microsyst. Technol.* 24, 4981 (2018).
110. L.S. Hung, C.W. Tang, M.G. Mason, P. Raychaudhuri, and J. Madathil, *Appl. Phys. Lett.* 78, 544 (2001).
111. M.Y. Lim, W.M.M. Yunus, Z.A. Talib, A. Kassim, C.F. Dee, and A. Ismail, *Am. J. Eng. Appl. Sci.* 3 (2010).
112. J. Park, Y. Kawakami, and S.H. Park, *J. Lightwave Technol.* 25, 2828 (2007).
113. J. Park, T. Kim, J. Lee, and D. Shin, *IEEE Photonics Technol. Lett.* 20, 1408 (2008).
114. J.J. Huang, Y.K. Su, F.S. Juang, Y.H. Liu, and S.J.L. Chang, *IEEE Photonics Technol. Lett.* 20, 1784 (2008).
115. Q. Wang, Z. Deng, J. Chen, and D. Ma, *IEEE Trans. Electron Devices* 57, 2221 (2010).
116. T. Dobbertin, M. Kroeger, D. Heithecker, D. Schneider, D. Metzdorf, H. Neuner, E. Becker, H.H. Johannes, and W. Kowalsky, *Appl. Phys. Lett.* 82, 284 (2003).
117. V. Bulovi, V.B. Khalfin, G. Gu, P. Burrows, D.Z. Garbuzov, and S.R. Forrest, *Phys. Rev. B* 58, 3730 (1998).
118. T. Shiga, H. Fujikawa, and Y. Taga, *J. Appl. Phys.* 93, 19 (2003).
119. C.K. Moon, S.Y. Kim, J.H. Lee, and J.J. Kim, *Opt. Express* 23, A279 (2015).
120. W.H. Kim, A.J. Mäkinen, N. Nikolov, R. Shashidhar, H. Kim, and Z. Kafafi, *Appl. Phys. Lett.* 80, 3844 (2002).
121. C.T. Lee, L.Z. Yu, and H.Y. Liu, *IEEE Photonics Technol. Lett.* 22, 272 (2010).
122. J.W. Huh, J. Moon, J.W. Lee, D.H. Cho, J.W. Shin, J.H. Han, and H.Y. Chu, *IEEE Photonics J.* 4, 39 (2012).
123. D. Wahyuningrum and A. Alni, *Procedia Chem.* 16, 586 (2015).
124. E.R. Triboni, M.R. Fernandes, J.R. Garcia, M.C. Carreira, R.G.D.S. Berlinck, P.B. Filho, and M. Cremona, *J. Taibah Univ. Sci.* 9, 579 (2015).
125. F. Krujatz, O. Hild, K. Fehse, M. Jahnel, and A. Werner, *Chem. Sci. J.* 7, 134 (2016).
126. J. Shi and C.W. Tang, *Appl. Phys. Lett.* 80, 3201 (2002).
127. S.W. Culligan, Y. Geng, S.H. Chen, K. Klubek, K.M. Vaeth, and C.W. Tang, *Adv. Mater.* 15, 1176 (2003).
128. C. Hosokawa, H. Higashi, H. Nakamura, and T. Kusumoto, *Appl. Phys. Lett.* 67, 3853 (1995).
129. W.J. Shen, B. Banumathy, H.H. Chen, and C.H. Chen, in *Proceedings of the International Display Manufacturing Conference* (2003), p. 741.
130. Z.Y. Xie, L.S. Hung, and S.T. Lee, *Appl. Phys. Lett.* 79, 1048 (2001).
131. K. Suzuki, A. Seno, H. Tanabe, and K. Ueno, *Synth. Met.* 143, 89 (2004).
132. J. Luo, Y. Zhou, Z.Q. Niu, Q.F. Zhou, Y. Ma, and J. Pei, *J. Am. Chem. Soc.* 129, 11314 (2007).
133. C.W. Tang, S.A. VanSlyke, and C.H. Chen, *J. Appl. Phys.* 65, 3610 (1989).
134. C.W. Ko and Y.T. Tao, *Appl. Phys. Lett.* 79, 4234 (2001).
135. P. Chen, L. Zhao, and Y. Duan, *J. Lumin.* 131, 2144 (2011).
136. H.J. Kim, M.H. Shin, H.G. Hong, B.S. Song, S.K. Kim, W.H. Koo, and Y.J. Kim, *J. Display Technol.* 12, 526 (2016).

137. I. Akimoto, S. Tsuzuki, H. Uzawa, M. Hinatsu, Y. Nishide, H. Osuga, and H. Sakamoto, *Phys. Status Solidi (c)* 8, 124 (2011).
138. J. Kido, K. Hongawa, K. Okuyama, and K. Nagai, *Appl. Phys. Lett.* 64, 815 (1994).
139. J. Kido, M. Kimura, and K. Nagai, *Science* 267, 1332 (1995).
140. M. Li, W. Li, J. Niu, B. Chu, B. Li, X. Sun, and Z. Hu, *Solid-State Electron.* 49, 1956 (2005).
141. B.W. D'Andrade, M.E. Thompson, and S.R. Forrest, *Adv. Mater.* 14, 147 (2002).
142. J. Thompson, R.I.R. Blyth, M. Mazzeo, M. Anni, G. Gigli, and R. Cingolani, *Appl. Phys. Lett.* 79, 560 (2001).
143. Y. Nishide, H. Osuga, K. Iwata, K. Tanaka, and H. Sakamoto, *Bull. Chem. Soc. Jpn.* 81, 1322 (2008).
144. Y. Sun, N.C. Giebink, H. Kanno, B. Ma, M.E. Thompson, and S.R. Forrest, *Nature* 440, 908 (2006).
145. J.H. Seo, J.H. Kim, J.H. Seo, *SID Symposium Digest of Technical Papers* p2-38, 1399 (2007).
146. J.H. Seo and J.H. Park, *Appl. Phys. Lett.* 90, 203507 (2007).
147. B. Kumar, B.K. Kaushik, and Y.S. Negi, *I.E.T. Circuits Devices Syst.* 8, 131 (2014).
148. J. Shi and C.W. Tang, *Appl. Phys. Lett.* 70, 1665 (1997).
149. W. Li, R.A. Jones, S.C. Allen, J.C. Heikenfeld, and A.J. Steckl, *J. Display Technol.* 2, 143 (2006).
150. K. Fehse, K. Walzer, K. Leo, W. Lövenich, and A. Elschner, *Adv. Mater.* 19, 441 (2007).
151. S.I. Na, S.S. Kim, J. Jo, and D.Y. Kim, *Adv. Mater.* 20, 1 (2008).
152. S.F. Tseng, W.T. Hsiao, K.C. Huang, and D. Chiang, *Appl. Phys. A* 112, 41 (2013).
153. B.H. Tong, Q.B. Mei, R.Q. Tian, M. Yang, Q.F. Hua, Y.J. Shi, and S.H. Ye, *RSC Adv.* 6, 34970 (2016).
154. B. Kumar, B.K. Kaushik, Y.S. Negi, S. Saxena, and G.D. Varma, *Microelectron. J.* 44, 736 (2013).
155. H. Klauk, U. Zschieschang, and M. Halik, *J. Appl. Phys.* 102, 074514-1 (2007).
156. J. Puigdollers, C. Voz, I. Martin, A. Orpella, M. Vetter, and R. Alcubilla, *J. Non-Cryst. Solids* 338, 617 (2004).
157. S. Acharya, J.P. Hill, and K. Ariga, *Adv. Mater.* 21, 1 (2009).
158. G. Horowitz, *Adv. Mater.* 10, 365 (1998).
159. H.N. Raval, S.P. Tiwari, R.R. Navan, S.G. Mhaisalkar, and V.R. Rao, *IEEE Electron Device Lett.* 30, 484 (2009).
160. H. Cho, H. Yoon, K. Char, Y. Hong, and C. Lee, *Jpn. J. Appl. Phys.* 49, 05EB08-1 (2010).
161. B. Kumar, B.K. Kaushik, and Y.S. Negi, *J. Comput. Electron.* 12, 765 (2013).
162. J.H. Kim, D.S. Kim, S.K. Kim, Y.Z. Yoo, J.H. Lee, S.W. Kim, and T.Y. Seong, *Ceram. Int.* 42, 3473 (2016).
163. D. Miao, S. Jiang, S. Shang, and Z. Chen, *Vacuum* 106, 1 (2014).
164. Y. Yoo and H.K. Jeong, *Chem. Commun.* 21, 2441 (2008).
165. V.Y. Butko, X. Chi, D.V. Lang, and A.P. Ramirez, *Appl. Phys. Lett.* 83, 4773-1 (2003).
166. D. Kumaki, T. Umeda, T. Suzuki, and S. Tokito, *Org. Electron.* 9, 921 (2008).
167. Y. Yun, C. Pearson, and M.C. Petty, *IET Circuits Devices Syst.* 3, 182 (2009).
168. W. Junfeng, J. Yadong, Y. Yajie, Y. Junsheng, and X. Jianhua, *Energy Procedia* 12, 609 (2011).
169. B. Kaushik, B. Kumar, Y.S. Negi, P. Mittal, in *Advances in Computer Science, Engineering & Applications*, vol. 167 (Springer, Berlin, Heidelberg, 2012), p. 125.
170. C.M. Keum, S. Liu, A. Al-Shadeedi, V. Kaphle, M.K. Callens, L. Han, and R.J. Twieg, *Sci. Rep.* 8, 699 (2018).
171. H.A. Becerril, M.E. Roberts, Z.H. Liu, J. Locklin, and Z.N. Bao, *Adv. Mater.* 20, 2588 (2008).
172. G. Giri, E. Verploegen, S.C. Mannsfeld, S. Atahan-Evrenk, D.H. Kim, S.Y. Lee, and Z. Bao, *Nature* 480, 504 (2011).
173. J.H. Kim, S.T. Williams, N. Cho, C.C. Chueh, and A.K.Y. Jen, *Adv. Energy Mater.* 5, 1401229 (2015).
174. J. Xu, Y. Wang, H. Shan, Y. Lin, Q. Chen, V.A.L. Roy, and Z. Xu, *ACS Appl. Mater. Interfaces* 8, 18991 (2016).
175. C. Grigoriadis, C. Niebel, C. Ruzie, Y.H. Geerts, and G. Floudas, *J. Phys. Chem. B* 118, 1443 (2014).
176. C. Xu, P. Cai, X. Zhang, Z. Zhang, X. Xue, J. Xiong, and J. Zhang, *Sol. Energy Mater. Sol. Cells* 159, 136 (2017).
177. M. Zvarik, D. Martinicky, L. Hunakova, I. Lajdova, and L. Sikurova, *Neoplasma* 60, 533 (2013).
178. Q. Zheng, F. You, J. Xu, J. Xiong, X. Xue, P. Cai, and L. Wang, *Org. Electron.* 46, 7 (2017).
179. M. Faustini, B. Louis, P.A. Albouy, M. Kuemmel, and D. Grosso, *J. Phys. Chem. C* 114, 7637 (2010).
180. M. Le Berre, Y. Chen, and D. Baigl, *Langmuir* 25, 2554 (2009).
181. S. Roland, R.E. Prud'homme, and C.G. Bazuin, *ACS Macro Lett.* 1, 973 (2012).
182. P. Mittal, Y.S. Negi, and R.K. Singh, *Microelectron. Eng.* 150, 7 (2016).
183. P. Mittal, Y.S. Negi, and R.K. Singh, *J. Comput. Electron.* 14, 828 (2015).
184. S. Negi, P. Mittal, and B. Kumar, *IET Circuits Devices Syst.* 13, 1255 (2019).
185. G.G. Malliaras and J.C. Scott, *J. Appl. Phys.* 83, 5399 (1998).
186. J. Chan, A.D. Rakic, Y.T. Yeow, and A.B. Djurisic, in *IEEE Conference on Optoelectronic and Microelectronic Materials and Devices* (2004), p. 53.
187. J. Park and Y. Kawakami, *J. Display Technol.* 2, 333 (2006).
188. P.W.M. Blom, M.J.M. de Jong, and S. Breedijk, *Appl. Phys. Lett.* 71, 930 (1997).
189. Y. Nagata and C. Lennartz, *J. Chem. Phys.* 129, 034709 (2008).
190. C. Pflumm, C. Gartner, and U. Lemmer, *IEEE J. Quantum Electron.* 44, 790 (2008).
191. C. Pflumm, C. Karnutsch, M. Gerken, and U. Lemmer, *IEEE J. Quantum Electron.* 41, 316 (2005).
192. D. Shin, Y. Kim, N. Chang, and M. Pedram, *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 32, 1017 (2013).
193. T.K. Wee and R.K. Balan, in *Proceedings of the first ACM International Workshop on Mobile Gaming* (2012), p. 25.
194. M.S. El-Nasr and S. Yan, in *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, ACE'06* (2006).
195. C.H. Lin, C.K. Kang, and P.C. Hsiu, in *2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC)* (2014), p. 1.
196. Y. Liu, Y. Chen, D. Buso, and G. Zissis, *IEEE Trans. Ind. Appl.* 52, 5219 (2016).
197. J. Jacobs, D. Hente, and E. Waffenschmidt, in *2007 IEEE Industry Applications Annual Meeting* (2007), p. 1147.
198. D. Buso, S. Bhosle, Y. Liu, M. Ternisien, C. Renaud, and Y. Chen, *IEEE Trans. Ind. Appl.* 50, 1459 (2014).
199. R.L. Lin, J.Y. Tsai, D. Buso, and G. Zissis, in *2015 IEEE Industry Applications Society Annual Meeting* (2015), p. 1.
200. C.L. Fan, Y.C. Chen, C.C. Yang, Y.K. Tsai, and B.R. Huang, *J. Display Technol.* 12, 425 (2016).
201. S. Negi, P. Mittal, B. Kumar, and P.K. Juneja, *Microelectron. Eng.* 218, 111154 (2019).
202. K.I. Miyamoto, K. Kaneko, A. Matsuo, T. Wagner, S.I. Kanoh, M.J. Schöning, and T. Yoshinobu, *Sens. Actuators B Chem.* 170, 82 (2012).
203. C.F. Werner, T. Wagner, K.I. Miyamoto, T. Yoshinobu, and M.J. Schöning, *Sens. Actuators B Chem.* 175, 118 (2012).
204. R. Shinar, Z. Zhou, B. Choudhury, and J. Shinar, *Anal. Chim. Acta* 568, 190 (2006).
205. P. Mittal, Y.S. Negi, and R.K. Singh, *J. Comput. Electron.* 14, 360 (2015).
206. E. Manna, T. Xiao, J. Shinar, and R. Shinar, *Electronics* 4, 688 (2015).
207. B.A. Prabowo, Y.F. Chang, Y.Y. Lee, L.C. Su, C.J. Yu, Y.H. Lin, and K.C. Liu, *Sens. Actuators B Chem.* 198, 424 (2014).
208. B. Choudhury, R. Shinar, and J. Shinar, *J. Appl. Phys.* 96, 2949 (2004).

209. J. Vučić, C. Kottke, S. Nerreter, K. Habel, A. Büttner, K.D. Langer, and J.W. Walewski, in *Proc. OFC NFOEC* (2010), p. 1.
210. P.A. Haigh, Z. Ghassemlooy, H. Le Minh, S. Rajbhandari, F. Arca, S.F. Tedde, and I. Papakonstantinou, *J. Lightwave Technol.* 30, 3081 (2012).
211. P.A. Haigh, Z. Ghassemlooy, and I. Papakonstantinou, *IEEE Photonics Technol. Lett.* 25, 615 (2013).

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

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Intrusion detection systems: A Comparison of soft computing techniques

To cite this article: Anamika Chauhan and Kapil Sharma 2020 *J. Phys.: Conf. Ser.* **1478** 012013

View the article online for updates and enhancements.

IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection–download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.100 on 22/05/2020 at 06:53

Journal of Physics: Conference Series **1478** (2020) 012013 doi:10.1088/1742-6596/1478/1/012013

Intrusion detection systems: A Comparison of soft computing techniques

Anamika Chauhan[1] and Kapil Sharma[1]

Delhi Technological University , New Delhi-42 India

Abstract. With rapid emergence of Internet of Things paradigm, there is a continual increase in the devices that are able to connect and exchange information, over the internet. This reachability has a negative consequence in the form of increased vulnerability of these devices to a variety of cyber attacks. The most efficient and pre emptive tool for this is an intrusion detection system. Since the attacks and threats continue to evolve over time, so the IDS should be able to learn as well to cope with the attacks. In this paper we have presented a comparison of some Soft computing techniques that can be used in designing of IDSs that can efficiently classify, detect and prevent attacks.

Keywords: Intrusion detection, cyber security.

1 Introduction

As Internet is becoming more and more pervasive in our everyday life, we have direct and indirect connectivity to the cyber world at all times. Therefore, Internet of Things concept technology and connectivity has probed further into our personal space. Though IoT has enabled smart environments, with applications and, that would enhance the life and experience for human life. It has on the flip side exposed most of our information exchange to malicious intruders and attackers. Thus most essential pre-requisite for such an environment would be its security and robustness.

With the number of devices and applications increasing exponentially, networking infrastructure has to deal with enormous amount of data traffic. This data will consist of useful as well as malicious data. This threat is manifold and has to be mitigated with deployment of new cyber security mechanisms. This mechanism must ensure security of data as well as regulatory and legal rights of users. Most of the attacks are basically intrusions into the network in the form of malware, bots, viruses, worms and Trojans.

The features of the security mechanism thus must ensure the following:

1. Maintain Privacy & security of data
2. Authentication and access control
3. Detecting malicious activity in the network
4. Warning and Mitigation of attacks

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

1.1 Types of IDS

The solution to most of the above is an intrusion detection system or further referred as IDS . It is a software designed to automate monitoring of packet traffic in network, and classify it into malicious or legitimate. In case malicious activity is detected the IDS must either warn administrator or take necessary actions to mitigate the attack. For future learning a log record must also be maintained.

An IDS works by examining vulnerabilities in the system and analyzing patterns based on known attacks. According to [1][2], IDS can be categorized as HIDS and NIDS i.e. host-based and network-based intrusion detection system respectively. HIDS operate on an individual system and possess the complete knowledge about the data and processing of that system. These are equipped with the servers and provide security to the mails, servers of a firm. HIDS provide high security as they have the complete knowledge of data and servers of a computer system. But here our concern is NIDS, NIDS operate on network. NIDS are deployed to handle the traffic over network. NIDS are used to restrict the malicious activities taking place over the internet, they does not replace the security implementations like firewall rather they add to them. The NIDS are used to detect those attacks which can't be detected by the basic security parameters. Now onwards we shall use the term IDS for NIDS.

Further IDS can be divided into two broad categories as signature based or anomaly based IDS. Anomaly based IDS analyze the traffic in the network and if in the received data, there is some deviation from the usual normal behavior then it raises a red flag but False alarm rate (FAR) is high with anomaly based IDS which makes its implementation quite onerous, the reason for the same is that attacker attacks the network with the values close to the training values of IDS. Whereas on the other hand Signature based IDS has its own database of known intrusions and it detects those intrusions only. [1] [3]

1.2 Implementation of IDS

As can be inferred from above Intrusion detection can be considered to be a classification problem, hence many Soft Computing techniques are used in order to build IDS like SVM, Random Forest, Swarm Optimization and Convolutional Neural Networks. These IDS are first trained on the basis of nature of intrusions then they are tested with some sample data after that these systems are deployed in the networks.

These IDS face some problems like the first problem is that nature of attack is variable, it changes vigorously and in order to cope with such changing attacks, it is important to make IDS which can detect such types of attacks. The second problem is of over-fitting, the machine learning methods usually face the problem of over-fitting.

Therefore it is immensely important to distribute the dataset into two parts, one for training and the other for testing. Also balance must be maintained among the features i.e. irrelevant features and redundant features must be taken care of. The third problem is labeling the dataset of intrusion. The dataset are so complex that labeling becomes a typical task. Efforts have been made to simplify the task of labeling the dataset. The problems faced are tackled by the IDS throughout its working phases.

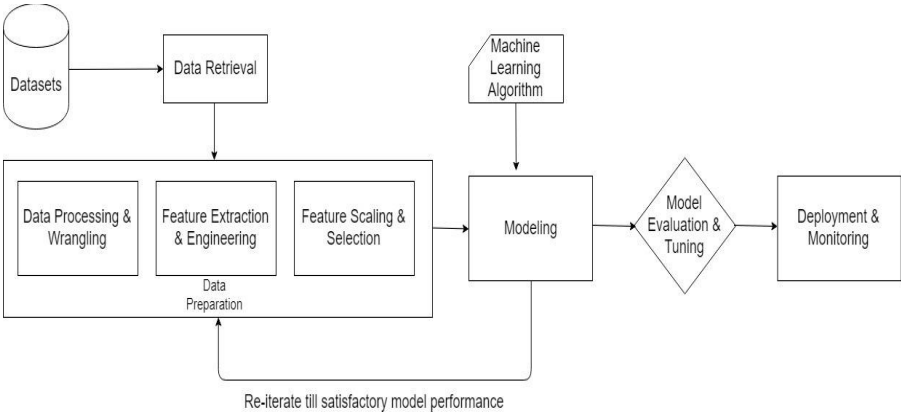


Fig.1. Implementation of IDS

Since new attacks are being designed every day, static or state-less models are unable to keep up. An effective way to improve and augment intrusion detection systems is Soft Computing techniques .The effectiveness of a Soft Computing based IDS directly depends on data and classification algorithm. Also since the data being monitored is substantial, irrelevant and redundant features must be minimized to maximize the performance of IDS [4] [5]. For this reason, a preprocessing is performed to select meaningful features and improve accuracy.

Selection of classification algorithm is also important. The accuracy and resource consumption are the two main parameters that must be considered when selecting a suitable algorithm.

2

2.1 Soft computing and classification

The idea behind classification is to predict the target class by analyzing the training dataset. The primary objective is to choose the appropriate classifier. As the choice of wrong classifier, often leads to misinterpretation of target class [6]. The accuracy of the solution, of the classification problem, depends on the selection of right classifier [5][7] The important factors used to measure the performance of any classification algorithm are:

1. Training time
2. Data fitting
3. Computational efficiency
4. Stability and
5. Learning rate

In this paper a comparison of six well known classification algorithms ; Support Vector Machine (SVM), Random Forest, Artificial Neural Network (ANN), Decision Tree, Logistic Regression and Naïve-Bayes is performed on the NSL-KDD dataset .

Artificial Neural Network (ANN): Artificial Neural Network is one of the most suitable technique that can be applied on IDS. An ANN mimics the behavior of the biological nervous systems or the brain of a human being , which is most complex information processing model. An ANN comprises of large number of interconnected processing elements (neurons), that are highly interconnected and work together to solve a specific problem. An ANN is composed of three layers: an Input layer, a hidden and an output layer. The ANN is trained in the first phase by learning and by reducing the error by adjusting the weight of connections that exist between neurons modeled for a specific application For different Activation function ANN results in different classification accuracy. ANN provides high classification accuracy for Logistic and tanh and low in Identity activation function. ANN is most suitable for high frequency attacks like flooding etc as that increases number of neurons in each hidden layer. The shortcoming of ANN is poor detection accuracy and precision in case of low-frequency attacks [3].

SVM: Support Vector Machine (SVM) is a supervised machine learning algorithm . SVM can be most efficient in solving classification problems as well as for regression analysis. But SVM is most suitable by way of design for binary classification. Compared to most ML techniques SVM gives better generalization capability as it is an eager learning algorithm. But the major shortcoming is low computational efficiency. Also its use is limited for very large data since it has a high training time. . Since SVM is based on the concept of hypothesis planes so decision boundaries are well defined and the algorithm detects outliers much better. A hypothesis plane is a plane that separates the input data into different class memberships.

In SVM the precision of classifier is dependent on kernel, the order of increasing precision is polynomial kernel followed by linear and sigmoid kernel giving nearly same precision and highest in RBF.

Random Forest: Random forest is also a supervised learning algorithm. Many similarities are there between SVM and Random forest algorithm. Both are used for solving classification problems as well as regression. The advantage of Random forest algorithm lies in its ability to classify large databases of data with high accuracy [1]. Random forest is also an eager learning algorithm. The reason for higher performance is that higher number of trees in the Random forest gives higher accuracy. But the downside of this is that larger forest consumes larger resources. So RF requires more computational time, memory and processor cycles. [8]

Decision Tree: like previous two techniques Decision Trees are also another variety of supervised classification algorithm that may be applied for solving both classification as well as regression problems. The algorithm can quickly learn complex decision boundaries even in large dataset. The algorithm is based on construction of a tree from training data in which interior nodes corresponds to any of the input feature and leaf nodes denotes a decision for the data item. Decision Tree algorithms performance is highly dependent on the noise in the training data. To solve this problem Pruning is used, pruning improves the performance of the Decision tree. It is the process of removing the branches that having low importance features. Cost complexity pruning and reduced cost pruning are considered the most efficient methods.

Logistic Regression: Logistic regression is probably the most commonly used data mining algorithm. It is a simple algorithm with low model building time. Logistic regression predicts the probability of problems with binary outcome. Logistics regression generates a logistic curve, which is limited to values between 0 and 1. Regularization doesn't improve the performance but it can improve the generalization performance of the classifier on the data set that are used to learn the model parameter . For regularization type Lasso the precision is 0.985 but for Ridge the precision is 0.984.

Naïve Bayes: Naïve Bayes is a simple and elegant powerful data mining algorithm used for classification problems. It follows the basic principle of basic bayes theorem and forms a Bayesian network. Simply put , a naïve bayes classifier makes an assumption that the presence of one particular feature in a class is not related to the presence of any other feature. Naïve bayes algorithm works efficiently with large datasets, but when the attributes in datasets are largely interrelated then it suffers from oversensitivity and reduces the performance of classifier.

3 Implementation

The datasets used by the methods discussed in this comparison are KDD '99 and NSL-KDD. KDD '99 was generated at MIT Lincoln Laboratory [9] [10] The attacks defined in this dataset are 4 major types of attacks that are encountered. It consists of 41 features out of which 38 have numerical values and 3 have non-numeric data. KDD '99 can be considered as an imbalanced dataset as it has around 80 percent data as attack data.

NSL-KDD is improvement to KDD '99. Both the dataset are widely used to serve the purpose of IDS since a long time. These two datasets are the benchmark in the field of intrusion detection system. NSL-KDD consists of training data KDDTrain+ and test data KDDTest+ as mentioned in [5]. In NSL-KDD U2R and R2L are also increased, although the increment is still insignificant. In the rest of the paper we shall comparison the methods for IDS.

In this paper authors use 20% of NSL-KDD training dataset with 10 folds cross validation to build model to classify the pattern is normal or attack.
The implementation was carried out in three phases as follows:
a) Data preprocessing

b) Model building and learning

c) Testing and Cross-validation

Data preprocessing: In this phase of implementation we remove redundant and null values from the data, as these redundant and null values were not adding any quality to model building.

Model building and learning: The refined dataset obtained after preprocessing phase is fed to train and build the models for the classification of the data.

Testing and Cross Validation: The results of the modeling phase are then used to compare amongst the various candidate classification algorithms. The parameters used in this paper are the True and false values of positive and negative classification. As well as the accuracy, precision and recall values of each mode.

The above phases were simulated for all possible attack types. The list of attacks and sub-categories for each type of attack are shown in table 1.

Table 1. Various attacks and their categories

Class	**Known Attack Subclasses**
DoS	Back, Neptune, pod, land, smurf, teadrop
Probe	Ipsweep, portsweep, nmap,satan
U2R	Buffer_overflow, perl, loadmodule,rootkit
R2L	ftp_write, imap, multihop, guess_passwd, phf, spy, warezclient,warezmaster

4 Performance Measures

While solving a problem, accuracy of classifier is an important performance measure, for selecting most appropriate technique for that problem. Performance of classifier model is predicted by how many input instances are correctly predicted by classifier. After implementation of a machine learning algorithm, it must be tested against metrics suitable for that dataset to calculate efficiency. The performance metrics used to measure prediction efficiency of classifier are accuracy, specificity, sensitivity and training time. The number of test data are correctly classified by classification model can be tabulated in the form of confusion matrix. The parameters of confusion matrices for binary classifier is shown on table 2.

Table 2. Parameters of confusion matrix

Actual Data Class	Predicted Data Class	
	Positive	Negative
Positive	True Positive (tp)	False Negative (fn)
Negative	False Positive (fp)	True Negative (tn)

The characteristics used for analysis are as follows :
1) ROC (Receiver operating characteristics): For making the curve between TPR and FPR. The ROC is calculated as the from the AUC , which is the area under the curve.
2) Sensitivity: tp / (tp+fn)
3) Specificity: tn / (fp+tn)
4) Precision: tp / (tp+fp)
5) Accuracy: (tp+tn)/ (tp+tn+fp+fn)
Another very important factor is training time, the time taken by classifier to build predictive model on given dataset to predict the different class labels.

5 Result

Table 3. EXPERIMENTAL RESULTS OF ANALYSED CLASSIFICATION ALGORITHMS

ALGORITHM	ACCURACY	CLASSIFICATION ACCURACY	PRECISION	RECALL
DECISION TREE	0.978	0.964	0.959	0.974
RANDOM FOREST	0.998	0.979	0.979	0.981
SVM	0.573	0.537	0.536	1.000
LOGISTIC REGRESSION	0.906	0.842	0.918	0.774
NAÏVE BAYES	0.958	0.864	0.879	0.866
NEURAL NETWORK	0.980	0.941	0.910	0.988

Table 4. CONFUSION MATRICES OF CLASSIFICATION ALGORITHMS

	Anomaly	Normal	∑
Anomaly	97.0%	4.1%	58630
Normal	3.0%	95.9%	67343
∑	57548	68425	125973

a) Decision Tree

	Anomaly	Normal	∑
Anomaly	97.9%	2.1%	58630
Normal	2.1%	97.9%	67343
∑	58445	67528	125973

b) Random Forest (RF)

	Anomaly	Normal	∑
Anomaly	98.6%	46.4%	58630
Normal	1.4%	53.6%	67343
∑	292	125681	125973

d) Logistic Regression(LR)

	Anomaly	Normal	∑
Anomaly	77.7%	8.0%	58630
Normal	22.3%	92.0%	67343
∑	69581	56392	125973

c) Support Vector

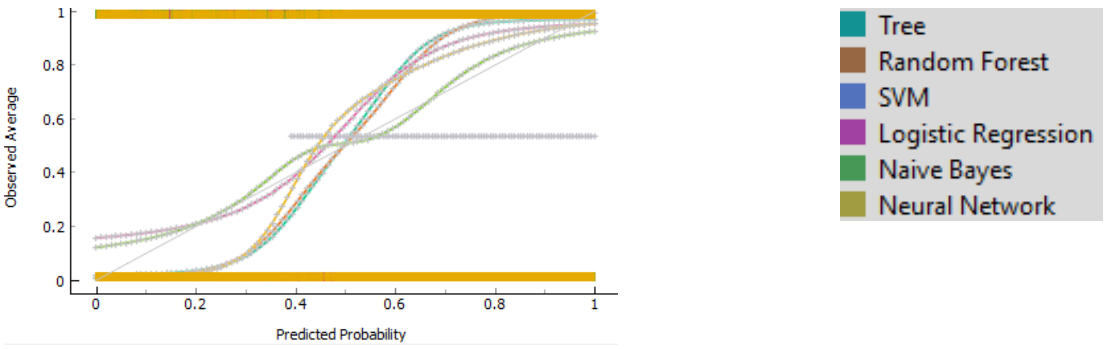
	Anomaly	Normal	∑
Anomaly	84.8%	12.1%	58630
Normal	15.2%	87.9%	67343
∑	59660	66313	125973

e) Naïve Bayes

	Anomaly	Normal	∑
Anomaly	98.4%	8.9%	58630
Normal	1.6%	91.1%	67343
∑	52961	73012	125973

f) Artificial Neural Network

Predicted class: Normal



Predicted class: Anomaly

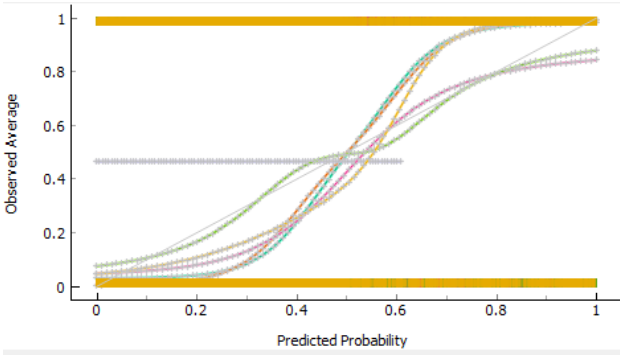


Fig.2. CALIBRATION PLOT OF CLASSIFICATION ALGORITHM

6 Conclusion

In this paper, we have implemented the six well known soft computing algorithms necessary for proper detection of network intrusion. We have also compared the performance of various classification techniques and concluded that SVM and ANN are most suitable for the purpose of intrusion detection. To reason for this is that SVM works even on small data samples, high dimensionality and ANN works well with both strong and weak classifiers. Both also work well as they need lesser training time and therefore converge faster.

Bibliography

[1] P. L. M. S. Sinclair C, " An application of machine learning to network intrusion," in In Computer Security Applications Conference(ACSAC), 1999.

[2] W. I. Troesch M, " Machine Learning for Network Intrusion Detection," 2014.

[3] H. Y. L. C. L. W. Tsai CF, "Intrusion detection by machine learning : A review," in Expert Systems with Applications. 2009 Dec 1;36(10):11994-2000., 2009.

[4] K. V. P. S. P. E. Chauhan H, "A comparative study of classification techniques for intrusion detection," in Computational and Business Intelligence (ISCBI), 2013.

[5] K. S. Garg T, "Comparison of classification techniques for intrusion detection dataset using WEKA," in In Recent Advances and Innovations in Engineering (ICRAIE), 2014.

[6] G. K. Bhoria MP, "An Imperial learning of Data Mining Classification Algorithms in Intrusion Detection Dataset," in International Journal of Scientific & Engineering Research, 2013.

[7] A. E. Gül A, "A feature selection algorithm for IDS," in Computer Science and Engineering (UBMK) International conference, 2017.

[8] M. M. Aung YY, "An analysis of random forest algorithm based network intrusion detection system.," in In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017.

[9] B. E. L. W. G. A. Tavallaee M, "A detailed analysis of the KDD CUP 99 data set," in Computational Intelligence for Security and Defense Applications, CISDA, 2009.

[10] D. M. S. D. K. K. A. P. Lahre MK, "Analyze different approaches for IDS using KDD 99 data set," in International Journal on Recent and Innovation Trends in Computing and Communication, ISSN., 2013.

[11] B. A. Choudhury S, " Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection," in Smart technologies and management for computing, communication, controls, energy and materials (ICSTM), 2015.

[12] K. H, "Algorithm used in Intrusion Detection System : a Review," in International Journal of Innovative Research in Computer and Communication Engineering, 2013.

2020 4th International Conference on Computational Intelligence and Networks (CINE) 978-1-7281-5688-0/20/$31.00 ©2020 IEEE 10.1109/CINE48825.2020.234394

Learning Interpretable Hidden State Structures for Handwritten Numeral Recognition

Seba Susan, Jatin Malhotra
Delhi Technological University, Shabad Daulatpur, Bawana Road, Delhi, India-110042
seba_406@yahoo.in

***Abstract*— A novel paradigm of learning interpretable hidden state structures from the Multi-Layer Perceptron (MLP) neural network is proposed, in this work, that considers hidden state activations as the feature vector. The classification task at hand is handwritten numeral recognition. The k-means Elbow method is used to analyze the interpretability of the clusters formed from each hidden layer. Only the hidden layer that forms well-defined and linearly separable clusters, corresponding to the different numeral classes, is considered for the feature extraction. An interpretable structure thus emerges from the hidden representations that are learnt using a suitable classifier. Direct learning of hidden states using the linear k-Nearest Neighbor (kNN) classifier and the Support Vector Machine (SVM) with linear kernel yields high accuracies. Two different MLP configurations are tested for the purpose. Patch-based hidden structure learning with feature fusion is found to further improve the results. Experiments on the benchmark MNIST handwritten numeral image dataset prove the efficacy of our interpretability-based approach as compared to existing works.**

Keywords—Multi-layer perceptron, hidden structure, handwritten numeral recognition, structure learning

I. INTRODUCTION

The recognition of handwritten digits and alphabets plays a significant role in the process of digitization of documents for deciphering handwritten information without human intervention. The challenges faced in the task are the poor, degraded quality of the document, the presence of external objects that occlude the text such as fingers, pen etc., and most often, poor lighting conditions [1]. Most of the experiments on handwritten digit recognition are of the supervised type and involve extracting intricate features from the digit image such as the contour feature vector based on chain code histogram [2], the curvature features involving distance and slope [3] and the fraction of black pixels in box-partitions of the numeral image [4]. Other than feature engineering, the recent advances made by deep neural networks have also made an impact s[5, 6]. Most of the experiments done by researchers, especially in deep learning, involve the benchmark MNIST dataset containing 70,000 images of handwritten English numerals [7] or the Persian numeral images containing 80,000 images [8]. Low-resource datasets are tackled by transfer learning that involves training the model with a high-resource dataset [9]. In this paper, we propose a novel perspective on the handwritten digit recognition task by interpreting hidden state structures in Multi-Layer Perceptron (MLP) neural networks as linearly separable entities and the states of a finite state automaton. Clustering is used to validate the existence of linearly separable hidden states prior to the feature extraction. The interpretable latent representation so derived, serves as the feature vector for classification. The organization of this paper is as follows. The literature on structure learning is reviewed in Section II, the proposed methodology of learning interpretable hidden state structures and patch-based learning is described in Section III, the experiments and results are discussed in Section IV and the overall conclusions are summarized in Section V.

II. RELATED WORK ON STRUCTURE LEARNING

Structure learning refers to the machine learning strategy where a structured object is learnt from the labeled training samples, instead of the class labels themselves [10]. Instances of 'structure' include vectors [11], trees [12], graphs [13] and sequences [14]. A structured perceptron is learnt in [15] for the tagging problem in natural language processing. A variant of the perceptron algorithm is suggested in [15] that learns the mapping between the input $x \epsilon X$ and output labels $y \epsilon Y$. The task is to learn the representation that maps the vector (x,y) to a d-dimensional feature vector. The derived feature serves as an indicator function that determines the tag. In preceding literature, this task was termed- rule extraction from Artificial Neural Networks (ANNs) [16]. Rules were added and updated in ANNs, simultaneously increasing the symbolic knowledge derived from the network structure. A dynamically growing hidden layer is incorporated in the ANN in [17] that increments the size of the hidden layer on observing a higher entropy of the input distribution. Frasconi *et al.* [18], introduced a state hidden layer in Radial Basis Function (RBF) with feedback connections in order to extract a trained automaton from the RBF network. The network state space is interpreted as a set of symbolic states of the automaton.

In recent times, the need for interpreting the existing machine learning algorithms has assumed importance [19] especially with the advent of deep neural networks [20] that are highly advanced forms of MLP. Interpretation in the form of finite state automaton is a major approach [21]. This concept of formation of linearly separable states is further explored in our work by considering the Multi-Layer Perceptron (MLP) as the base or parent network for extracting the hidden state structure, akin to deep feature extraction and transfer learning mechanism in Computer Vision [28].

978-1-7281-5688-0/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2020 at 05:42:43 UTC from IEEE Xplore. Restrictions apply.

Clustering of hidden state activations is performed as a litmus test to detect meaningful state entities, that helps to decode the classification task more efficiently.

III. Learning Interpretable Hidden State Structure for Handwritten Numeral Recognition

Our work progresses on the lines of Frasconi *et al.* [18] and Hinton *et al*. [24, 25] by deriving an interpretable hidden state structure from a perceptron network for handwritten numeral classification. This is offered as a viable solution rather than learning directly from the input representations.

A. Hidden structure learning

Learning the structure of the network was basically advocated by Hinton [24] and this was the basis for the introduction of Autoencoders [25]. The hidden state activations of a Long Short Term Memory (LSTM) are considered as the features to be learnt in [26], in place of the input representation from video frames. In the field of deep networks, semantic feature extraction from layers of pre-trained deep networks has been investigated [28]. This representation is then decoded using another LSTM for predicting the future video sequence. The general architecture of a MLP network is shown in Fig. 1. The hidden state vector extracted from the last fully connected layer of the MLP is a linear weighted sum of the input vector $\mathbf{x}$ followed by a non-linear function as shown in (1) and (2). b_0 is the network bias.

$$h_i = \tanh\left(\mathbf{w}.\mathbf{x} + b_0\right) \tag{1}$$

$$\bar{\boldsymbol{h}} = \{h_1, h_2, \ldots h_N\} \tag{2}$$

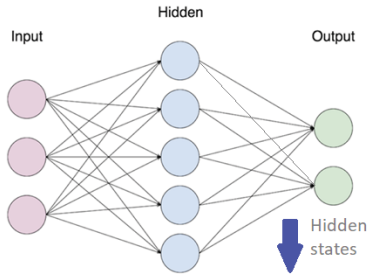


Fig. 1. Hidden state vector extracted from an example 3-6-2 MLP network

B. Framework: The Multi-Layer Perceptron architecture

A 784-35-10 Multi-Layer Perceptron (MLP) network (**Net 1**) is the framework for the first phase of our experiments on the benchmark MNIST dataset of handwritten English numerals. Each grayscale image of this dataset is of size 28×28 which is converted to a linear 784-dimensional vector that is given as input to the MLP. There are 10 classes of numerals ranging from 0 to 9. The single hidden layer comprising of 35 neurons is a fully connected layer. The activations of these 35 hidden neurons observed over time, as training progresses, reflect the outcome of the training when they start assuming distinct, linearly separable clusters for the 10 classes. A feedforward-backpropagation procedure is used for the training phase and a simple feedforward in the testing phase computes the class labels using a softmax function, such that the sum of the probabilities of output nodes is one. As per [27], the weights start stabilizing much before the network error does, hence learning the hidden state vector rather than the input representation is a viable learning paradigm and forms the basis for the investigation done in this paper. In the second phase of our experiments, a larger (deeper) network of size 196-35-35-35-10 (**Net 2**), with three hidden layers containing 35 neurons each, is used for training patches extracted from the MNIST images, as elucidated in the next sub-section.

C. Learning interpretable hidden state structures: Concept and Methodology

Clustering the hidden state activations emanating from each hidden layer is the next step, once the training is complete. The training samples are subjected to a single forward pass through the trained model for procuring the hidden state activations. For an interpretable structure, these states should be clustered into linearly separable well-defined clusters. Clustering is a useful strategy in computer vision to segment the visual information into semantically meaningful linearly separable classes [31, 32]. The k-means Elbow test [33, 37] is a graphical procedure for determining the apt number of clusters from the inflection point in the graph. The Elbow test plots the average distortion score for c=2 to c=15 clusters. The distortion score is the sum of squared differences of each data point from its respective cluster center after the clustering process is over. A smooth curve with "no Elbow" indicates that the data cannot be clustered or used for feature extraction.

Processing image patches followed by decision- or feature-level fusion is a successful approach as observed in [34] where facial images are processed via grids to detect emotions. Numeral image patches processed by a cascade of classifiers aid in improved classification in [23, 29]. A patch-based training procedure is adopted in the second phase of our experiments. Fig. 2 (a) shows the layout of the 14×14 patches in the numeral image. The five patches are displayed for an instance of the numeral '6' in Fig. 2 (b). A set of five MLP classifiers (each of configuration 196-35-35-35-10) are trained on the respective patches. The features extracted from the interpretable hidden layer in all the five classifiers are then fused by concatenation for further classification.

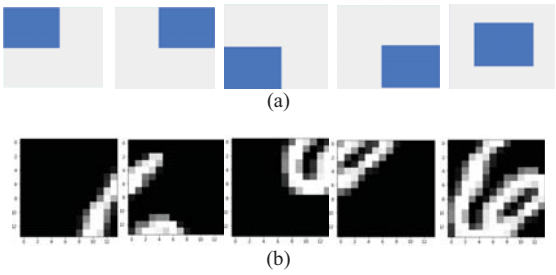


Fig. 2. (a) The layout of the five image patches each covering a quarter area of the image (b) The five patches shown for the numeral '6'

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2020 at 05:42:43 UTC from IEEE Xplore. Restrictions apply.

The proposed learning methodology is summarized below.

1. Train the MLP network (**Net 1**, **Net 2**) for 200 epochs with a learning rate of 0.1
2. Onto the trained model, present the training and testing samples for deriving the hidden state vector using a single forward pass.
3. Cluster the hidden state vectors emanating from each hidden layer, for number of clusters c=2 to c=15.
4. Determine the inflection point if any, using the k-means Elbow test. Select the hidden layer(s) whose activations form an inflection point, for extracting the 1×35 feature vector of hidden state activations from each training and test image.
5. Learn the 35-dimensional feature vector using a Support Vector Machine (SVM) or k-Nearest Neighbor (k-NN) classifier (with k varying from 1 to 9). A linear kernel for SVM (with C=1) is used.

IV. Results and Discussions

A. Experimental setup

The MNIST handwritten numeral dataset [22] is used for the experiments. The MNIST dataset comprises of 60,000 training images and 10,000 images for testing, segregated at the source. There are ten classes of numerals representing the digits 0,…,9; some samples of images are shown in Fig. 3.

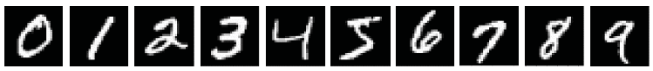


Fig. 3. Some samples of MNIST handwritten numeral image dataset

Each numeral image is of size 28×28 that when converted to 1D, yields a 784-dimensional input vector. The size of each patch, in the second phase of our experiments involving **Net 2**, is 14×14 that translates to a 196-dimensional input vector. The hyperparameters for the Multi-Layer Perceptron (MLP), and the device specifications are listed below.

Hyperparameters (Multi-layer perceptron)

Layer1 size (INPUT LAYER) = 784, [Layer1 size (INPUT LAYER) = 196 (for patch)], Layer2 size (HIDDEN LAYER) = 35 (Additional 2 layers added for Phase 2), Layer3 size (OUTPUT LAYER) = 10, Learning Rate = 0.1, Epoch = 200.

Device Specifications

Processor: Intel(R) Core (TM) i5-5200U CPU @ 2.20GHz, Memory: RAM 8.00GB DDR3L-1600 SDRAM, Graphics card: NVIDIA GeForce 940M (2 GB DDR3L dedicated).

B. Discussion on results

The performance of various classifiers on the handwritten numeral dataset are compared in Table I with the proposed approach and also with some state-of-the-art techniques namely, Deep Structured Energy Based Models (Zhai *et al.,* 2016) [30]; Positive-Generative Adversarial Network (Chiaroni *et al.* 2018) [35] and the Rank Pruning algorithm (Northcutt *et. al.,* 2017) [36]. The F1-score metric is a harmonic mean computed as

$$\boldsymbol{F1-score} = 2(\boldsymbol{precision} * \boldsymbol{recall}) / (\boldsymbol{precision} + \boldsymbol{recall}) \quad (3)$$

The **precision** and **recall** computations are based on the number of True Positives (TP), False Positives (FP) and False Negatives (FN) as shown below.

$$\boldsymbol{precision} = \boldsymbol{TP} / (\boldsymbol{TP} + \boldsymbol{FP}) \quad (4)$$

$$\boldsymbol{recall} = \boldsymbol{TP} / (\boldsymbol{TP} + \boldsymbol{FN}) \quad (5)$$

The MLP took 2740 secs to train and the hidden state computation just took a few seconds more. The results of clustering the 35-dimensional hidden state activations from **Net 1** corresponding to the 60,000 training samples of MNIST dataset are shown in Fig. 4. The clusters are, in general, observed to be linear separable from Fig. 4, and the k-Nearest Neighbor (kNN) classifier is expected to perform well under the circumstances. A 10×35 training vector comprising of the ten centroids is compiled and the Euclidean distance of each training and test sample from the ten centroids is used as the 10-dimensional feature vector for learning an SVM classifier, the F1-score of which is observed to be 0.95 from Table I. Direct classification using k-NN classifier leads to a marginal increase of scores to 0.96.

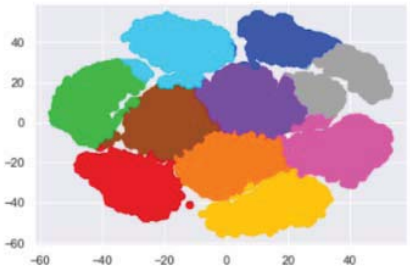
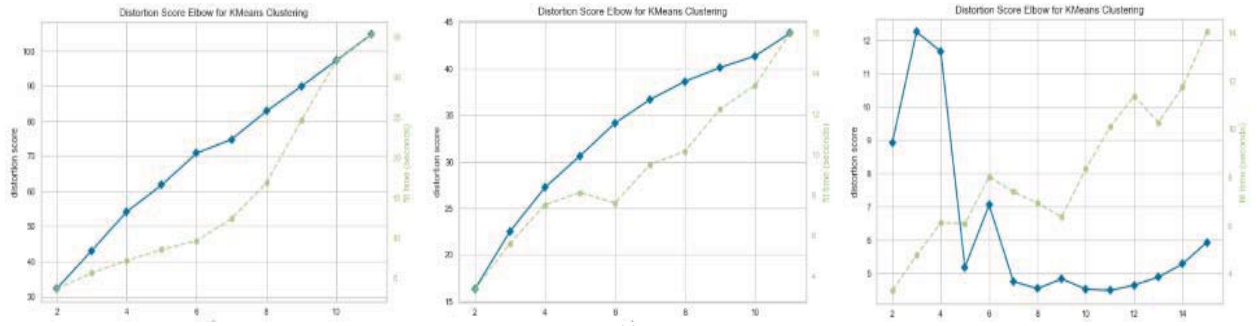


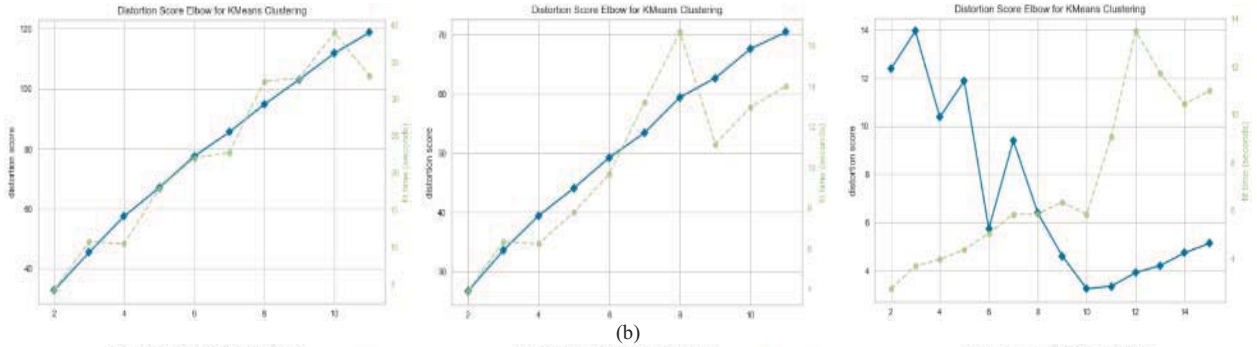
Fig. 4. The clustering of hidden state activations from **Net 1** after feature dimension is reduced to 2 by feature-reduction method t-SNE

The results of k-means Elbow test on **Net 2** are shown in the form of graphs in Fig. 5 (a-e), for the clusters of hidden states derived from the five trained (patch-based) classifiers respectively. The third hidden layer is observed to be interpretable (in all cases) by forming well defined clusters, as observed from the last column in Fig. 5, and therefore is used for extracting hidden representations for the patch-based learning. Whereas hidden layers 1 and 2 form smooth curves with no distinct inflection points and hence are not interpretable. The results in Table I indicate that our method competes favorably with the state-of-the-art, with the highest F1-score of 0.9711 achieved for our patch-based approach (**Net 2**) + SVM, as observed from the last row in Table I. The scores for k-NN classifier are slightly lesser at 0.9676. The comparison amongst the digit accuracies for this case is shown in Table II. Accuracies are noted to be high for '0', '1' and '6' numerals.

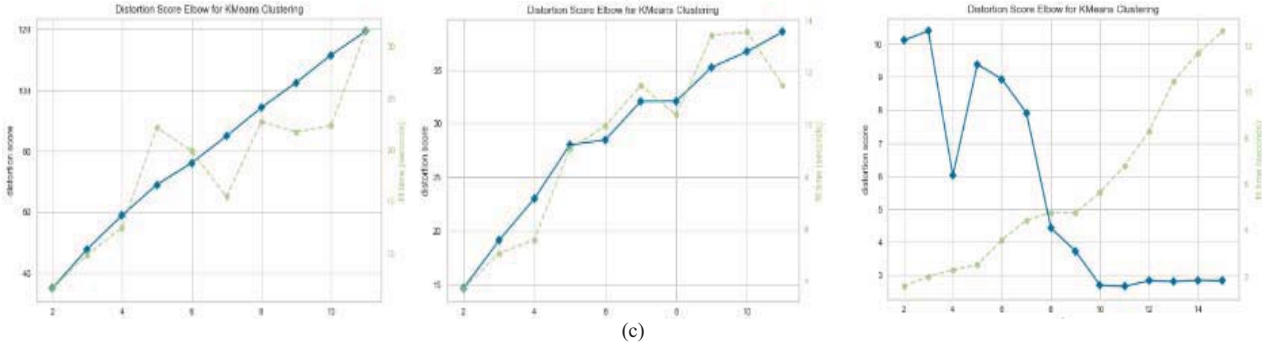
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2020 at 05:42:43 UTC from IEEE Xplore. Restrictions apply.



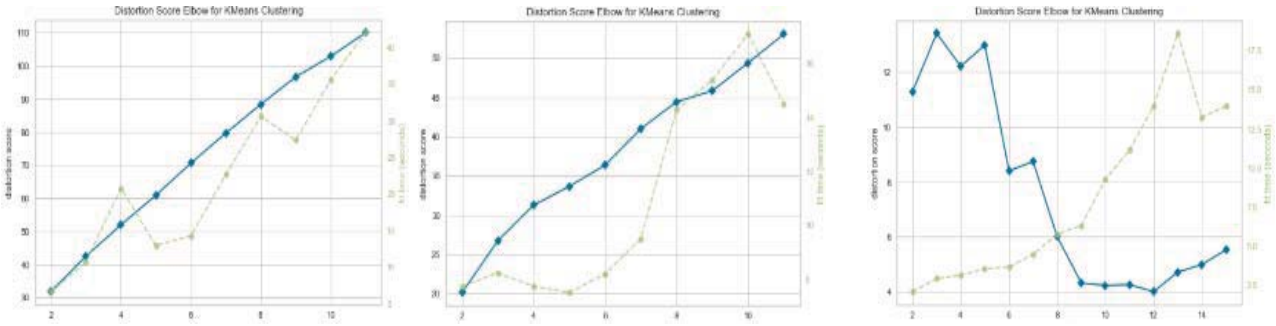
(a)



(b)

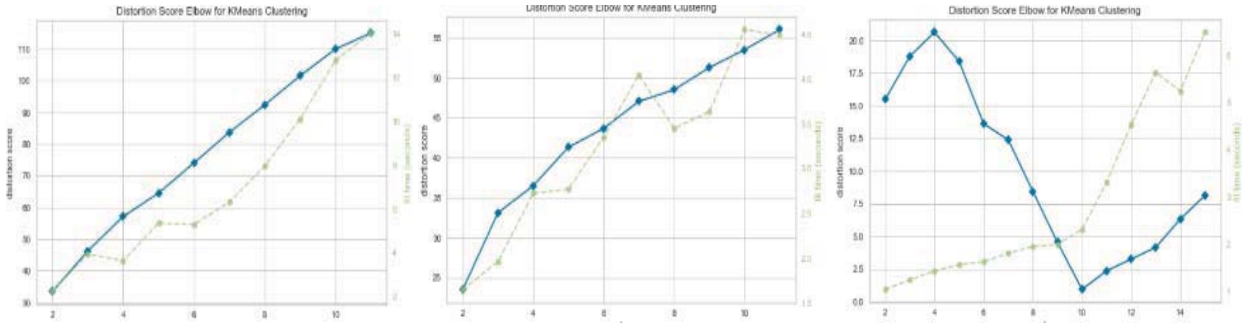


(c)



(d)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2020 at 05:42:43 UTC from IEEE Xplore. Restrictions apply.



(e)

Fig.5. The kmeans Elbow graphs for determining the interpretable hidden layer (*Clustering results from Left to right: Hidden state activations of Hidden Layer 1; Hidden state activations of Hidden Layer 2; Hidden state activations of Hidden Layer 3*), shown for the five image patch layouts [(a)-(e)]

TABLE I. THE OUTPUT PERFORMANCE SCORES FOR MNIST DATASET (PROPOSED STRUCTURE LEARNING METHOD- HIGHLIGHTED IN GRAY)

Dataset	Method	Training Feature dimension	F1-score
MNIST	Deep Structured Energy Based Models (Zhai *et al.,* 2016) [30]	60000 x784	0.9689
MNIST	Positive-Generative Adversarial Network (Chiaroni *et al.,* 2018) [35]	60000 x784	0.97
MNIST	Rank Pruning (Northcutt *et al.,* 2017) [36]	60000 x784	0.97
MNIST	Direct MLP classifier (**Net 1**)	60000 x 784	0.9477
MNIST	Direct Naïve Bayes classifier	60000 x784	0.5658
MNIST	Hidden state clustering (**Net 1**) + distance from 10 centroids + SVM	60000 x 10	0.95
MNIST	Hidden state vector (**Net 1**) + kNN (k=1)	60000 x 35	0.96
MNIST	Patch-based Hidden state vector (**Net 2**) + kNN (k=7)	60000 x35	0.9676
MNIST	Patch-based Hidden state vector (**Net 2**) + SVM	60000 x35	0.9711

TABLE II. A COMPARISON OF 0…9 DIGIT ACCURACIES FOR THE MNIST DATASET [PATCH-BASED HIDDEN STATE VECTOR (NET 2)+ SVM]

Numeral Category	Precision	Recall	F1-score	Samples
0	0.97	0.99	0.98	980
1	0.98	0.99	0.98	1135
2	0.97	0.97	0.97	1032
3	0.96	0.98	0.97	1010
4	0.97	0.97	0.97	982
5	0.97	0.97	0.97	892
6	0.98	0.97	0.98	958
7	0.97	0.96	0.96	1028
8	0.97	0.95	0.96	974
9	0.97	0.96	0.97	1009

V. CONCLUSION

A novel perspective of hidden structure learning for handwritten numeral recognition is proposed in this paper. A shallow multi-layer perceptron with a single hidden layer of 35 neurons is instantiated for the first phase of our experiments. The set of hidden state activations is explored as a viable option for training classifiers. The k-means Elbow method is the litmus test for deciding the interpretability of hidden state activations from the presence of inflection points in the cost function graph. Only the hidden layer that forms separable structures after clustering is used for the feature extraction. Experiments on the benchmark MNIST numeral dataset achieves high accuracies at a reduced feature dimension as compared to the baseline methods. The second phase of our experiments is regarding patch-based learning of numeral images, wherein five separate (and deeper) classifiers train the respective image patches. The hidden state representations derived from the interpretable hidden layer (found to be the third and final hidden layer in the hierarchy) is passed as the feature vector to the SVM and kNN classifiers. A deep network implementation of our interpretable structure learning scheme forms the future scope of our work.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2020 at 05:42:43 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Susan, Seba, and KM Rachna Devi. "Text area segmentation from document images by novel adaptive thresholding and template matching using texture cues." *Pattern Analysis and Applications* (2019): 1-13.

[2] Boukharouba, Abdelhak, and Abdelhak Bennia. "Novel feature extraction technique for the recognition of handwritten digits." *Applied Computing and Informatics* 13, no. 1 (2017): 19-26.

[3] Prasad, Binod Kumar, and Goutam Sanyal. "Novel features and a cascaded classifier based Arabic numerals recognition system." *Multidimensional Systems and Signal Processing* 29, no. 1 (2018): 321-338.

[4] Susan, Seba, and Veerendra Singh. "On the discriminative power of different feature subsets for handwritten numeral recognition using the box-partitioning method." In *2011 Annual IEEE India Conference*, pp. 1-5. IEEE, 2011.

[5] Cireşan, Dan Claudiu, Ueli Meier, Luca Maria Gambardella, and Jürgen Schmidhuber. "Deep, big, simple neural nets for handwritten digit recognition." *Neural computation* 22, no. 12 (2010): 3207-3220.

[6] Qiao, Junfei, Gongming Wang, Wenjing Li, and Min Chen. "An adaptive deep Q-learning strategy for handwritten digit recognition." *Neural Networks* 107 (2018): 61-71.

[7] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86, no. 11 (1998): 2278-2324.

[8] Gader, Paul D., Magdi Mohamed, and Jung-Hsien Chiang. "Handwritten word recognition with character and inter-character neural networks." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 27, no. 1 (1997): 158-164.

[9] Zoph, Barret, Deniz Yuret, Jonathan May, and Kevin Knight. "Transfer learning for low-resource neural machine translation." *arXiv preprint arXiv:1604.02201* (2016).

[10] BakIr, Gökhan, Thomas Hofmann, Bernhard Schölkopf, Alexander J. Smola, and Ben Taskar, eds. *Predicting structured data*. MIT press, 2007.

[11] Liu, Weiwei, Ivor W. Tsang, and Klaus-Robert Müller. "An easy-to-hard learning paradigm for multiple classes and multiple labels." *The Journal of Machine Learning Research* 18, no. 1 (2017): 3300-3337.

[12] Wang, Kai, Zhaoyan Ming, and Tat-Seng Chua. "A syntactic tree matching approach to finding similar questions in community-based qa services." In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 187-194. ACM, 2009.

[13] Chen, Liang-Chieh, Alexander Schwing, Alan Yuille, and Raquel Urtasun. "Learning deep structured models." In *International Conference on Machine Learning*, pp. 1785-1794. 2015.

[14] Liu, Weiwei, and Ivor Tsang. "On the optimality of classifier chain for multi-label classification." In *Advances in Neural Information Processing Systems*, pp. 712-720. 2015.

[15] Collins, Michael. "Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms." In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 1-8. Association for Computational Linguistics, 2002.

[16] Tickle, Alan B., Robert Andrews, Mostefa Golea, and Joachim Diederich. "The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks." *IEEE Transactions on Neural Networks* 9, no. 6 (1998): 1057-1068.

[17] Susan, Seba, and Mayank Dwivedi. "Dynamic growth of hidden-layer neurons using the non-extensive entropy." In *2014 Fourth International Conference on Communication Systems and Network Technologies*, pp. 491-495. IEEE, 2014.

[18] Frasconi, Paolo, Marco Gori, Marco Maggini, and Giovanni Soda. "Representation of finite state automata in recurrent radial basis function networks." *Machine Learning* 23, no. 1 (1996): 5-32.

[19] Biran, Or, and Courtenay Cotton. "Explanation and justification in machine learning: A survey." In *IJCAI-17 workshop on explainable AI (XAI)*, vol. 8. 2017.

[20] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.

[21] Hou, Bo-Jian, and Zhi-Hua Zhou. "Learning with Interpretable Structure from RNN." *arXiv preprint arXiv:1810.10708* (2018).

[22] http://yann.lecun.com/exdb/mnist/ (*Online: Last accessed on 02/06/2019*)

[23] Aly, Saleh, and Ahmed Mohamed. "Unknown-Length Handwritten Numeral String Recognition Using Cascade of PCA-SVMNet Classifiers." *IEEE Access* 7 (2019): 52024-52034.

[24] Salakhutdinov, Ruslan, and Geoff Hinton. "Learning a nonlinear embedding by preserving class neighbourhood structure." In *Artificial Intelligence and Statistics*, pp. 412-419. 2007.

[25] Hinton, Geoffrey E., and Richard S. Zemel. "Autoencoders, minimum description length and Helmholtz free energy." In *Advances in neural information processing systems*, pp. 3-10. 1994.

[26] Srivastava, Nitish, Elman Mansimov, and Ruslan Salakhutdinov. "Unsupervised learning of video representations using lstms." *arXiv preprint arXiv:1502.04681* (2015).

[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] Jia, Yangqing, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. "Caffe: Convolutional architecture for fast feature embedding." In *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675-678. ACM, 2014.

[29] Gomez, Lluis, Anguelos Nicolaou, and Dimosthenis Karatzas. "Improving patch-based scene text script identification with ensembles of conjoined networks." *Pattern Recognition* 67 (2017): 85-96.

[30] Zhai, Shuangfei, Yu Cheng, Weining Lu, and Zhongfei Zhang. "Deep Structured Energy Based Models for Anomaly Detection." In *International Conference on Machine Learning*, pp. 1100-1109. 2016.

[31] Susan, Seba, and Ankit Kumar. "Auto-segmentation using mean-shift and entropy analysis." In *2016 3rd International conference on computing for sustainable global development (INDIACom)*, pp. 292-296. IEEE, 2016.

[32] Susan, Seba, and Monika Sharma. "Automatic texture defect detection using Gaussian mixture entropy modeling." *Neurocomputing* 239 (2017): 232-237.

[33] Ng, Andrew. "Clustering with the k-means algorithm." *Machine Learning* (2012).

[34] Susan, Seba, and Gitin Kakkar. "Decoding facial expressions using a new normalized similarity index." In *2015 Annual IEEE India Conference (INDICON)*, pp. 1-6. IEEE, 2015.

[35] Chiaroni, Florent, Mohamed-Cherif Rahal, Nicolas Hueber, and Frédéric Dufaux. "Learning with a generative adversarial network from a positive unlabeled dataset for image classification." In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1368-1372. IEEE, 2018.

[36] Northcutt, Curtis G., Tailin Wu, and Isaac L. Chuang. "Learning with confident examples: Rank pruning for robust classification with noisy labels." *arXiv preprint arXiv:1705.01936* (2017).

[37] Syakur, M. A., B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto. "Integration k-means clustering method and Elbow method for identification of the best customer profile cluster." In *IOP Conference Series: Materials Science and Engineering*, vol. 336, no. 1, p. 012017. IOP Publishing, 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 22,2020 at 05:42:43 UTC from IEEE Xplore. Restrictions apply.

Journal of the Brazilian Society of Mechanical Sciences and Engineering (2020) 42:302
https://doi.org/10.1007/s40430-020-02375-6

TECHNICAL PAPER

Material removal analysis of hybrid EDM-assisted centrifugal abrasive flow machining process for performance enhancement

Parvesh Ali[1] · R. S. Walia[2] · Qasim Murtaza[1] · Ranganath Muttanna Singari[1]

Received: 4 December 2018 / Accepted: 21 April 2020
© The Brazilian Society of Mechanical Sciences and Engineering 2020

Abstract
Finishing of metallic machine components is a prime requirement for the better performance and longer product life cycle. To get highly finished machine components, a number of conventional and non-conventional finishing processes have evolved in recent times to overcome constraints due to shape and properties of materials. This paper discusses a new hybrid technique, Thermal additive centrifugal abrasive flow machining (TACAFM) which is a combination of centrifugal force-assisted abrasive flow machining and electrical discharge machining (EDM) process. In this process abrasive particles move coaxially with workpiece and also rotate inside the hollow workpiece with the help of a rotating system and also generated thermal effect as in EDM process. Due to the movement of abrasive particles along with rotation, Coriolis effect plays a major role in determining the position of abrasives at any time. Therefore, in the present study, Coriolis component effect is incorporated in the mathematical model to predict the material removal in TACAFM. Also, its parameters, viz current, abrasive concentration, pressure, duty cycle, and rotation of electrode, were optimized using response surface methodology. A simulation model is also presented using Ansys® 15 software to analyze the effect of temperature around the work surface when varying gap between the electrode and work surface along with rotational speed of electrode. Experimental results show a good agreement with the mathematical model. The experiments conducted on the developed process show average 44.34% improvement in material removal and average 39.74% improvement in surface finish, respectively.

Keywords Current · Finishing · Rotation · Thermal spark mechanism · Hybrid machining

1 Introduction

Due to the rapid development in finishing technology, there is a high demand of accurate and precise dimension components in the industry. Even a small scratch can be responsible for failure of a component in dynamic loading as the scratch becomes stress raiser. Recently, finishing industry has been using conventional finishing processes such as honing and lapping, which restricts the workpiece to a particular shape. To overcome the shape factor, there is a need to develop non-conventional finishing processes such as abrasive flow machining (AFM). AFM process has an edge over the conventional processes because it uses a semisolid media such as combination of polymer, gel and abrasives which flow at high pressure through the restrictive path and provide surface finishing by removing the material in the form of micro/nano chips. The non-conventional finishing process is being used for nano finishing of the complex machined geometry of workpiece surfaces that cannot be obtained through conventional finishing technique. Jain et al. [1] reported that the abrasive laden media used for finishing was viscous enough to hold the abrasive particles together for abrasion mechanism in AFM process.

Sankar et al. [2] stated that AFM process removes the material in the range of 1–10 μm. However, large imperfections in the material cannot be corrected by this process because it removes low amount of material uniformly from the workpiece surface. Therefore, AFM process has a limitation of low material removal that leads to the development of new hybrid techniques such as clubbing AFM

Technical Editor: Lincoln Cardoso Brandao.

✉ R. S. Walia
waliaravinder@yahoo.com

✉ Qasim Murtaza
qasimmurtaza@gmail.com

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

[2] Department of Production and Industrial Engineering, Punjab Engineering College, Chandigarh, India

Published online: 15 May 2020

process with other conventional and non-conventional processes in order to improve the overall productivity of developed process. Walia et al. [3] provided a centrifugal motion to the media while it flowed through the workpiece surface and found the centrifugal motion increased the number of dynamic abrasive particles, resulting in the increase of contact area between workpiece and abrasive media. Kim and Kim [4] used AFM finishing process to analyze the deburring effect on the surface of spring collets made of chrome molybdenum and concluded that high viscous media gave better deburring effect in comparison with medium and low viscous media. Singh and Shan [5] provided magnetic effect around a hollow workpiece surface and used magnetic abrasive particles in the media that contact the workpiece surface easily and reported increase in material removal. Singh [6] concluded from the simulation and experimental results, that the magnetic field effect is more prominent at low value of extrusion pressure. Sankar et al. [7] incorporated a drill bit in the media flow path which directed the media to flow along the helical passageway in the restrictive area that corresponded with the improvement in active abrasive particles. Dabrowski et al. [8] developed electrochemical-assisted AFM for flat surfaces and obtained a significant change in material removal due to the chemical action. Brar et al. [9] used this hybrid process for cylindrical surfaces and found 1.5 times improvement in the material removal.

Singh and Walia [10] developed a centrifugal magnetic-assisted AFM. The researchers used the combination of magnetic field and different shapes of rod to create a centrifugal action in the flow of media and got best results at 0.3 Tesla of magnetic field. Vaishya et al. [11] used the combination of electrochemical and centrifugal effect during finishing and found that it saves 20–30% time in comparison with other AFM techniques. Gorana et al. [12] proposed a mathematical model to calculate the force exerted on the single abrasive particle during the finishing process and stated that abrasive concentration and forces acting on abrasive particles affect deformation of material. Mittal et al. [13] performed finishing on EDM-generated surface using AFM process and found that material of the product and extrusion pressure were critical parameters in the finishing process. Wang et al. [14] designed a new helical passageway to increase the media flow paths and found that the gap between the helical edge and the workpiece surface, and the number of helical turns have a major role in the material removal. Kishan et al. [15] conducted experiment on unidirectional AFM set up and explained that the media viscosity and finishing time gave better performance during machining. Venkatesh et al. [16] found that media flow in axial direction is better in case of ultrasonic-assisted AFM in comparison with the conventional AFM process. Marzban et al. [17] applied a spin motion to the rotating workpiece and found significant change in material removal. Fu et al. [18] concluded that media streamline follows an irregular path near the trailing edge during the finishing process.

Jain et al. [1] calculated the stresses generated on the workpiece surface during the flow of media using the finite element method. It was observed that stresses generated on the workpiece surface and media properties did not change with respect to time and space. Jain et al. [1] also concluded that material removal increases with the abrasive concentration and piston pressure; however, abrasive particle mesh size has an adverse effect on the material removal. Jain et al. [19] used neural network approach for modeling and selecting the optimum input AFM parameters. The researchers compared optimized results of AFM process using neural network and GA method and observed close relation between them. Stief and Haan [20] concluded from the model that when the abrasive particles come in contact with the finishing surface, a force is developed between both of them. This force develops stresses on the abrasive particles. Rajeshwar et al. [21] developed a model to analyze the media flow characteristics and found a consistent relation between the media characteristics and the simulation results. Walia et al. [22] concluded from the simulation results that the resultant pressure was maximum at the entry and minimum at the exit point, however media velocity was maximum near the Centrifugal force generating rod and further decreased with the media flow. Wani et al. [23] predicted the surface roughness of the workpiece in magnetic abrasive flow finishing process and got better surface roughness of the workpiece. Das et al. [24] stated that higher pressure produced a smaller plug flow region in magnetorheological abrasive flow finishing (MRAFF) for a particular magnetic field. Ali et al. [25] finished brass workpiece using CNT particles in addition with alumina as abrasive particles in AFM process and developed a three-dimensional model using Ansys® and Fluent® software to identify the critical condition of parameters during finishing. It was found that material removal of the process was enhanced using sharp CNT particles. Also, it was found that velocity of the media was maximum at the center where abrasive particles contacted the workpiece surface. Singh et al. [26, 27] enhanced the material removal rate of EDM process by providing ultrasonic vibration in the flushing region. This caused easily removal of debris from the machining area which increased the material removal rate. The author also performed a finite element analysis of the developed process and compared it with the EDM results. The result showed that pressure was maximum under the tool tip during the upward movement of workpiece.

Although AFM process gives good finishing, its material removal is low. In this research paper an attempt was made to club thermal spark mechanism with the CFAAFM process to enhance the material removal and overall process productivity. This new hybrid process developed is termed

as thermal additive centrifugal AFM (TACAFM) process. In conventional AFM process only, axial force is responsible for the material removal. When clubbing centrifugal motion to the media of AFM, one additional force comes in play. Therefore, in centrifugal-assisted abrasive flow machining (CFAAFM) both axial and radial forces are involved in material removal process. Axial force acts along the media flow direction and radial flow acts perpendicular to the surface. Total force exerted by the abrasive particles will be the resultant of axial and radial forces. The developed TACAFM process is a combination of CFAAFM process and EDM process. In this process abrasive particles are moving coaxially with workpiece and also rotating inside the hollow workpiece with the help of a rotating system. Due to movement of abrasive particles along with rotation, Coriolis effect plays a major role in determining the position of abrasives at any time. Whenever a mass moves in rotating system, it is subjected to a Coriolis force which acts perpendicular to the direction of motion and axis of rotation. Therefore, two types of forces (radial and Coriolis) act perpendicular to the surface and are responsible for the material removal mechanism in TACAFM process. Due to the Coriolis effect, abrasive particles can impact with a larger force and different attack angle over the finishing surface. Most researchers have not considered the effect of Coriolis force during the finishing process in AFM; however, it is an important effect which should be considered for the better response characteristics (i.e. material removal and % improvement in surface finish). This paper also presents a mathematical model for calculating the material removal considering the Coriolis effect in Thermal additive Centrifugal AFM process.

2 Principle of thermal additive centrifugal abrasive flow machining process

Researchers have hybridized AFM process with other processes to overcome its limitation of low material removal. In AFM process a huge amount of forces are involved in removing the material from the surface and a lot of energy is lost due to friction and restriction in the path of flowing media. This motivated authors to develop a new process which can reduce the energy loss and forces involved during the finishing process. This paper introduces a new hybrid technique, i.e., thermal additive centrifugal abrasive flow machining (TACAFM) which uses the thermal spark mechanism along with the conventional material removal mechanism of AFM process. The TACAFM experimental setup included the EDM principle in addition to CFAAFM process. For the spark generation in EDM process, a potential difference is applied between the rotating electrode and the workpiece with a non-conductive environment through abrasive laden media (mixture of media with kerosene based oil) between them. For regular spark generation, gap between both the poles (rotating electrode and workpiece) should be less than 0.25 mm.

Due to potential difference between both the poles, an electrical field is being established. When this electric field is being established between both poles, electrostatic forces develop on the free electrons. The electrons emitted from the rotating electrode accelerate towards the workpiece surface through the abrasive laden media due to repulsive nature of similar polarity. When the electron starts moving toward the workpiece surface, they collide with the media molecules and cause ionization of the media molecules as shown in Fig. 1b. As the electrons further accelerate, due to collision more positive ions and electrons are being generated. This cyclic process increases the concentration of positive and negative ions in the media between the rotating electrode and workpiece surface. The concentration of these ions becomes so high that the matter existing in the channel can be characterized as plasma as shown in Fig. 1b.

The positive ions start flowing towards rotating electrode and negative ions target to the workpiece surface. The movement of ions and electrons can be observed as a spark between the electrode and workpiece gap. The spark generated in the TACAFM process melts the workpiece surface, enabling abrasive particles to take away the softened material easily. The rotation of either electrode or workpiece is necessary to maintain uniformity in shape inside the hollow cavity. With the flow of media, if there is a contact between both the poles, then a large amount of current discharge occurs that removes the material from both the poles and reduces the force required for indentation of abrasive particles.

During the generation of sparks in the gap, the material is eroded from both poles (rotating electrode and workpiece surface), which can change the gap between the poles. This will affect the frequency of spark and minimizes the thermal effect developed over the surface. Regular electrode changes are therefore necessary to maintain the required gap between the rotating electrode and workpiece surface. In TACAFM process evaporation, melting and abrasion are the principal techniques for material removal. As the spark is generated between the rotating electrode and workpiece surface, oxide layers start forming over the surface due to the presence of atmospheric gases in the finishing zone. The oxide layers developed over the surface leads to poor electrical conductivity of the workpiece. This could hamper the machining process. Therefore, continuous flow of media is necessary to maintain almost oxygen-free environment in the finishing zone.

The spiral shape of the rotating electrode performs two functions, first, preserving the necessary gap between the tip of the rotating electrode and workpiece surface and second, allowing the media to pass through the gaps between the

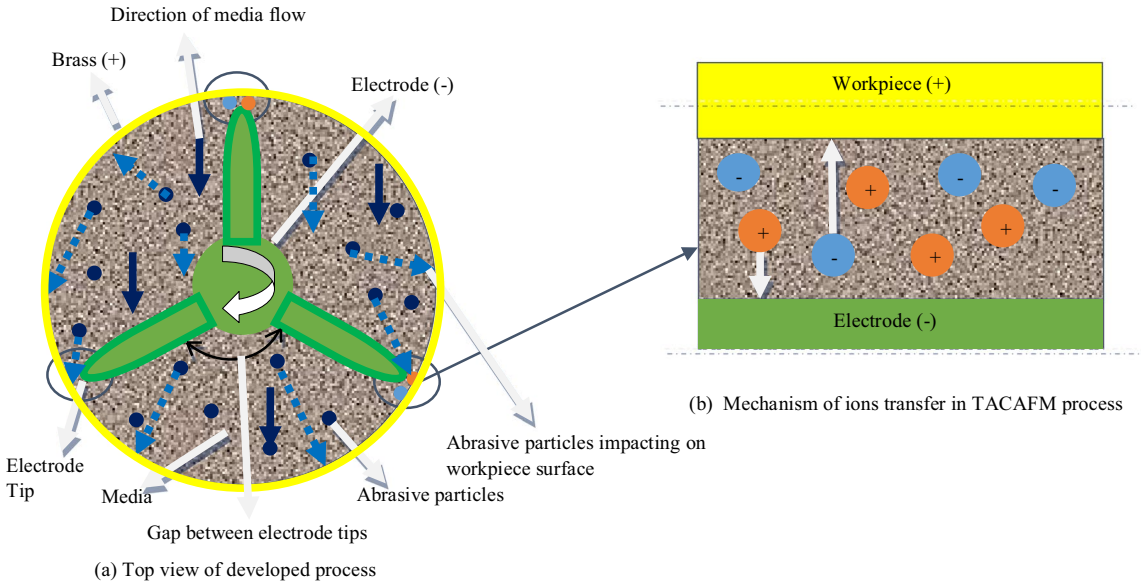


Fig. 1 Principle of thermal additive centrifugal AFM process

tips of the electrode and between the electrode tip and inner workpiece surface. During the electrode rotation, surface material melts in a uniform way and abrasive particles in the media are also continuously participating in the material removal mechanism by interacting with the whole inner surface. Thus, material removal will occur on both inside and outside portion of the discharge gap.

3 Experimental setup

The arrangement of the developed experimental setup of TACAFM process is as shown in Fig. 2. The experimental setup included a fixture for holding the workpiece and guiding the media from one media cylinder to the other, gears for rotating the electrode, bearings, three-phase induction motor, power drive to control the rotation of the electrode, power supply system and manual analog controller to generate spark on maintaining the gap between the rotating electrode and workpiece surface. The power supply system transforms the alternating current of the main supply into the pulsed direct current required for producing the spark. It also senses the voltage between the electrode and workpiece that helps the controller to sense the gap between both the poles. The negative electrode was made to rotate through gears to maintain the uniformity in work surface. In this experimental set up, the driving motor was three-phase induction motor with rpm-controlled AC power drive, attached with a driving gear made of mild steel on the motor shaft. The rotating electrode was attached through aluminum gear which was further rotated through two intermediate gears between the three-phase induction motor and rotating electrode. One of the intermediate gears was made of Metalon which is non-conductive in nature. The non-conductive gear stops the current from reversing back to the motor. The fixture was divided into three parts as shown in Fig. 2iv. It included bearings, gears for electrode rotation, workpiece, electrode and terminals for power supply. The fixture was made of Nylon, consisted two terminals for positive and negative power supply. Workpiece was attached to the positive power supply, while rotating gear was attached to the negative supply. The fixture was also used to hold the workpiece and to guide the media. For spark generation, AFM setup was integrated with EDM power supply and EDM controller as shown in Fig. 2i. For pulse power supply manual type of analog controller (25 Ampere) was used. The function of the controller is to supply the current whenever the gap is maintained between both the poles for producing the spark. The function and specification of the components used in the TACAFM process are as follows-

Components	Function and specifications
Hydraulic power pack	It causes the back and forth movement of piston in the hydraulic cylinder. It consists of motor, reservoir, filter and hydraulic pump along with accompanying hydraulic circuit Maximum working Pressure = 210 kg/cm^2

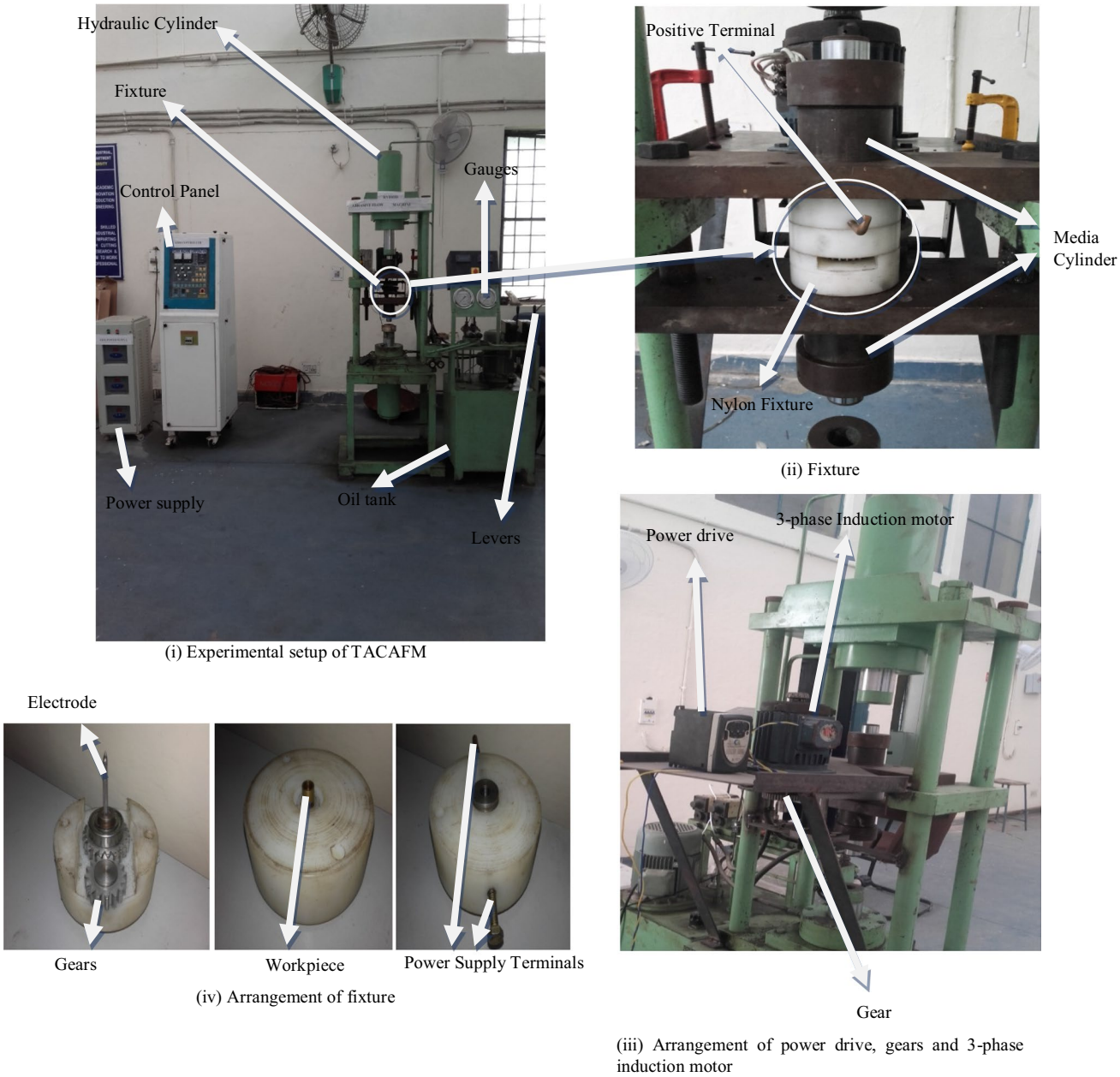


Fig. 2 Experimental setup for thermal additive centrifugal AFM process

Components	Function and specifications
Hydraulic cylinders	In the AFM setup, there are two vertical cylinders which are co axially opposite to each other. In this the piston moves to and fro due to pressure difference in the cylinder barrel. The barrel is closed on one side by cylinder bottom and other end by cylinder head called as gland. The cylinder acts as a mechanical actuator by driving the piston through the action of a pressurized hydraulic fluid to generate a unidirectional force Hydraulic cylinder Bore diameter = 130 mm, Hydraulic cylinder Stroke = 96 mm

Components	Function and specifications
Media cylinders	The media cylinder consists the mixture of the gel, polymer and the abrasive particles which is forced to flow through the hollow workpiece Max. Media Flow volume = 290 cm^3
Fixture	The fixture is made of Nylon. It holds the workpiece and causes the media to flow through the workpiece. The fixture is made of Nylon because it has good wear properties

Components	Function and specifications
Power drive	Power drive is a device that controls the speed of an electrical motor. It controls the speed by changing the frequency of electrical supply to the motor. The three-phase voltage in the electrical grid connected to motor creates a rotating magnetic field in it Power = 0.75 kW/1 HP, Input = AC 1 Ph, 200–240 V
Power supply and control panel	The function of the power supply is to provide correct voltage, frequency and current to the load from a source of electric current. Controller is used to sense the gap between both the poles Current = 25 A, Power Supply = 415 V, 3 Ph, 50 Hz, Power = 3 kV A
Lever	Lever controls the movement of piston in both upward and downward direction
Three-phase induction motor	The induction motor rotates the driving gear for power transmission Power = 1 HP, RPM = 800
Gear set	Gear set used to rotate the electrode inside the hollow cavity
Power supply terminals	Through fixture, terminal was connected to electrode and workpiece for supplying the power to both poles

4 Mathematical modeling of TACAFM process for material removal

In AFM process, abrasive particles in the media pass through the restrictive path with a high extrusion pressure and perform micro cutting action. A large amount of force is exerted by the abrasive particles for indentation on the workpiece surface. The developed TACAFM process uses thermal spark mechanism, which melts the workpiece surface uniformly and soft work surface material can be easily carried by the abrasive particles in the media. This reduces the force required for abrasive particles indentation which improves the productivity of AFM process. While performing the finishing, abrasive particles in the media are subjected to Coriolis force due to sliding of abrasive particles on the rotating electrode. In this paper, authors report a developed mathematical model for the Coriolis force taken into account along with other forces on abrasive particles for indentation and material removal.

The assumptions used in this mathematical model are as follows:

- The plastic deformation of abrasive particle cutting edge is very small, hence neglected.
- Each active abrasive particle participating in indentation will perform same penetration depth.
- All the abrasive particles having spherical shape produce hemispherical crater during indentation.
- Heat flow in abrasive media has a steady state, one dimensional, uniform heat generation and constant thermal conductivity.
- Effect of thermal process is assumed to be uniform over inner surface of workpiece due to the rotation of electrode.

4.1 Coriolis force effect on abrasive particle

Total acceleration of a point with respect to some other point in a rigid link is the vector summation of its normal and tangential components. This is applicable when the distance between the two points is fixed while considering the relative acceleration of the two points on a moving rigid link. When the distance between the two points is changed i.e. if the second point slides, the total acceleration will contain one additional component called as Coriolis acceleration component. Figure 3 shows the top view of electrode which may be considered as a link and abrasive particles sliding on the electrode surface. The shifting of the abrasive particle on the link is as shown in Fig. 3.

Consider an abrasive particle B on a rotating electrode (link MN). When the link MN is rotating in anticlockwise direction with an angular velocity of 'ω', the abrasive particle will move outward with a linear sliding velocity 'v' as shown in Fig. 3. Let the angle turned by the link MN in time dt be $d\theta$ which occupies the new position MP with the abrasive particle at position E. The movement of abrasive particle from B to E is due to the following effect.

- B to C due to the angular velocity ω of the link MN.
- C to D due to the outward velocity 'v' of the abrasive particle.
- D to E due to the acceleration perpendicular to the link, i.e., Coriolis acceleration.

In TACAFM process, electrode is rotating to provide centrifugal effect and abrasive particles are sliding on rotary electrode which is seen to develop similar effect as Coriolis acceleration. Total acceleration of abrasive particles during the finishing process in TACAFM is equal to the acceleration of abrasive particles due to extrusion pressure provided for indentation considering the thermal effect and Coriolis acceleration due to sliding and rotation of abrasive particles in rotating electrode.

The effect of Coriolis force on the abrasive particles is shown in Fig. 4 at different instant of time. Figure 4i show the top view of electrode rotating at an angular velocity of ω inside the hollow work surface. Considering the abrasive particle at a distance of z from the center of rotation at $t = t$ in the z axis direction, the tangential velocity will be

perpendicular to the radius of rotation at any point during the rotation of abrasive particle. After time $t = t + \Delta t$, the position of the abrasive particle is changed which can be seen in Fig. 4ii. If the abrasive particles are affected only by the rotation of electrode, the position of abrasive particle will be in the locus of the electrode rotation. But the position of abrasive particle is shifted and it follows both translation and rotary motion. This is caused due to the Coriolis acceleration involved during the flow in TACAFM process as abrasive particles are sliding in the rotating electrode. There may be a chance that abrasive particles will move in a circular path with the rotation of the electrode which can be seen in Fig. 4iii.

When the electrode has rotated through an angle $\Delta\theta$, the abrasive will additionally move Δz in downward direction.

At $t = t$, $V_{\tan} = z \cdot \omega$ [refer Fig. 4i]
At $t = (t + \Delta t)$, $V_{\tan} = (z + \Delta z) \cdot \omega$ [refer Fig. 4ii]

$V_{\tan}$ at $t = (t + \Delta t)$ increases due to increase in radius with rotation. This increase in $V_{\tan}$ is a part of Coriolis acceleration.

$$\begin{aligned} a_{\text{coriolis}-\Delta z} &= [V_{\tan\,(\text{at } t=t+\Delta t)} - V_{\tan\,(\text{at } t=t)}]/\Delta t \\ &= [(z + \Delta z) \cdot \omega - z \cdot \omega]/\Delta t \\ &= \Delta z \cdot \omega/\Delta t = \omega \cdot (\Delta z/\Delta t) = \omega \cdot V_z \end{aligned} \quad (1)$$

For a small time interval Δt, $\Delta\theta$ will be small, then the direction of this acceleration will be tangential in x-direction.

$$a_{\text{coriolis}-\Delta z} = \omega \cdot V_{\hat{z}i} \quad (2)$$

The abrasive particle velocity on electrode has changed its direction, though in this case velocity is constant.

For small time interval, $\Delta\theta$ will be small and V_z length approaches to arc length (Fig. 5).

$$\Delta V_z = \Delta\theta \cdot V_z$$

The acceleration for a small change in rotational angle:

$$a_{\text{coriolis}-\Delta\theta} = \Delta V_z/\Delta t = \Delta\theta \cdot V_z/\Delta t = V_z \cdot \omega$$

The direction of acceleration is exactly equal to $a_{\text{coriolis}-\Delta z}$ i.e. in the x-direction.

Thus, $a_{\text{coriolis}-\Delta\theta} = \omega \cdot V_{\hat{z}i}$.

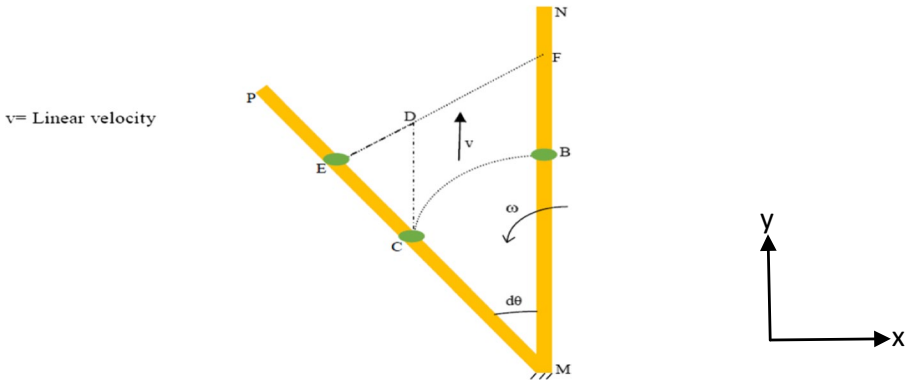


Fig. 3 Top view of abrasive particle sliding on the rotating electrode

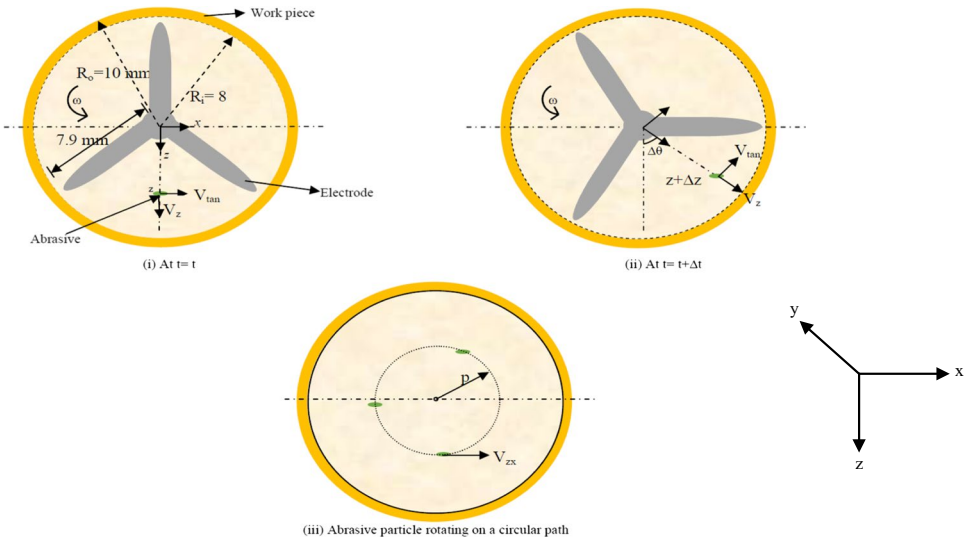


Fig. 4 Top view of media flow path showing the position of abrasive particle in TACAFM process

Total coriolis acceleration is combination of both acceleration i.e.:

$$a_{\text{coriolis}} = a_{\text{coriolis}-\Delta z} + a_{\text{coriolis}-\Delta\theta} = 2V_{zx} \cdot \omega \tag{3}$$

There is a chance that some of the abrasive particle will move on circular path of the electrode. According to this condition, the tangential speed of the abrasive will increase by V_{zx} in comparison with effect of ω alone.

$$V_{\tan} = \omega \cdot p + V_{zx} \quad \left[\text{refer Fig. 4iii}\right] \tag{4}$$

The absolute rotational velocity of abrasive particle will be $\omega_{zx} = (V_{\tan}/p)$ a_{radial} and a_{coriolis} are aligned and direction of both is towards center.

Adding together should produce the normal acceleration.

$$a_{\text{radial}} + a_{\text{coriolis}} = \omega^2 \cdot p + 2\omega \cdot V_{zx} \tag{5}$$

During the flow of media in TACAFM process through the restrictive path, abrasive particles are subjected to three types of forces which include axial force, centrifugal force and Coriolis force. The direction of axial force will be along the axis of the media flow, while centrifugal force and Coriolis force will act in radial direction towards the work surface. The axial force includes forces due to the extrusion pressure of media, weight of the abrasive particles and velocity of the weight of the abrasive particles while radial force includes Centrifugal and Coriolis forces. The resultant force acting on the abrasive particle will be the resultant of axial force and centrifugal force assisted with Coriolis force. These forces cause abrasive particles to hit the workpiece surface with a different approach angle rather than the initial approach angle which can be clearly shown in Fig. 6. The effect of various forces on the abrasive particle during the finishing process is as shown in Fig. 6. These all forces make change in approach angle which increases the number of dynamic abrasive particles coming in contact with workpiece surface, hence increase the material removal.

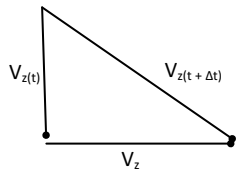
In TACAFM process, spark generated between the electrode and workpiece surface melts the surface material of workpiece which can be easily carried away by abrasive particles. This requires less amount of force for indentation because abrasive particles can easily take away the soften material. The forces acting on the abrasive particles while performing the finishing are represented in Fig. 7.

Marochkin [28] proposed an equation for predicting the normal force subjected to the abrasive particle penetrating the workpiece surface which is given as follows:

$$F = C \cdot A \cdot \sigma$$

C = flow constraint factor, A = projected area of abrasive particles, σ = uniaxial flow stress of the workpiece material.

Fig. 5 Change in V_z's direction



According to Arpaci et al. [29], if spark is produced between the electrode and workpiece surface for time τ, then force developed due to thermal effect in TACAFM process for hollow cylindrical workpiece will be:

$$P_{\text{thermal}} = \frac{2\pi K \cdot \left(T_{\max} - T_{\text{w}}\right) \cdot \tau}{\ln\left(\frac{R_2}{R_1}\right)}$$

where K = thermal conductivity of the electrode, $T_{\max}$ = maximum temperature produced at the tip of the electrode, T_{w} = temperature of the workpiece, R_2 = Outer radius of the workpiece, R_1 = inner radius of the workpiece.

The abrasive particles will move through the restrictive passage due to extrusion pressure of media. The force acting on the dynamic abrasive particles due to extrusion pressure will be

$$= \sum_{\theta_i=\theta_{\text{a}}}^{0} (C \cdot A \cdot \sigma)\text{Sin}\theta_i \cdot N_{\text{d}} \left[\text{refer Fig. 7}\right],$$

where N_{d} = number of dynamic abrasive particles.

In TACAFM process, the generated spark melts the workpiece surface which reduces the force required by abrasive particles for indentation. Then the force required for indentation due to extrusion pressure considering the thermal effect of spark will be

$$= \left[\sum_{\theta_i=\theta_{\text{a}}}^{0} (C \cdot A \cdot \sigma)\text{Sin}\theta_i \cdot N_{\text{d}} - P_{\text{thermal}}\right]$$

(where θ_{a} = Approach angle)

The abrasive particles in the media have some weight which acts towards downward direction. This force also contributes in the axial force provided for indentation.

Therefore, total force developed on the abrasive particles under extrusion pressure and weight considering the thermal effect will be (refer Fig. 8).

$$= \left[\sum_{\theta_i=\theta_{\text{a}}}^{0} (C \cdot A \cdot \sigma)\text{Sin}\theta_i \cdot N_{\text{d}} - P_{\text{thermal}}\right] \pm W \tag{6}$$

where W = weight of the dynamic abrasive particles, (+) sign indicates weight of the abrasive particle along the direction of media flow, (−) sign indicates weight of the abrasive particle opposite the direction of media flow.

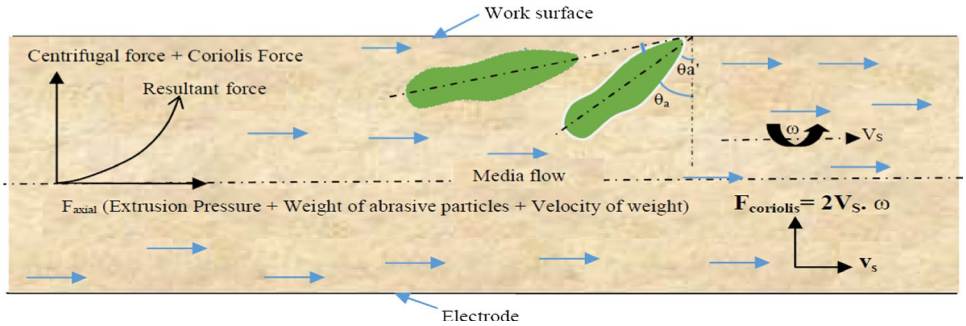
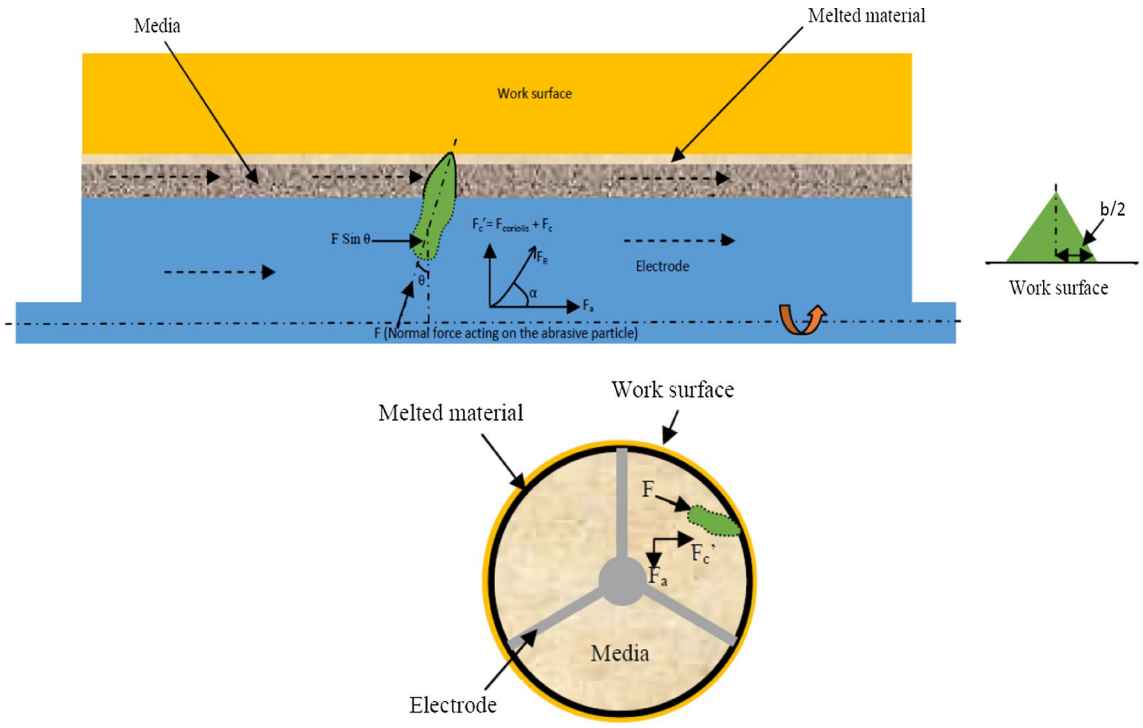


Fig. 6 Effect of coriolis acceleration during the media flow. Where θ_a = Initial approach angle, θ_a' = Change in approach angle due to Coriolis acceleration

Where θ_a= Initial approach angle, θ_a'= Change in approach angle due to Coriolis acceleration



Where θ = Angle of indentation of abrasive particle with the machining surface

Fig. 7 The resultant force involved in the TACAFM process and side view of the process. Where θ = angle of indentation of abrasive particle with the machining surface

Fig. 8 Forces developed on abrasive particles during upward and downward motion of media

Total Coriolis force due to sliding of abrasive particles on the rotating electrode

$$= m \cdot [2V_{zx} \cdot \omega + \omega^2 \cdot p] \quad \text{(from Eq. 5)} \tag{7}$$

where m = total mass of abrasive particles.

This force will be applied by the abrasive particles contacting the workpiece surface. Hence the actual force required for indentation will be provided by the dynamic abrasive particles which are actually in contact with the workpiece surface. Walia et al. [30] proposed an equation for calculating the number of dynamic abrasive particles as given below:

$$N_{\mathrm{d}} = \left[4 \cdot V_a \cdot \left(D \cdot d - d^2\right)/V_o \cdot D^2\right] \tag{8}$$

So, force exerted by dynamic abrasive particles under Coriolis acceleration will be:

$$F_{\mathrm{coriolis}} = 2m \cdot V_{zx} \cdot \omega \cdot N_{\mathrm{d}} + m \cdot \omega^2 \cdot p \cdot N_{\mathrm{d}} \tag{9}$$

F_{a} is the axial thrust which forces the abrasive particles to move along the axis of the workpiece while F_c will apply a normal force on the abrasive particle towards the work surface. Centrifugal force will force the abrasive particles towards the interior surface of the workpiece. Walia et al. [31] reported that the centrifugal force acting on dynamic abrasive particles can be calculated by

$$F_c = 3\pi\mu d_a V_c \cdot N_{\mathrm{d}} \tag{10}$$

(where V_c = velocity of abrasive particle normal to the workpiece axis).

Total force exerted by the abrasive particle in the direction normal to the work surface

$$\begin{aligned} F_c' &= F_{\mathrm{coriolis}} + F_c = (2m \cdot V_{zx} \cdot \omega \cdot N_{\mathrm{d}} \\ &\quad + m \cdot \omega^2 \cdot p \cdot N_{\mathrm{d}} + 3\pi\mu d_a V_c \cdot N_{\mathrm{d}}) \end{aligned} \tag{11}$$

The resultant force developed in thermal additive centrifugal AFM will be

$$\begin{aligned} F_c' &= F_{\mathrm{coriolis}} + F_c = (2m \cdot V_{zx} \cdot \omega \cdot N_{\mathrm{d}} \\ &\quad + m \cdot \omega^2 \cdot p \cdot N_{\mathrm{d}} + 3\pi\mu d_a V_c \cdot N_{\mathrm{d}}) \end{aligned}$$

where α = angle between the resultant and axial force, α represents the direction of the resultant force.

$$F_R = \sqrt{\left[\left\{\left[\sum_{\theta_i=\theta_{\mathrm{a}}}^{0}(C \cdot A \cdot \sigma)\mathrm{Sin}\theta_i \cdot N_{\mathrm{d}} - P_{\mathrm{thermal}}\right] \pm W\right\}\right]^2 \pm (2m \cdot V_{zx} \cdot \omega \cdot N_{\mathrm{d}} + m \cdot \omega^2 \cdot p \cdot N_{\mathrm{d}} + 3\pi\mu d_a V_c \cdot N_{\mathrm{d}})^2} \tag{12}$$

For this case, abrasive particles are assumed to be of spherical shape having diameter d_g, penetration caused by the abrasive particle can be calculated by the equation as given below (refer Fig. 9):

$$\begin{aligned} AC^2 &= BC^2 + AB^2 \\ \left(\frac{d_g}{2}\right)^2 &= r^2 + \left(\frac{d_g}{2} - x\right)^2 \\ r^2 &= d_g * x - x^2 = x(d_g - x) \\ r &= \sqrt{x \cdot (d_g - x)} \end{aligned} \tag{13}$$

from Fig. 9:

$$x = \frac{d_g}{2} - \sqrt{\left(\frac{d_g}{2}\right)^2 - r^2}$$

If r represents the radius of projected area of indentation, x is indentation depth and H_{w} is the workpiece hardness, then resultant indenting force-

$$F_R = H_{\mathrm{w}} \cdot \pi \cdot r^2 \tag{14}$$

According to Jain et al. [1], value of indentation,

$$x = \frac{d_g}{2} - \sqrt{\left(\frac{d_g}{2}\right)^2 - \frac{F_r}{H_{\mathrm{w}} \cdot \pi}} \tag{15}$$

The cross-sectional area of grooved developed by the abrasive particles can be obtained from the geometry:

$$A' = \frac{d_g^2}{4}\sin^{-1}\frac{2\sqrt{x \cdot (d_g - x)}}{d_g} - \sqrt{x \cdot (d_g - x)} \cdot \left(\frac{d_g}{2} - x\right) \tag{16}$$

Volume of material removal by an abrasive particle will be the multiplication of area 'A' and contact length 'L_i' of abrasive particle with the finishing surface.

$$V_{\mathrm{a}} = \left[\frac{d_g^2}{4}\sin^{-1}\frac{2\sqrt{x \cdot (d_g - x)}}{d_g} - \sqrt{x \cdot (d_g - x)} \cdot \left(\frac{d_g}{2} - x\right)\right] \cdot L_i \tag{17}$$

Let N be the number of dynamic abrasive particles per unit contact area. Jain et al. [1] reported that total number of abrasive particles participating in indentation per stroke (N_{s}) can be determined as:

$$N_{\mathrm{s}} = 2\pi \cdot R_{\mathrm{w}} \cdot N \cdot v_{\mathrm{f}} \cdot l_{\mathrm{s}}/v_{\mathrm{p}} \tag{18}$$

where l_{s} shows the stroke length, v_{p} represents piston velocity, v_{f} represents the velocity of flow through the restrictive

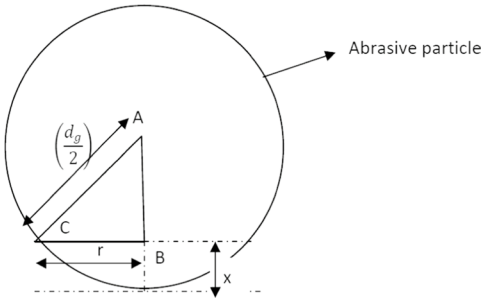


Fig. 9 Abrasive particle indentation on the work surface

(d_g= Diameter of abrasive particle)

(where x = Abrasive Indentation),

(r = abrasive particle projected length)

passage, R_w represents radius of the cylindrical-shaped workpiece. Using continuity equation,

$$\frac{v_f}{v_p} = \frac{R_c^2}{R_w^2} \tag{19}$$

So,

$$N_s = 2\pi \cdot R_w \cdot N \cdot l_s \cdot \frac{R_c^2}{R_w^2} = 2\pi \cdot N \cdot l_s \cdot \frac{R_c^2}{R_w} \tag{20}$$

Volume of material removed in ith stroke (V_i) is:

$$V_i = N_s.\left[\frac{d_g^2}{4}\sin^{-1}\frac{2\sqrt{x\cdot(d_g-x)}}{d_g} - \sqrt{x\cdot(d_g-x)}.\left(\frac{d_g}{2}-x\right)\right]\cdot L_i \tag{21}$$

If R_a^i is the initial surface roughness and R_a^f is the surface roughness obtained after ith stroke.

$$L_i = \left[1 - \frac{R_a^f}{R_a^i}\right]\cdot L_w \tag{22}$$

Total material removal in 'n' number of cycle-

$$\text{Vol} = \sum_{i=1}^{2n} V_i$$

$$\text{Vol} = N_s.\left[\frac{d_g^2}{4}\sin^{-1}\frac{2\sqrt{x\cdot(d_g-x)}}{d_g} - \sqrt{x\cdot(d_g-x)}\cdot\left(\frac{d_g}{2}-x\right)\right]\cdot\sum_{i=1}^{2n}\left[1-\frac{R_a^f}{R_a^i}\right]\cdot L_w \tag{23}$$

Weight of material removal in 'n' number of cycle is:

$$W_{(\text{CFAAFM})} = \rho_w \cdot N_s \cdot \left[\frac{d_g^2}{4}\sin^{-1}\frac{2\sqrt{x\cdot(d_g-x)}}{d_g} - \sqrt{x\cdot(d_g-x)}\cdot\left(\frac{d_g}{2}-x\right)\right]\cdot\sum_{i=1}^{2n}\left[1-\frac{R_a^f}{R_a^i}\right].L_w \tag{24}$$

ρ_w = density of work material.

This equation gives material removal for AFM process assisted with centrifugal forces.

Total material removal = Material removal of CFAAFM + Material removal in EDM mechanism

$$\text{MR}_{\text{TACAFM}} = f(\text{CFAAFM} + \text{EDM})$$

Material removal in EDM is mainly due to the intense localised heating provided by the heat source for a small time frame. This heating leads to the melting and crater formation.

Energy developed by a single spark [32], will be:

$$E_S = V \cdot I \cdot T_{on} \tag{25}$$

where V = supply voltage, I = current, T_{on} = pulse on time.

Some of the spark energy consumes in heating the media and the remaining energy is distributed between the impinging electrons and ions. Thus, the energy available as heating the work surface:

$$E_W = k_1 \cdot E_S$$

It can be assumed that material removal in a single spark is proportional to the spark energy.

$$\text{MR}_{\text{EDM}} \alpha\, E_S \alpha\, E_W$$

So

$$\text{MR}_{\text{EDM}} = k_2 \cdot E_S = k_2 \cdot V \cdot I \cdot T_{on} \tag{26}$$

So total material removal in developed hybrid TACAFM process-

$$\mathrm{MR}_{\mathrm{TACAFM}} = \rho_w \cdot N_s \cdot \left[\frac{d_g^2}{4} \sin^{-1} \frac{2\sqrt{x \cdot (d_g - x)}}{d_g} - \sqrt{x \cdot (d_g - x)} \cdot \left(\frac{d_g}{2} - x \right) \right] \cdot \sum_{i=1}^{2n} \left[1 - \frac{R_a^f}{R_a^i} \right] \cdot L_w + k_2 \cdot V \cdot I \cdot T_{on} \quad (27)$$

The simulation procedure involves determination of resultant force exerted by the abrasive particles on the surface for micro cutting and amount of material removal in TACAFM process. Such findings can be used to correlate to experimental results that can help in improving the productivity and efficiency of the process. The simulation results also help in identifying the critical parameters affecting the material removal mechanism. Equation 12 clearly shows that the resultant force exerted by the abrasive particles is mainly dependent upon uniaxial flow stress of material, intensity of rapid discharge, abrasive mesh size and dynamics number of abrasive particles. Material removal in TACAFM process can be theoretically calculated from Eq. 27. It is clearly evident from the equation that material removal of the developed process can be increased by enhancing the intensity of spark, pulse on time, number of process cycles and contact length. TACAFM process uses spark energy to melt the surface in addition with the abrasion mechanism. This discharge energy is uniformly distributed over the surface due to rotation of electrode. It makes easier for the abrasive particles to take away the molten/semi molten material requiring lesser amount of resultant force. Also, a centrifugal motion is provided to the media using electrode rotation which increase the dynamic motion of abrasive particles in the media and increase the contact region between the abrasive particle and workpiece surface. The increase in material removal can improve the surface integrity of the component which corresponds better functional performance and service life.

5 Modeling of TACAFM process

Spark generated in Thermal additive centrifugal Abrasive Flow Machining process requires less energy to remove material in comparison with the conventional AFM. The temperature produced from the spark will remove material from both the electrodes. So, in this process temperature distribution around the workpiece is major point of concern, when the electrode was either stationary or rotating. A three-dimensional model of thermal additive centrifugal abrasive flow machining was made and imported in ANSYS® 15 software for further analysis. The three-dimensional model includes the rotating electrode, AFM boundary, workpiece and abrasive particles. Figure 10 shows the model of components with their assembly. After importing it into the ANSYS® software, meshing was generated of the size of $6e^{-3}$ m. The inlet and outlet were selected for the analysis part. The properties taken for the analysis is represented in Table 1.

Some assumptions considered during the model analysis are as follows:

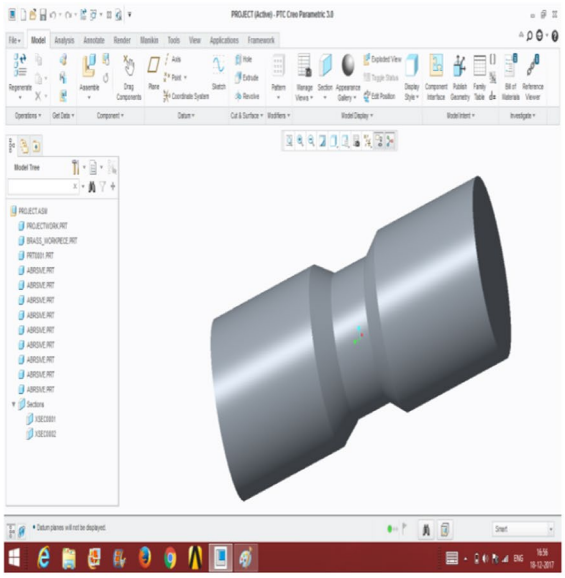
- Model is pressure-based model.
- Analysis is based on the Green Gauss cell based.
- Formulation used is absolute velocity and implicit.
- Media flows in a three-dimensional path and symmetric to the path of hollow surface.
- Media is homogeneous and isotropic.
- Flow is considered as laminar flow and steady in nature.

Meshing was done in ANSYS® 15 software, and for analysis, it was imported in FLUENT® software. Figure 11 shows the fluid body of the TACAFM process.

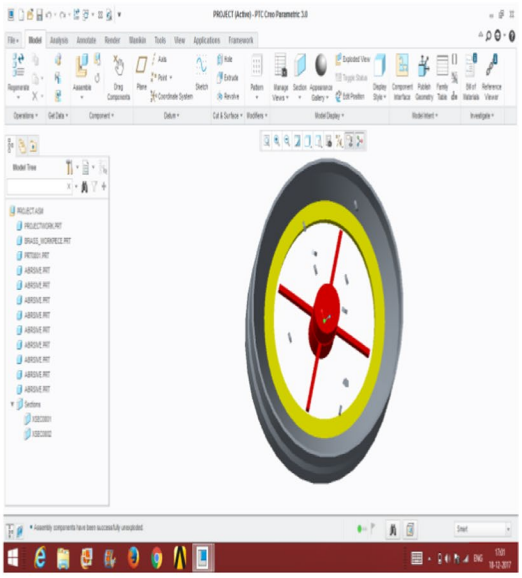
Pressure on one side of the boundary was taken as 40 MPa, named as inlet. Outlet pressure was assumed as 20 MPa. The media flows due to pressure difference created between the two hydraulic cylinders. The gap between the electrode tip and the workpiece surface was being varied under two different conditions of electrode (1) when electrode was stationary and (2) when electrode was rotating one. Electrode rotation was varied to know the temperature distribution around the work surface. Kazimierz et al. [33] stated that the temperature of the plasma channel in EDM was about 8000–12,000 K at 35–50 A of current. Larger value of current increases the material removal and damages the surface integrity. This analysis was done for lower value of current to provide good surface integrity. So, the electrode temperature was taken according to Kansal et al. [34], around 3500 K for analysis and workpiece temperature was taken as 298 K.

6 Optimization of TACAFM process using response surface methodology

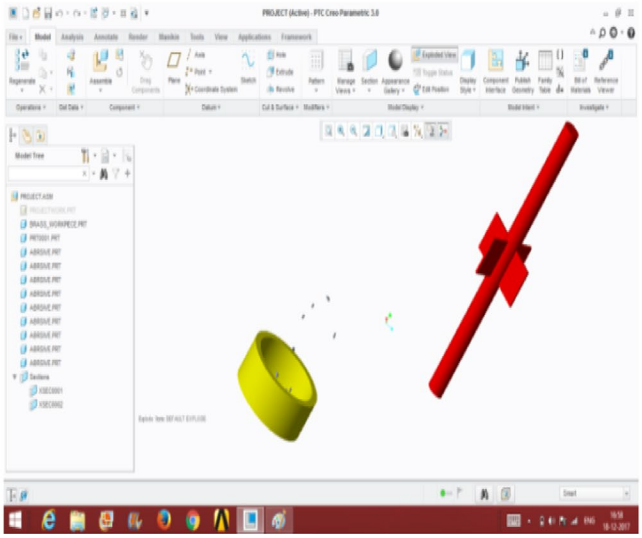
In this design, all the points which are at the same distance from the region center have same standard error. If point (0, 0, …, 0) represents the region centre for which Y and X relation was under study. From any of the experimental result, the standard errors e_r of Y can be defined at any point on the fitted surface. Due to rotatability condition, the value of e_r will be same for all the points which are at the same distance of k from the region centre. As stated by Draper [35], for this condition following equation holds good.



(i) 3- dimensional model of TACAFM setup



(ii) Top View of model



(iii) Electrode, work piece and abrasives used in the model

Fig. 10 Model of components used in TACAFM process

$$x_1^2 + x_2^2 + x_3^2 + \cdots + x_k^2 = k^2 = \text{constant}$$

In the present experimental work, material removal has been considered for evaluating the finishing performance. The experiments were conducted on the TACAFM setup which was shown in Fig. 2. Current, duty cycle, rotation, extrusion pressure, and abrasive concentration were taken as variable and these chosen parameters and their limits are as given in Table 2. The workpiece used for these experiments was brass. Thirty-two experiments were performed as per central composite design. As shown in Table 3, workpiece weight was measured by precision weighing machine (SHIMADZU AUX220) having 0.1 mg resolution. The measurement of material removal was carried out by the difference in the workpiece weight before machining and after machining. Surface roughness was measured by Taylor Hobson precision machine having resolution of 0.01 µm. Surface roughness was taken at five different positions of the workpiece and average of these values was taken for the initial and final value of surface roughness.

Table 1 Properties of the materials

Component	Material	Properties
Electrode	Copper	Density = 8940 kg/m^3 Thermal conductivity (K) = 385 W/m K Specific heat = 376 J/kg K
Workpiece	Brass	Density = 8300 kg/m^3 Poison's ratio = 0.31 K = 109 W/m K Specific heat = 401 J/kg K
Abrasive	Aluminium Oxide	Specific heat = 451 J/kg K K = 12 W/m K Density = 3950 kg/m^3
Media	Polyborosiloxane	K = 0.22 W/m K Density = 985 kg/m^3 Specific heat = 2025 J/kg K

Table 2 Process parameters with their limits

Factors	Levels				
	−2	−1	0	+1	+2
A-Current (Amp)	0	4	8	12	16
B-Duty cycle (fraction)	0.63	0.68	0.73	0.78	0.83
C-Rotation (rpm)	100	150	200	250	300
D-Pressure (MPa)	5	10	15	20	25
E-Abrasive Concentration (fraction)	0.2	0.3	0.4	0.5	0.6

Constant parameters: Polymer to Gel ratio: 1:1, Workpiece Material: Brass, Abrasive type: Al_2O_3, Mesh Size: 180, Media flow volume: 290 Cm^3, Temperature: 32 ± 2 °C

7 Results and discussion

7.1 Validation of mathematical model

In conventional AFM process, material is removed by abrasion mechanism caused by the direct contact of the abrasive particle with the workpiece surface. The material removal of conventional AFM is very low. To overcome this limitation Conventional AFM process was integrated with centrifugal force termed as centrifugal force-assisted AFM process (CFAAFM). In CFAAFM process, centrifugal force was developed due to the rotation of CFG rod. This centrifugal force acts in normal direction to the axial movement of the media towards workpiece surface. In CFAAFM process, total force acting on the abrasive particle was the resultant of centrifugal force and axial force. This process requires lesser number of cycles for getting desired material removal. In TACAFM process material was removed by the thermal spark mechanism which melted the surface material and hence can be easily taken away by the abrasive particles. This process was a combination of CFAAFM and EDM process. The rotation of the electrode generates centrifugal force and abrasive particles are sliding on the rotating electrode which develops Coriolis force. In TACAFM process total force acting on the abrasive particle was the resultant of axial force and centrifugal force associated with the Coriolis force. Figure 12 shows that the material removal in case of TACAFM process decreases with increase in mesh size of the abrasive particles. The reason for decrease in the material removal may be due to decrease in the contact region between the abrasive particle and workpiece surface and less indentation force applied by the abrasive particles. The experimental results of TACAFM process was compared with the conventional AFM process and modeling result of

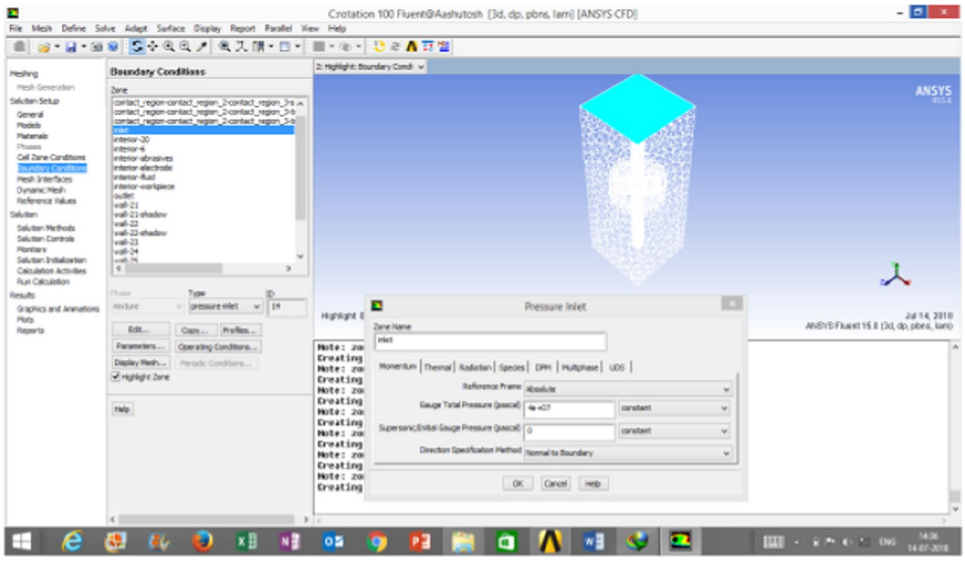


Fig. 11 Fluid body of TACAFM process

Table 3 Central composite design for the measured experimental results and actual factors

Std	Run	Current (A)	Duty cycle	Rotation (rpm)	Pressure (MPa)	Abrasive concentration	Material removal (mg)	% Improvement in R_a
1	3	4	0.68	150	10	0.5	15.54	32.26
2	13	12	0.68	150	10	0.3	27.86	27.56
3	4	4	0.78	150	10	0.3	14.49	25.34
4	12	12	0.78	150	10	0.5	31.24	36.45
5	2	4	0.68	250	10	0.3	18.56	31.46
6	32	12	0.68	250	10	0.5	30.45	37.73
7	15	4	0.78	250	10	0.5	18.57	36.57
8	22	12	0.78	250	10	0.3	36.24	42.35
9	10	4	0.68	150	20	0.3	13.35	29.94
10	5	12	0.68	150	20	0.5	27.54	39.47
11	6	4	0.78	150	20	0.5	22.34	28.91
12	1	12	0.78	150	20	0.3	32.33	36.87
13	27	4	0.68	250	20	0.5	22.23	16.78
14	24	12	0.68	250	20	0.3	28.89	28.18
15	18	4	0.78	250	20	0.3	15.67	33.46
16	17	12	0.78	250	20	0.5	34.41	38.49
17	7	0	0.73	200	15	0.4	8.55	15.76
18	29	16	0.73	200	15	0.4	38.81	29.18
19	11	8	0.63	200	15	0.4	21.12	32.89
20	25	8	0.83	200	15	0.4	27.65	41.13
21	20	8	0.73	100	15	0.4	26.65	28.64
22	23	8	0.73	300	15	0.4	31.25	31.35
23	8	8	0.73	200	5	0.4	24.45	39.13
24	21	8	0.73	200	25	0.4	30.12	34.46
25	14	8	0.73	200	15	0.2	24.46	27.54
26	30	8	0.73	200	15	0.6	29.91	29.75
27	31	8	0.73	200	15	0.4	27.64	27.79
28	19	8	0.73	200	15	0.4	25.56	30.21
29	16	8	0.73	200	15	0.4	28.61	25.42
30	26	8	0.73	200	15	0.4	25.67	23.29
31	28	8	0.73	200	15	0.4	25.58	22.34
32	9	8	0.73	200	15	0.4	27.35	24.46

TACAFM process. The experimental results show 44.34% average improvement in material removal compared to the conventional AFM process. Figure 12 also shows that the experimental results of TACAFM process was slightly lower compared to the modeling result. The modeling results were showing 18.78% average deviation from the experimental results. Figure 13 shows that as the value of current was increased, material removal also increased. The reason for this may be that on increasing the current, intensity of spark was increased and more amount of energy was involved in melting the workpiece surface which gave more material removal. Figure 14 shows the SEM image of the surface obtained from the Centrifugal force-assisted AFM process which shows less material removal from the surface obtained from TACAFM process, on comparing the SEM results obtained after TACAFM process. The heat generated due to spark allowed oxides formation on the workpiece surface as shown in Fig. 15. The high temperature developed due to the heat generation melted the surface material. The melting of the surface material is clearly shown in Fig. 16. Figure 17 shows the spark spots developed on the surface. Figure 18 shows the SEM image of surface produced after finishing through TACAFM process. The experimental results obtained in TACAFM process on increasing the value of current were also shown in Fig. 13, which were slightly lower in comparison with the model results.

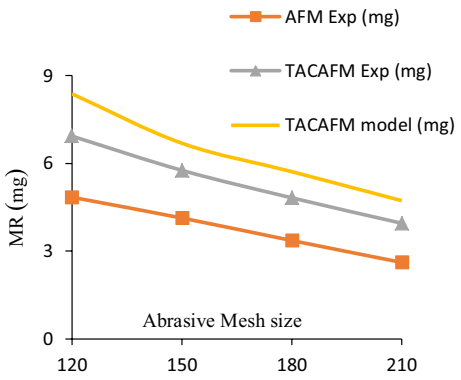


Fig. 12 Effect of abrasive particle mesh size on material removal for conventional and hybrid forms of AFM [at $\rho_w = 8300$ kg/m^3, $R_a^i = 2.2$, $R_a^f = 1.2$, $L_w = 16$ mm, $k_2 = 0.6$, $V = 415$ V, $I = 10$ A, $T_{on} = 35$ µs]

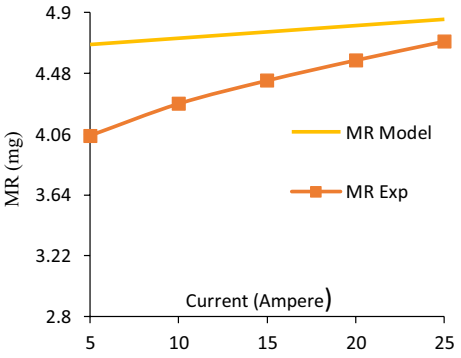


Fig. 13 Effect of current on material removal for TACAFM process [at $\rho_w = 8300$ kg/m^3, $R_a^i = 2.2$, $R_a^f = 1.2$, $L_w = 16$ mm, $k_2 = 0.6$, $V = 415$ V, mesh size = 210, $T_{on} = 35$ µs]

7.2 Effect of Abrasive mesh size and current on percentage improvement in surface finish

Surface roughness was measured by Taylor Hobson precision machine having resolution of 0.01 µm. Surface roughness was taken at five different positions of the workpiece and average of these values was taken for the initial and final value of surface roughness. Figure 19 shows the Taylor Hobson precision machine. Each experiment was repeated for three times and average of their values was taken as the final observed roughness. Figure 20a, b shows the surface roughness graphs of the workpiece before and after the finishing process. From the literature review [36, 37], it was found that abrasive mesh size is a key parameter in Abrasive Flow Machining towards material removal and surface roughness. Current in EDM process also has a greater impact on the material removal which

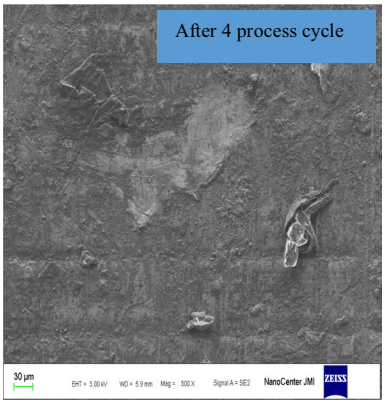


Fig. 14 Microstructure of workpiece surface after CFAAFM process

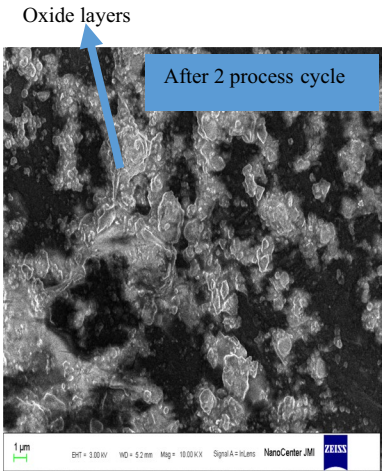


Fig. 15 Oxide layers formation in TACAFM process

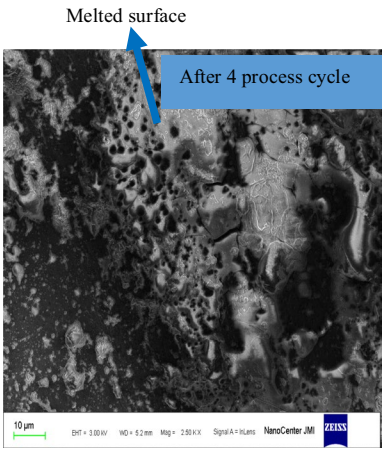


Fig. 16 Melting of workpiece surface in TACAFM process

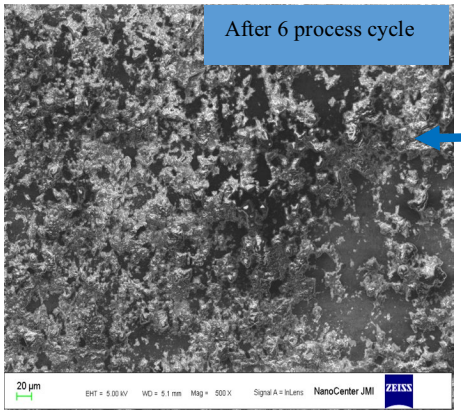


Fig. 17 SEM image showing sparks spot on the surface in TACAFM process

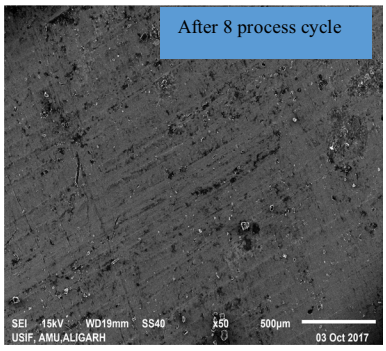


Fig. 18 SEM image of finished surface after TACAFM process

will significantly affect the surface integrity of the component [38, 39]. Therefore, effect of abrasive mesh size and current intensity was studied for $\%\Delta R_a$ in the present investigation. Figure 21 shows that TACAFM process provides better surface finish in comparison with conventional AFM process on increasing the abrasive mesh size. As the abrasive mesh size increases, the size of abrasive particle is decreased, and it will remove tiny amount of material from the surface due to lesser contact region. In TACAFM process, due to melting of the surface material abrasive particles have to impart lesser force for removing the similar amount of material from the surface. This increases the material removal and thus provides better surface finish. The experimental result showed average 39.74% improvement in surface finish after using the thermal spark mechanism. Figure 22 shows that as the value of the current increases, percentage improvement in surface finish initially is increased due to increase in the material removal but with further increase in current, surface finish decreased due to large crater formed. The experimental results show that up to 8 Amp, there is increasing trend in improvement in surface finish and after 8 Amp percentage improvement in surface finish is decreased. This may be further explained, due to current increase more material melted at work surface, which is further removed by abrasive particles. But once the amount of molten metal is reached at certain critical value, abrasive particles are not able to remove all the molten material that solidified on the work surface, results in formation of recast layer on work surface that reduces the percentage improvement of surface roughness. On increasing the material removal, surface finish goes better but after achieving more material removal will damage the surface finish and deteriorates the surface quality. Habib et al. [35] found that on increasing current, surface finish got better as more discharge energy available at the discharge point with increasing current. Sultan et al. [39] performed an experiment on EDM and found that on increasing the current, surface finish firstly increased and later decreased. The same trend was observed from the experimental results obtained in the developed thermal additive centrifugal AFM setup.

7.3 Temperature distribution on the workpiece surface

The effects of temperature distribution according to the gap between the electrode tip and workpiece surface was shown in Fig. 23.

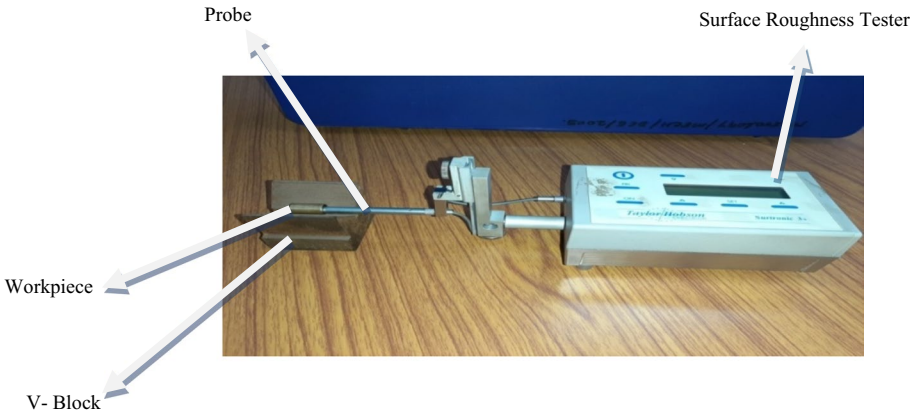
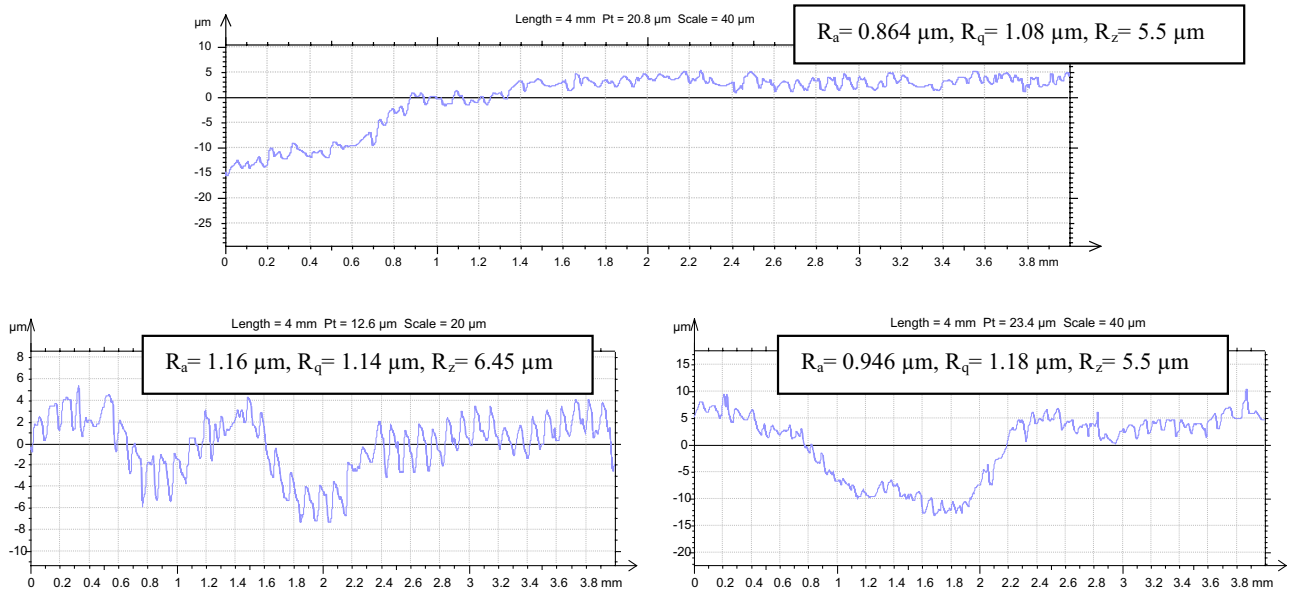
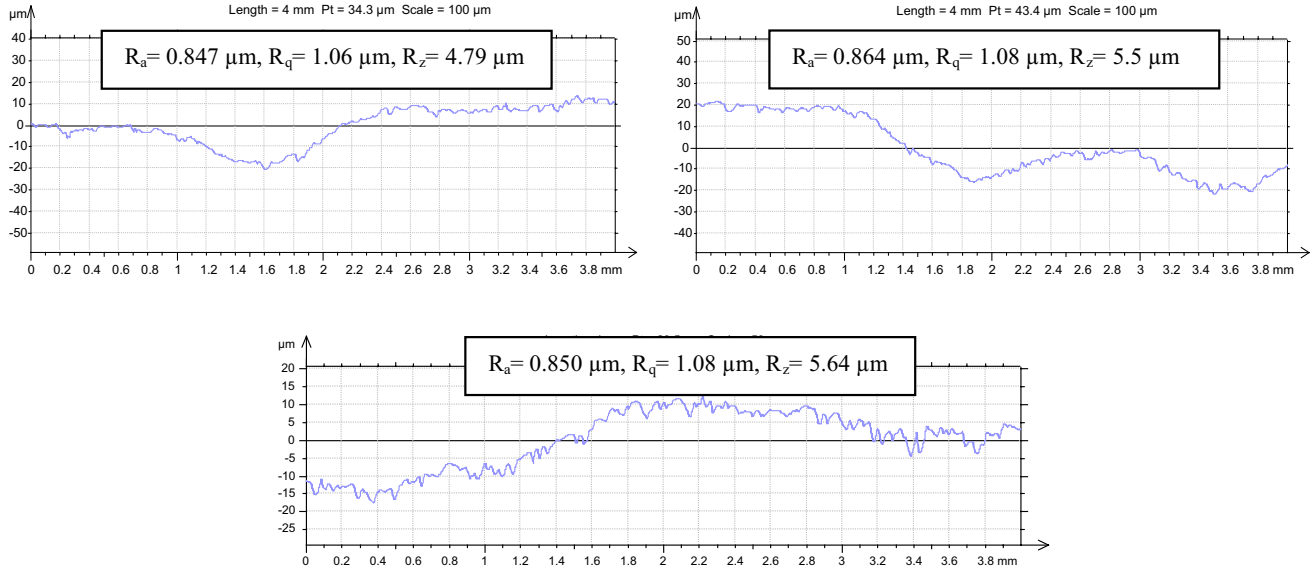


Fig. 19 Taylor Hobson precision machine for measurement of surface roughness



(a) Surface roughness of workpiece before machining



(b) Surface roughness of workpiece after machining

Fig. 20 Surface roughness of workpiece before and after finishing

Figure 23 shows that when the electrode was stationary, spark will only be generated at that place where electrode tip was closer to the workpiece surface. The results also show maximum temperature on the workpiece surface close to the electrode tip. Therefore, to maintain uniformity on the surface, rotation of either workpiece or electrode was necessary. In this analysis, electrode was taken as rotating one. Due to the rotation of the electrode, whole spark energy cannot be concentrated at a single spot which utilizes the energy to melt the workpiece surface in a uniform way. It can be clearly analyzed from the model results that with increase in rotational speed of electrode, temperature starts becoming uniform but after a certain level when rotational speed was further increased the temperature starts increasing on the workpiece surface. The reason for the sudden temperature rise may be due to spark generated at the tip will be at the same position in very less time and more energy

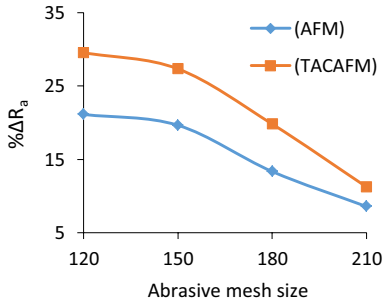


Fig. 21 Effect of abrasive particle mesh size on % ΔR_a for conventional and TACAFM process [at 5 A current]

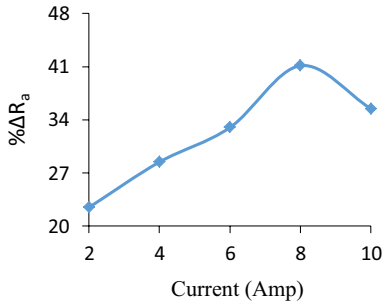


Fig. 22 Effect of current on % ΔR_a in the developed TACAFM process

interaction will occur between the tip and the workpiece surface. The other reason may be due to higher resistance in the flow of media with the increase in the rotational speed of the electrode. The rise in the temperature on the workpiece surface can be clearly observed at 1000 rpm of rotational speed.

The model result also shows that whenever the gap between the electrode tip and work surface increases, the range of temperature generated on the workpiece surface decreases. This decrease in temperature range was due to less amount of heat transfer because of more convection resistance with increase in the gap between the rotating electrode and surface. The modeling result also shows that the entering velocity of the media was continuously decreasing from inlet to the outlet, as shown in Fig. 24. Figure 24 shows that as the media starts flowing its energy was continuously decreasing due to friction between the media particles. When media interacts with the work surface its energy was increased due to conversion of pressure energy into the kinetic energy. After the impact again, the energy starts decreasing.

The measured values of response, i.e., material removal and % improvement in R_a corresponding to central composite design matrix are shown in Table 3. The effect of variable parameters on the response has been analyzed by response surface and contour plots.

7.4 Effects of process parameters on material removal

The 3D response surface and contour plots for the effect of variable parameters on material removal has been established according to the fitted model:

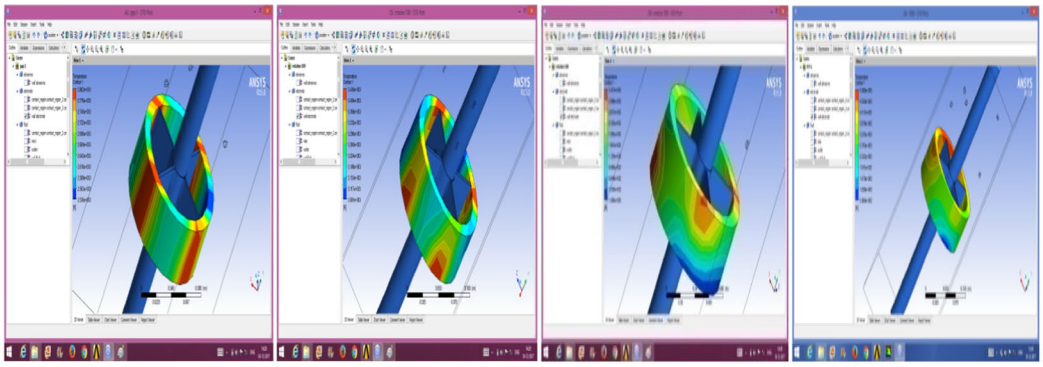
$$\begin{aligned}\text{Material Removal} = {} & -179.374 - 0.143374 \times \text{Current} \\ & + 512.404 \times \text{Duty Cycle} + 0.0246083 \\ & \times \text{Rotation} - 0.75525 \times \text{Pressure} \\ & + 0.56875 \times \text{Abrasive Conc} \\ & + 5.65313 \times \text{Current} \times \text{Duty Cycle} \\ & - 2.85781 \times \text{Current} \times \text{Abrasive Conc} \\ & + 2.20375 \times \text{Pressure} \times \text{Abrasive Conc} \\ & - 0.0676674 \times \text{Current}^2 - 362.571 \\ & \times \text{Duty Cycle}^2\end{aligned}$$

Table 4 shows the analysis of variance table for material removal having model F value as 49.79, which represents the model is significant. There was 0.01% chance for obtaining the larger F value which may be due to noise. If P was less than 0.05, it shows significant model terms. For material removal terms, A, B, C, E, AB, AE, DE, A^2, B^2 were the significant model terms with their percentage contribution as 83.05%, 3.35%, 2.54%, 1.95%, 1.43%, 1.46%, 1.36%, 2.45%, and 1.72%, respectively. The values greater than 0.1000 shows that the model terms were insignificant in nature. If insignificant model terms exist in large number, then by using model reduction method, model can be improved. The F value of 1.88 represents that the lack of fit was not significant in nature. This case has 25.07% chance that a larger value of Lack of Fit F value occurs which may be due to noise. The Predicted R^2 of 0.8890 was in sensible relation with the adjusted R^2 of 0.9403, which shows that the difference was less than 0.2. Adeq Precision estimates the S/N ratio. The ratio of more than 4 was advisable. The ratio of 29.113 represents an adequate signal for the analysis.

Figure 25 shows a magnificent acceptability of the regression model. Every observed value was comparable to the predicted value which can be taken from the model as shown in Fig. 25. Figure 25 shows a fair consistency of the predicted line with the actual values. The actual value points were nearer to the predicted line which means that the experimental values are satisfying the predicted results.

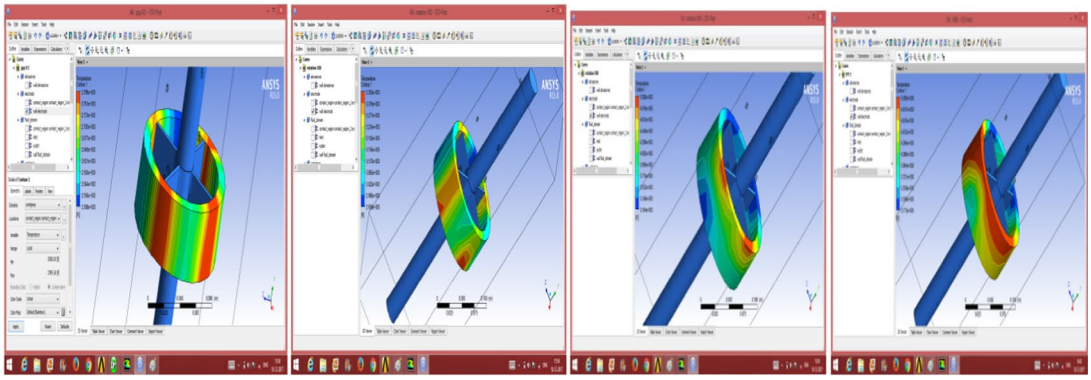
Figure 26a shows effect of duty cycle and discharge current on material removal. The surface plot reveals that material removal increases with discharge current. When the discharge current was increased, the discharge energy density increases, meaning that higher energy per pulse was available for melting the surface area. This causes rapid melting of the surface in spark region corresponding to formation of

(a) When the gap between the electrode and workpiece is = 0.1 mm



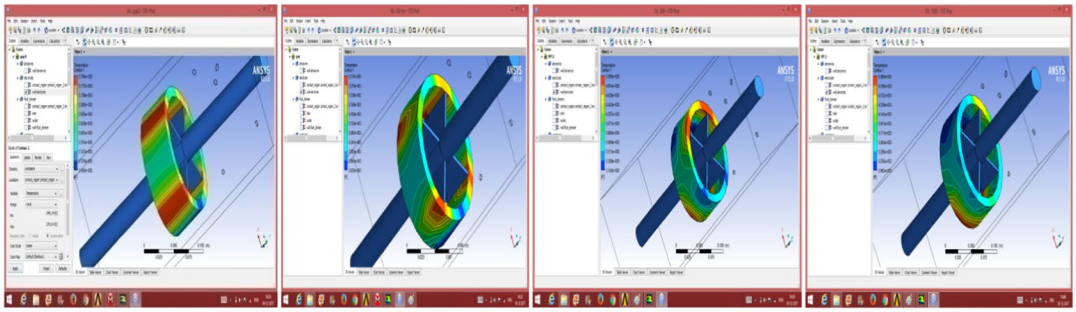
(i) Electrode stationary (ii) electrode rpm= 100 (iii) rpm= 500 (iv) rpm= 1000

(b) When the gap between the electrode tip and work surface is = 0.5 mm



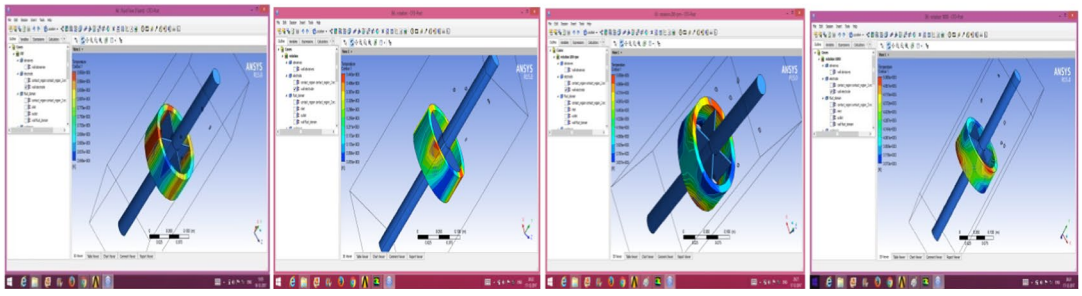
(i) Electrode stationary (ii) electrode rpm= 100 (iii) rpm= 500 (iv) rpm= 1000

(c) When the gap between the electrode tip and work surface is = 0.8 mm



(i) Electrode stationary (ii) electrode rpm= 100 (iii) rpm= 500 (iv) rpm= 1000

(d) When the gap between the electrode tip and work surface is = 1 mm



(i) Electrode stationary (ii) electrode rpm= 100 (iii) rpm= 500 (iv) rpm= 1000

Fig. 23 Temperature distribution around work surface at different rotational speed of electrode

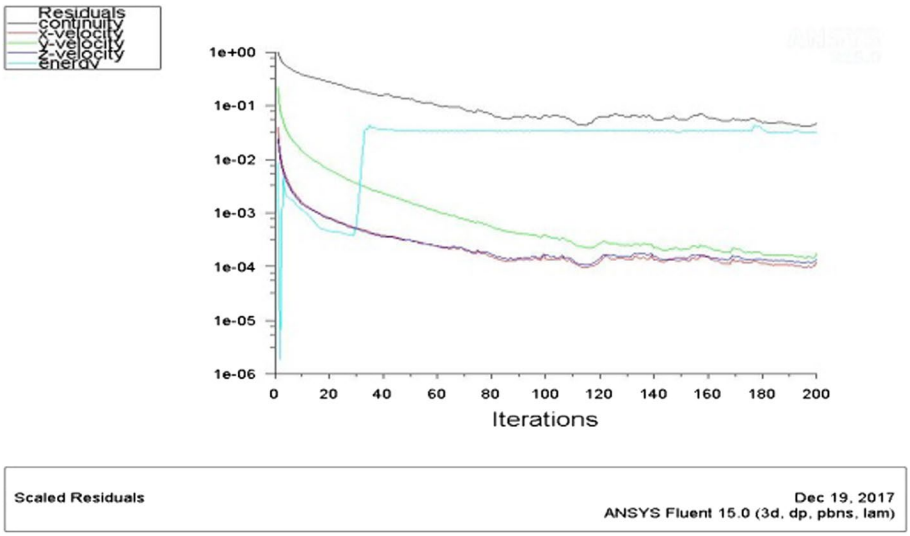


Fig. 24 Residuals of energy and velocity of media

Table 4 Pooled ANOVA table for material removal of TACAFM process

Source	C. V	R. V	S.S	D.F	M.S	*F* value	Prob. > *F*
Model	+26.90	−179.37439	1424.58	10	142.46	49.79	< 0.0001
A	+7.03	−0.143374	1186.24	1	1186.24	414.56	< 0.0001
B	+1.41	+512.40429	47.97	1	47.97	16.76	0.0005
C	+1.23	+0.024608	36.33	1	36.33	12.70	0.0018
D	+0.6312	−0.755250	9.56	1	9.56	3.34	0.0818
E	+1.08	+0.568750	27.80	1	27.80	9.72	0.0052
AB	+1.13	+5.65313	20.45	1	20.45	7.15	0.0142
AE	−1.14	−2.85781	20.91	1	20.91	7.31	0.0133
DE	+1.10	+2.20375	19.43	1	19.43	6.79	0.0165
A^2	−1.08	−0.067667	35.01	1	35.01	12.23	0.0021
B^2	−0.9064	−362.57143	24.54	1	24.54	8.58	0.0080
Residual			60.09	21	2.86		
Lack of fit			51.53	16	3.22	1.88	0.2507
Pure error			8.56	5	1.71		
Cor total			1484.67	31			
SD	1.69		R^2	0.9595			
Mean	25.41		Adjusted R^2	0.9403			
C.V. %	6.66		Predicted R^2	0.8890			
Press	164.81		Adeq precision	29.1133			

C.V Coded value, *R.V* real value, *S.S* sum of squares, *D.F* degree of freedom, *M.S* mean square

deeper discharge craters which can be easily taken away by the abrasive particles and increases the material removal. Yan et al. [40] also reported increase in material removal with the value of current in EDM process due to the formation of deeper craters on the surface.

Figure 26a also shows increase in material removal with duty cycle. The increase in duty cycle means discharge was applied for a longer duration which increases the energy concentration on the surface corresponding to larger material removal. Miller et al. [41] also stated increase in material removal with the duty cycle on machining the advanced materials through Wire EDM process.

Figure 26b shows effect of abrasive concentration and discharge current on material removal. As the abrasive concentration was increased number of dynamic abrasive particles, which were actually involved in the cutting

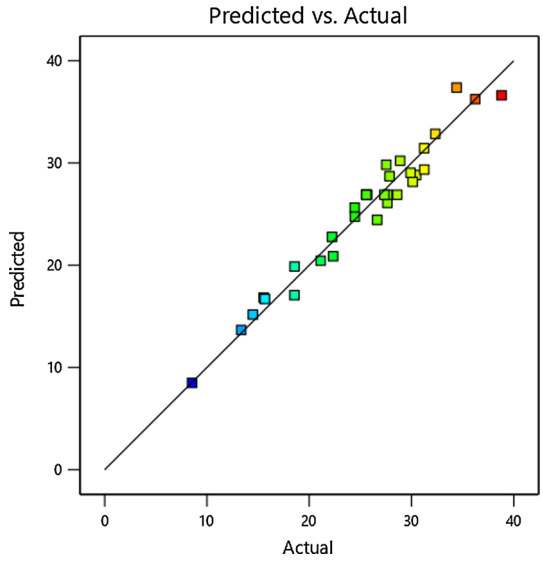


Fig. 25 Plot of actual versus predicted response of material removal

process were increased. This improves the contact region between the abrasives and the finishing surface. The increase in the contact region corresponded to higher material removal. Gorana et al. [42] found the similar results and stated increase in material removal with the abrasive concentration.

Figure 26c shows the effect of extrusion pressure and abrasive concentration on the material removal. Figure shows as the extrusion pressure increases material removal was increased. The reason for more material removal was that when the extrusion pressure was increased the force exerted by the abrasive particles was increased. This increase in the force applied by the abrasive particle causes deeper penetration, allowing increase in material removal. Gorana et al. [42] found the similar results towards material removal on increasing the extrusion pressure.

The optimum values for the input parameters and their corresponded responses were analyzed by the software and the details were presented in Table 5. For single-factor optimization, other response variables were neglected. For the validation of optimized responses confirmation experiments were performed as shown in Table 5. It was seen from the experimental validation, predicted values were very close to the experimental values.

7.5 Effect of process parameters on percentage improvement in R_a (% increased R_a)

The effect of variable parameters on percentage improvement in R_a using contour plots and 3-D response surface have been established according to the fitted model:

$$\begin{aligned}\%\ \text{increased } R_a = {} & +756.49 - 0.603542 \times \text{Current} \\ & - 1882.27 \times \text{Duty Cycle} - 0.599127 \\ & \times \text{Rotation} - 0.427792 \times \text{Pressure} \\ & + 5.62917 \times \text{Abrasive Conc.} \\ & + 3.57188 \times \text{Current} \times \text{Abrasive Conc.} \\ & + 0.9595 \times \text{Duty Cycle} \times \text{Rotation} \\ & - 0.011195 \times \text{Rotation} \times \text{Pressure} \\ & - 0.29075 \times \text{Rotation} \times \text{Abrasive Conc.} \\ & - 2.6375 \times \text{Pressure} \times \text{Abrasive Conc.} \\ & + 1187.21 \times \text{Duty Cycle}^2 + 0.000485708 \\ & \times \text{Rotation}^2 + 0.116571 \times \text{Pressure}^2.\end{aligned}$$

Table 6 shows the ANOVA for percentage improvement in R_a having the model F value as 16.49. This represented the model as significant. There was a 0.01% chance for having a large F value due to noise. If value of P was less than the 0.0500, model terms became significant. In this model A, B, D, AE, BC, CD, CE, DE, B^2, C^2, D^2 were significant model terms. The values more than 0.1000, showed that model terms were insignificant. If more number of insignificant model terms exists, then the model can be improved by the reduction process. The lack of fit, F, value of 0.45 showed this as insignificant relative to the pure error. This case had 87.79% chance of having large Lack of Fit F value due to noise. The Predicted R^2 of 0.8073 was in a close relation with the Adjusted R^2 of 0.8750; i.e. the difference between them is less than 0.2. Adeq Precision shows the value of S/N ratio. The ratio more than 4 was advisable. The ratio of 15.891 represented an adequate signal.

Figure 27 shows a magnificent acceptability of the regression model. Each observed value was comparable to the predicted value obtained from the model represented as in Fig. 27. Figure 27 indicated a close relationship between predicted line and actual values. The actual values obtained were nearer to the predicted line which clearly indicates that experimental values were satisfying the predicted values.

In conventional AFM process, material removal occurs due to the abrasion mechanism by the sharp cutting edges particles. This process is subjected to the axial force in the direction of the media flow due to pressure difference created between the two opposite hydraulic cylinders. Centrifugal-assisted AFM process uses CFG rod to develop centrifugal force in the media flow path. Centrifugal-assisted AFM process is associated with two types of forces (i.e., axial force and radial force). Axial force is exerted in the direction of flow of media while radial force acts in perpendicular direction to the workpiece surface. CFAAFM process requires lesser number of cycles for

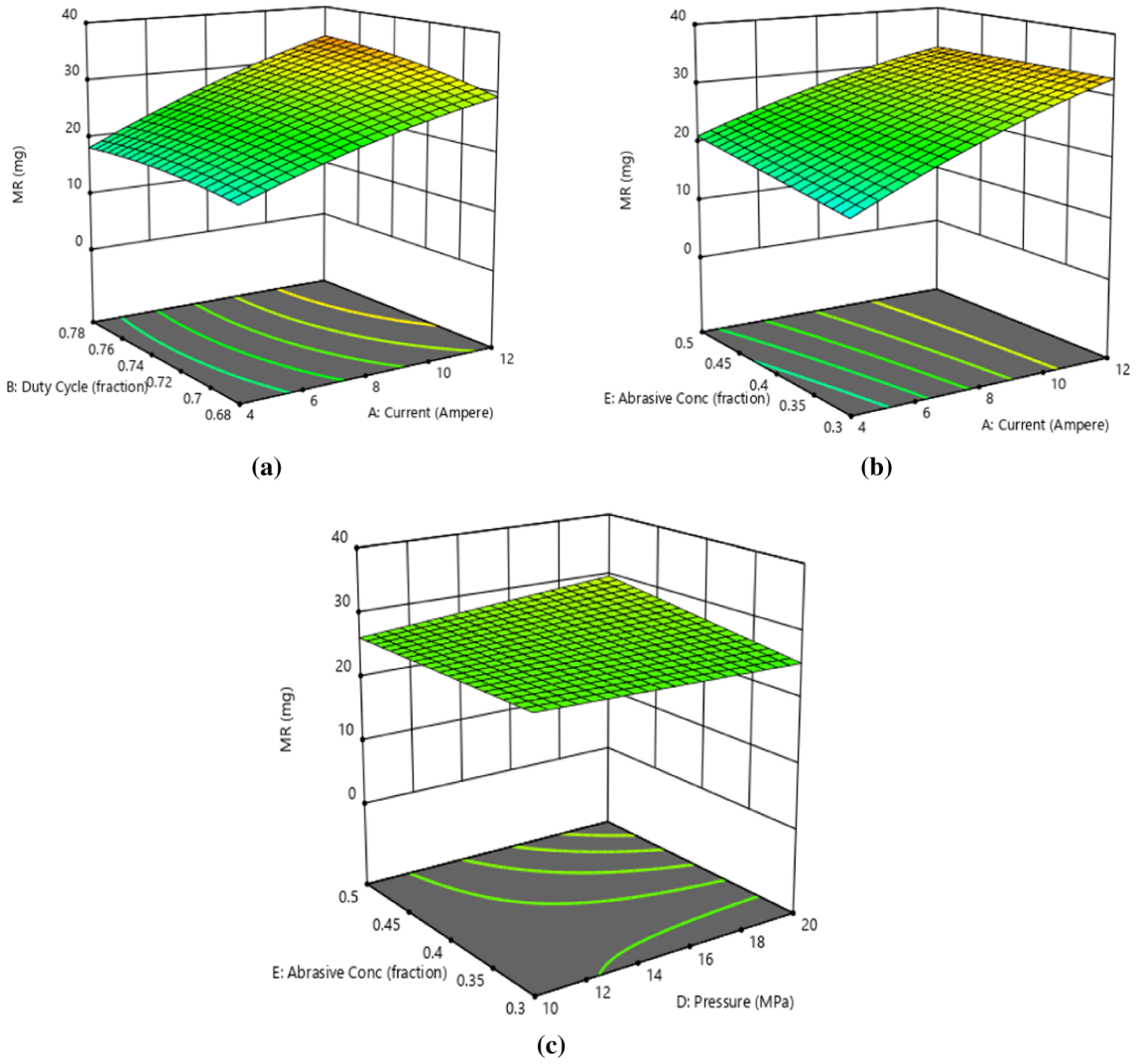


Fig. 26 Effect of significant parameters on the response (material removal)

Table 5 Single-factor optimization and comparative study of optimized outcomes and experimental facts of process variables (material removal)

Optimization type	Objective	Optimized process variables					Response (predicted)	Response (experimental)	Desirability
		Current (Amp)	Duty cycle (fraction)	Rotational speed (rpm)	Pressure (MPa)	Abrasive concentration (fraction)			
Single response	To maximize MR	12	0.78	250	10	0.3	36.5705 mg	36.571 mg	0.926

removing the same material from the surface in comparison with the conventional AFM process.

TACAFM process is the combination of two processes that is centrifugal-assisted AFM and EDM process. The process uses the thermal spark energy for melting the surface. The generated spark between both poles developed a high temperature on the workpiece surface. This caused formation of the oxide layers on the surface. Figure 28a shows surface produced by centrifugal-assisted AFM process. As the spark was generated, the oxide layers on the surface started forming and further melted the surface material which can be easily understood through Fig. 28b, c. The

Table 6 Pooled ANOVA table for percentage increase in surface roughness of TACAFM process

Source	C.V	R.V	S.S	D.F	M.S	*F* value	Probability > *F*
Model	+24.81	+756.48984	1230.77	14	87.91	16.49	< 0.0001
A	+3.30	−0.603542	261.49	1	261.49	49.06	< 0.0001
B	+2.15	−1882.27417	110.68	1	110.68	20.76	0.0003
C	+0.5683	−0.599127	7.75	1	7.75	1.45	0.2443
D	−1.12	−0.427792	30.29	1	30.29	5.68	0.0291
E	+0.6633	+5.62917	10.56	1	10.56	1.98	0.1773
AE	+1.43	+3.57188	32.66	1	32.66	6.13	0.0241
BC	+2.40	+0.959500	92.06	1	92.06	17.27	0.0007
CD	−2.80	−0.011195	125.33	1	125.33	23.51	0.0002
CE	−1.45	−0.290750	33.81	1	33.81	6.34	0.0221
DE	−1.32	−2.63750	27.83	1	27.83	5.22	0.0354
B^2	+2.97	+1187.20833	260.21	1	260.21	48.82	< 0.0001
C^2	+1.21	+0.000486	43.55	1	43.55	8.17	0.0109
D^2	+2.91	+0.116571	250.87	1	250.87	47.07	< 0.0001
Residual			90.61	17	5.33		
Lack of fit			47.27	12	3.94	0.4544	0.8779
Pure error			43.34	5	8.67		
Cor total			1321.38	31			
SD	2.31		R^2	0.9314			
Mean	30.79		Adjusted R^2	0.8750			
C.V. %	7.50		Predicted R^2	0.8073			
Press	254.59		Adeq Precision	15.8913			

C.V Coded value, *R.V* real value, *S.S* sum of squares, *D.F* degree of freedom, *M.S* mean square

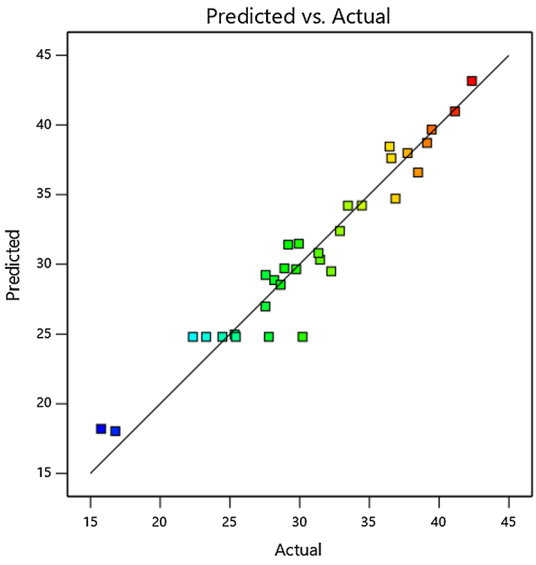
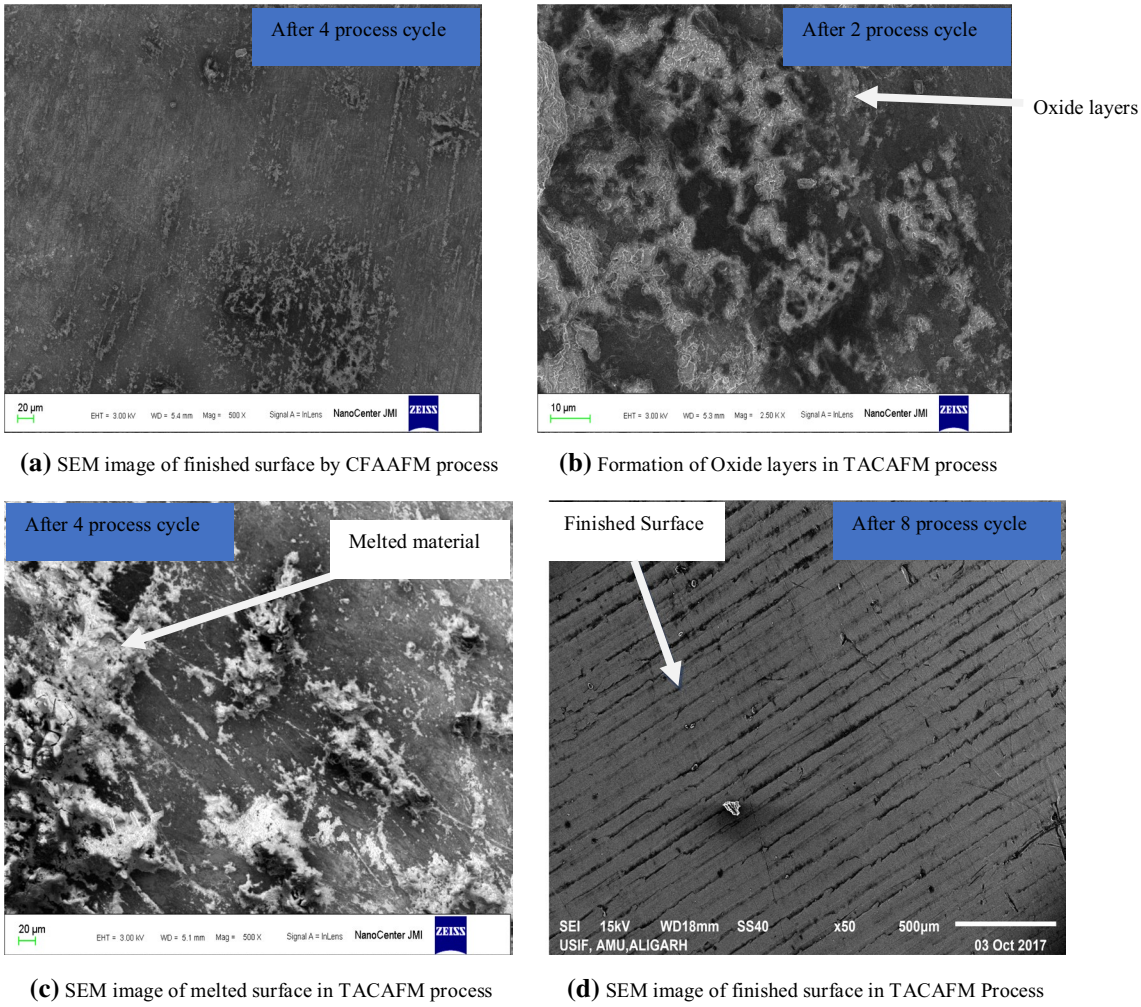


Fig. 27 Plot of actual versus predicted response of % improvement in R_a

finished surface produced after TACAFM process can be understood through Fig. 28d.

Figure 29a shows that when the current and abrasive concentration value increases, % ΔR_a also increased. When the discharge current value was increased, the discharge energy density increased, meant that higher energy per pulse would be available for melting the surface material. This caused rapid melting of the surface in spark region corresponding to the formation of deeper discharge craters [40], which can be easily taken away by the abrasive particles and improved the surface finish. On increasing the abrasive concentration, active number of abrasive particles increased that improved the contact area between the sharp abrasive particles and the workpiece surface. More number of sharp cutting edges involved in the abrasion process easily removed the peaks of surface which resulted in good level of surface finish. The results obtained were comparable to the outcome of experiments available in literature review [42].

Figure 29b shows increase in percentage improvement in R_a with rotational speed and duty cycle while extrusion pressure, current and abrasive concentration were kept constant. As the duty cycle was increased, it meant that the discharge energy was available for longer duration. This increased the temperature developed on the surface and softens more material. The abrasive particles easily dig on the surface and provide good surface integrity. When the electrode was stationary, abrasive particles were subjected to the axial force provided by the media pressure. This axial force caused the abrasive particle to roll along the workpiece surface [43].



(a) SEM image of finished surface by CFAAFM process

(b) Formation of Oxide layers in TACAFM process

(c) SEM image of melted surface in TACAFM process

(d) SEM image of finished surface in TACAFM Process

Fig. 28 SEM images of surfaces developed during CFAAFM and TACAFM process

But on increasing the rotational speed of the electrode a centrifugal force was developed in the media flow path. This centrifugal force increased the resultant force at which the abrasive particles were impacting the workpiece surface. This resulted in increase of active abrasive particles which were involved in abrasion process. More number of active abrasive particles increased the energy available to break the atomic bonds and developing new finished surface by the movement of atoms from the fixed position. This corresponded more material removal and provided better surface finish [44]. Walia et al. [31] obtained the similar trend from the experimental observations in centrifugal-assisted AFM process.

Figure 29c shows increase in percentage improvement in R_a with the extrusion pressure. In TACAFM process two forces were involved in material removal (i.e. axial force and radial force). Axial force acted in direction of media flow while radial force in perpendicular direction to the workpiece surface. The resultant of these two forces was the total force acting on the abrasive particles which impact on the workpiece surface. At lower value of extrusion pressure both axial and radial force values were very less and abrasive particles were not able to shear the peaks of the surface. But as the extrusion pressure is increased, resultant force acting on the abrasive particles was also increased, which corresponded to increase in percentage improvement in R_a value. Jain and Jain [43] also found similar pattern, i.e. increase in percentage improvement in R_a value with the extrusion pressure.

Figure 29d shows increase in percentage improvement in surface roughness with increase in rotational speed and abrasive concentration while current, duty cycle and extrusion pressure are kept constant. This is due to increase in active abrasive particles and contact region between the sharp cutting edges and workpiece surface. Figure 29e shows increase in surface finish with abrasive concentration while surface

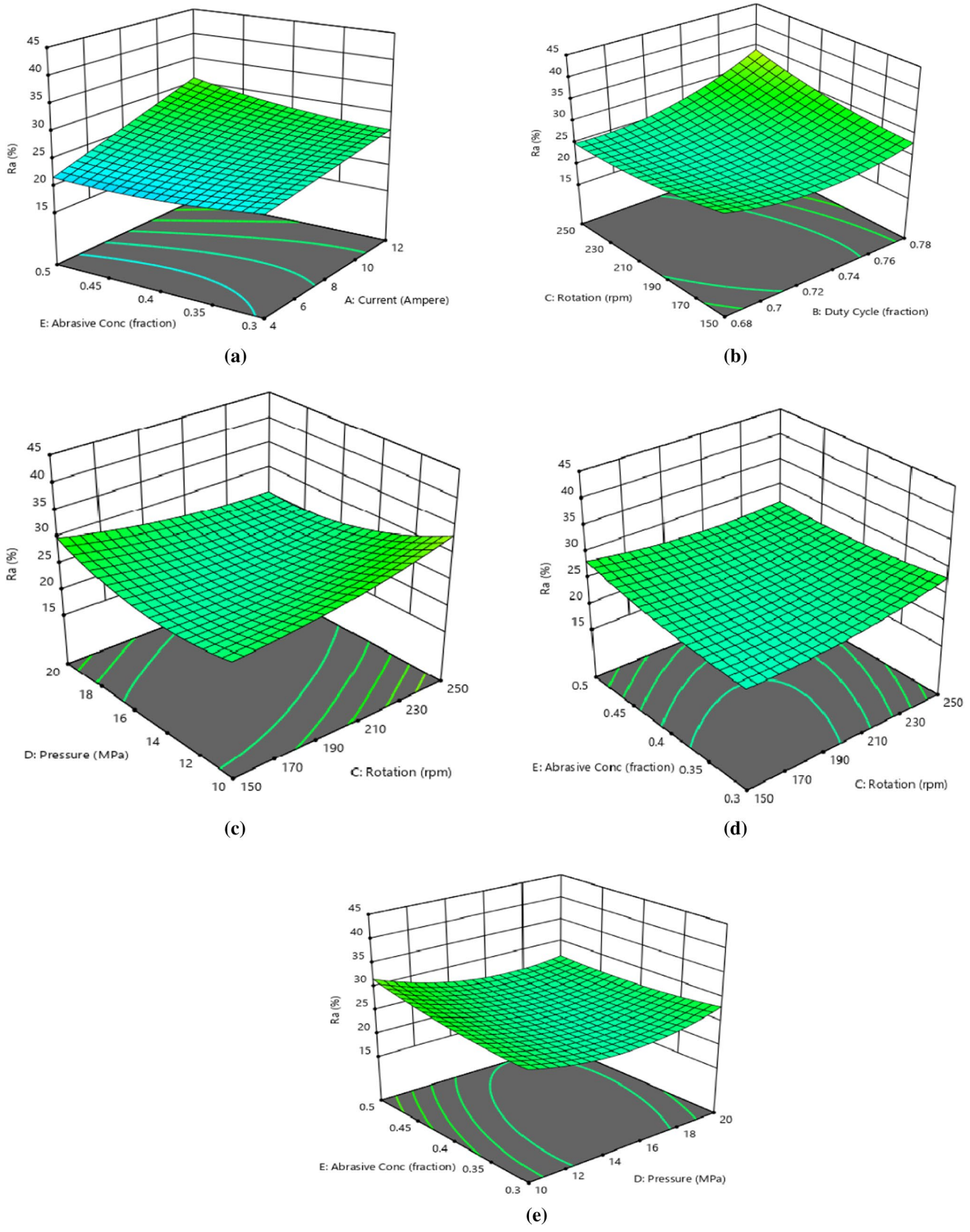


Fig. 29 Effect of process parameters on percentage improvement in R_a

Table 7 Single factor optimization and comparative study of optimized outcomes and experimental facts of process variables (% improved R_a)

Optimization type	Objective	Optimized process variables					Response (predicted)	Response experimental	Desirability
		A	*B*	*C*	*D*	*E*			
Single response	To maximize % improvement in R_a	12	0.78	250	10	0.3	42.3887%	42.389%	1

finish initially decreases with pressure but later it increases. The reason may be that initially many abrasive particles could not come in contact with the surface but as the pressure increases; a greater number of abrasive particles came in contact with the surface and increases both the material removal and surfaces finish.

The optimum values for the input parameters and their corresponded responses were analyzed by the software and the details are presented in Table 7. For single factor optimization, other response variables were neglected. For the validation of optimized responses, confirmation experiments were performed as shown in Table 7. It was seen from the experimental validation, predicted values were very close to the experimental values.

8 Conclusions

The following conclusions can be made from the experimental and modeling study:

- Mathematical model has been developed using Coriolis force effect to predict material removal of developed TACAFM process. Developed mathematical model results show, average 18.78% variation in material removal that were calculated when compared to the experimental results of TACAFM process.
- From 3D modeling using ANSYS®, it can be concluded that as the gap between the electrode and workpiece surface is increased, the temperature on the workpiece surface decreases. This may be due to less spark frequency between them.
- The computational results show that at higher rotational speed, temperature of work surface was also increased due to increase in the spark intensity and increased obstruction in the flow of media.
- The experimental results of TACAFM process show average 44.34% improvement in material removal and 39.74% improvement in surface finish, respectively when compared to conventional AFM process.
- The optimum values of material removal and percentage improvement in surface finish are found to be 36.57 mg and 42.39% respectively.
- Current intensity and duty cycle have the major effect on finishing performance (83.05% and 3.35% respectively) in comparison with the other input parameters.
- The current, duty cycle, abrasive concentration, and pressure have the significant effect on material removal.

References

1. Jain RK, Jain VK, Dixit PM (1999) Modeling of material removal and surface roughness in abrasive flow machining process. Int J Mach Tool Manuf 39(12):1903–1923
2. Ravi Sankar M, Jain VK, Ramkumar J (2008) Abrasive flow machining (AFM): an overview. In: INDO-US workshop on smart machine tools, intelligent machining systems and multi-scale manufacturing
3. Walia RS, Shan HS, Kumar P (2008) Morphology and integrity of surfaces finished by centrifugal force assisted abrasive flow machining. Int J Adv Manuf Technol 39:1171–1179
4. Kim J, Kim K (2004) Deburring of burrs in spring collets by abrasive flow machining. Int J Adv Manuf Technol 24:469–473
5. Singh S, Shan HS (2002) Development of magneto abrasive flow machining process. Int J Mach Tool Manuf 42:953–959
6. Singh S (2002) Studies in metal finishing with magnetically assisted abrasive flow machining. PhD thesis, IIT Roorkee
7. Ravi Sankar M, Mondal S, Ramkumar J, Jain VK (2009) Experimental investigations and modeling of drill bit guided abrasive flow finishing (DBG-AFF) process. Int J Adv Manuf Technol 42(7–8):678–688
8. Dabrowski L, Marciniak M, Szewczyk T (2006) Analysis of abrasive flow machining with an electrochemical process aid. Proc Inst Mech Eng Part B J Eng Manuf 220:397–403
9. Brar BS, Walia RS, Singh VP (2015) Electrochemical-aided abrasive flow machining (ECA2FM) process: a hybrid machining process. Int J Adv Manuf Technol 79(1–4):329–342
10. Singh R, Walia RS, Suri NM (2015) Parametric optimization of centrifugal-magnetic force assisted abrasive flow machining process using utility concept. Int J Res Eng Technol 4(8):382–388
11. Vaishya R, Walia RS, Kalra P (2015) Design and development of hybrid electrochemical and centrifugal force assisted abrasive flow machining. Mater Today Proc 2(4–5):3327–3341
12. Gorana VK, Jain VK, Lal GK (2006) Forces prediction during material deformation in abrasive flow machining. Wear 260(1):128–139
13. Mittal S, Kumar V, Kumar H (2015) Experimental investigation and optimization of process parameters of Al/SiC MMCs finished by abrasive flow machining. Mater Manuf Process 30:902–911
14. Wang AC, Cheng KC, Chen KY, Lin YC (2014) Enhancing the surface precision for the helical passageways in abrasive flow machining. Mater Manuf Process 29:153–159

15. Jai Kishan S, Mali HS, Garg V (2017) Experimental investigation on unidirectional abrasive flow machining of trim die work piece. Mater Manuf Process. https://doi.org/10.1080/10426914.2017.1364847
16. Venkatesh G, Sharma AK, Kumar P (2015) On ultrasonic assisted abrasive flow finishing of bevel gears. Int J Mach Tools Manuf 89:29–38
17. Marzban MA, Hemmati SJ (2017) Modeling of abrasive flow rotary machining process by artificial neural network. Int J Adv Manuf Technol 89:125–132
18. Fu Y, Wang X, Gao H, Wei H, Li S (2016) Blade surface uniformity of blisk finished by abrasive flow machining. Int J Adv Manuf Technol 84:1725–1735
19. Jain RK, Jain VK (2000) Optimum selection of machining conditions in abrasive flow machining using neural network. J Mater Process Technol 108(1):62–67
20. Steif PS, Haan JJ (1998) Abrasive wear due to slow flow of a concentrated suspension. Wear 219(2):177–183
21. Rajeshwar G, Kozak J, Rajurkar KP (1994) Modeling and computer simulation of media flow in abrasive flow machining process. Manuf Sci Eng 68(2):965–971
22. Walia RS, Shan HS, Kumar P (2006) Finite element analysis of media used in the centrifugal force assisted abrasive flow machining process. Proc Inst Mech Eng Part B J Eng Manuf 220:1775–1785
23. Wani AM, Yadava Vinod, Khatri Atul (2007) Simulation for the prediction of surface roughness in magnetic abrasive flow finishing (MAFF). J Mater Process Technol 190:282–290
24. Das M, Jain VK, Ghoshdastidar PS (2008) Fluid flow analysis of magnetorheological abrasive flow finishing (MRAFF) process. Int J Mach Tools Manuf 48(3):415–426
25. Ali P, Walia RS, Murtaza Q, Ranganath MS (2019) Experimentation and modelling of CNT additive abrasive media for micro finishing. Measurement. https://doi.org/10.1016/j.measurement.2019.107133
26. Singh J, Walia RS, Satsangi PS, Singh VP (2012) Hybrid electric discharge machining process with continuous and discontinuous ultrasonic vibrations on workpiece. Int J Mech Syst Eng 2(1):22–33
27. Walia RS, Shan HS, Kumar P (2004) Effect of providing a rotating rod inside the hollow cylindrical workpiece on the material removal in AFM. In: Proceedings of 21st AIMTDR conference, VIT, Vellore, India, pp 143–148
28. Marochkin VN (1959) The limiting plastic state in indenting and compressing a truncated cone. Frict Wear Mach 13:79–131
29. Arpaci VS, Selamet A, Kao SH (1999) Introduction to heat transfer. Prentice Hall Publication, Upper Saddle River. ISBN 0-13-391061-X
30. Walia RS, Shan HS, Kumar P (2008) Determining dynamically active abrasive particles in the media used in centrifugal force assisted abrasive flow machining process. Int J Adv Manuf Technol 38(11):1157–1164
31. Walia RS, Shan HS, Kumar P (2009) Enhancing AFM process productivity through improved fixturing. Int J Adv Manuf Technol 44:700–709
32. https://nptel.ac.in/content/storage2/courses/112105127/pdf/LM-39.pdf. Accessed 05 Dec 2019
33. Kazimierz A, Karol M, LabuzStefan MA, Marek M (1996) The temperature of a plasma used in electrical discharge machining. Plasma Sources Sci Technol 5:736–742
34. Kansal HK, Singh Sehijpal, Kumar Pradeep (2008) Numerical simulation of powder mixed electric discharge machining (PMEDM) using finite element method. Math Comput Model 47:1217–1237
35. Draper NR (1982) Center points in second order response surface designs. Technometrics 24(2):127–133
36. Walia RS, Shan HS, Kumar P (2006) Parametric optimization of centrifugal force-assisted abrasive flow machining (CFAAFM) by the Taguchi method. Mater Manuf Process 21(4):375–382
37. Gorana VK, Jain VK, Lal GK (2006) Prediction of surface roughness during abrasive flow machining. Int J Adv Manuf Technol 31(3–4):258–267
38. Habib Sameh S (2009) Study of parameters in electrical discharge machining through response surface methodology approach. Appl Math Model 33:4397–4407
39. Sultan T, Kumar A, Gupta RD (2014) Material removal rate, electrode wear rate and surface roughness evaluation in die sinking EDM with hollow tool through response surface methodology. Int J Manuf Eng. https://doi.org/10.1155/2014/259129
40. Yan BH, Tsai HC, Huang FY (2005) The effect in EDM of a dielectric of a urea solution in water on modifying the surface of titanium. Int J Mach Tools Manuf 45:194–200
41. Miller SF, Shih AJ, Qu J (2004) Investigation of the spark cycle on material removal rate in wire electrical discharge machining of advanced materials. Int J Mach Tools Manuf 44:391–400
42. Gorana VK, Jain VK, Lal GK (2004) Experimental investigation into cutting forces and active grain density during abrasive flow machining. Int J Mach Tool Manuf 44:201–211
43. Jain RK, Jain VK (1999) Simulation of surface generated in abrasive flow machining process. Robot Comput Integr Manuf 15:403–412
44. Walia RS, Shan HS, Kumar P (2006) Abrasive flow machining with additional centrifugal force applied to the media. Mach Sci Technol 10:1–14

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

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

Modeling and Predicting Piped Water Theft using Machine Learning Approach

Simon Peter Khabusi
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India

Rajni Jindal
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India

***Abstract*— Water theft is a major challenge to water distribution companies in most communities around the world. This results into unbalanced water flow and damage to the water pipes. The resulting high level of nonrevenue water does not only lower income generation but also the quality of service. The recent advances in the field of Machine Learning have seen its latest application to a wide range of fields. This work proposes a Random Forest (RF) prediction Model to accurately detect water theft along the distribution network. Data for training and testing purposes was collected through experimentation over a 3 hour period in 10 seconds intervals using a system prototype of two Arduino microcontrollers programmed and interfaced with water flow rate sensors with the ability to adopt a normal flow rate value after a small time delay. Therefore flow rate fluctuations outside a range of ±2 litres/minute were taken as abnormal flow. The proposed RF model was evaluated on four statistical measures namely; accuracy, precision, recall, and F-measure, and compared with three competitive approaches that is, logistic regression (LR), Support vector machine (SVM) and K-nearest Neighbour (KNN). Experimental results show no significant differences observed in accuracy and F-measure among the four models, while the proposed RF gives a higher value of recall. Conclusively, the proposed classifier has advantages compared with the other approaches in terms of reliable feature importance estimate and efficiency in test error estimation without incurring the cost of repeated model training associated with cross-validation.**

Keywords—Water theft, flowrate, Arduino, machine learning, supervised learning

I. INTRODUCTION

Water scarcity is a major challenge around the globe [1]. The United Nations 2015 report on world water development highlighted limitation of clean and safe drinking water as a major issue across Europe and North America, Asia and the Pacific, the Arab region, Latin America and the Caribbean, and Africa [1]. The United Nations Department of Economic and Social Affairs [UNDESA] report published in 2013 states that water crisis is not only a natural but also a human-made phenomenon [6]. Physical scarcity has resulted from natural phenomenon, that is; climate change and global warming whereas people's inability to utilize adequate water sources has resulted from economic scarcity which is common among low income earners in developed countries and more prevalent in under developed communities, especially in Africa [2].

World Health Organization (WHO) report of 2015 [3] shows that only 50% or less of people in rural Africa have access to clean drinking water and improved sanitation in spite of the fact that this is an issue that concerns all aspects of the wellbeing of persons, including their health, agricultural activities, educational development, economic productivity, even peace and stability [2]. In most communities, water is not only scarce but also of low quality [4] due to rapid urbanization and inefficiency in Government systems to supply safe and clean piped water to all citizens [5].

Despite these challenges, African states are striving hard to extend piped water to all communities [5]. The sector reforms and Investments report published by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) in 2015 showed a significant increase in piped water connections over the recent years in Africa. Among the countries surveyed, Burkina Faso had the highest piped water coverage and this was at 76%, followed by Kenya at 70%, Zambia at 66%, Tanzania at 59% and Uganda at 56% [5].

Therefore, due to financial inability to afford and limited piped water supply, water theft is prevalent in most communities in Uganda especially the urban centers [6]. Its usage in almost each and every aspect of life has made its distribution and supply of utmost importance to researchers. Homes, Hospitals, public offices, construction companies and Industrial centers need a steady constant supply of water to ensure sustainability of their activities. Hence any scarcity could lead to great losses of money and even lives. Large water consuming entities allocate a great percentage of their budget on water [7] and even still, the normal daily water distribution per service pipe could most likely not meet their water demands. These result in a number of illegal activities such as illegal connections, meter bypass, meter tampering, meter reversals, and vandalism to ensure that more water can be obtained at lower or no costs [8].

Some customers use suction pumps to increase water supply to their service lines there by rendering low/no supply to other clients and eventually destroying pipes due to extreme uncontrolled pressures. Such water volumes could under normal circumstances not be availed [9] considering the normal water supply procedure in a given pressure zone. Such attempts interfere with the distribution water pressure which results into poor water supply and leaks resulting from pipe bursts due to extreme pressures

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

exerted by the suction pipes [9][10][11]. Piped water theft is widely spread and it is perpetrated by all classes of people [12].

In Uganda for example, National Water and Sewerage Corporation (NWSC), a government parastatal body mandated with supplying clean and safe piped water to citizens incurs about 30% to 35% of its total revenue in water theft [13]. The initial attempts involving use of police to enforce the law against piped water theft, community policing and regular staff supervisions to reduce the Non-Revenue water (NRW) have not yielded much success. In response to this dire need, a lot of initiatives have been undertaken by different researchers to provide a mechanism for automated water supply, monitoring and timely reporting of water theft activities along the distribution network. The current state of research involves use of remote flow rate sensors laid along water distribution pipelines to capture the water flow rate values. Figure 1 shows the conceptual framework of the process of water theft detection in existing applications.

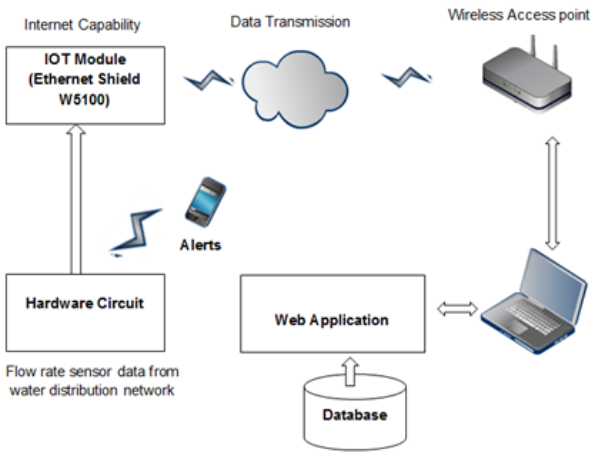


Fig. 1. Conceptual framework of existing piped water theft detection systems

A network of flowrate sensors interfaced with PLC is used to capture the flow patterns along the water distribution system. A range of normal water distribution flowrate is determined and coded on the PLC. Any variation outside the set range is treated as theft and reported to authorities via a sms using a GSM modem interfaced with PLC. Different technologies such as IOT, SCADA are used to relay the sensor data to a remote database serving a web application which takes a record of the daily patterns in water distribution [14]. The remotely controlled server running web application enables the authorities to look into the state of the network whenever it is required. The programs that run on these systems are fixed and rule-based yet the water theft problem is very dynamic and sophisticated in nature. This calls for more intelligent approaches to the problem hence the proposal of this Machine Learning framework for water theft prediction using Random Forest classifier.

Prediction models have been used greatly in different disciplines to address different issues. However, their performance are focused on similar aspects, that is, (i) use of data sets that are appropriate and with enough data (ii) prediction model performance assessed using appropriate performance metrics (iii) using statistical tests to evaluate the reliability of the results (iv) using data different from the training data to validate the predicted models[15].

In this work, we model and predict piped water theft using data collected over a period of 3 hours during an experiment. Water from a reservoir was let in the interconnected distribution network and its state was varied using suction pump and diversionary flow. The proposed model can be incorporated in existing systems at the server end to address the shortcomings of drawing conclusions based on fixed hardcoded rules. A discussion of the related work, materials and methods, project description, modeling and prediction, results and discussion, conclusion and future work have been presented in the proceeding sections.

II. RELATED WORK

Significant number of research initiatives has been undertaken in response to water theft challenges that face water distribution companies around the globe. To the best of our knowledge, till to-date, no work has been done on application of Machine Learning Predictive Models in piped water theft detection. The authors in [14] use Arduino for central processing interfaced with water flow rate sensors, solenoid valve and flow meter. The microcontroller captures water flow rate readings which are compared with the fixed set value of water flow. Any reading that deviates from the set norm value is regarded as theft. A mobile application called Cayenne is also proposed for prepaid water purchase. With the use of Ethernet Shield W5100, data from the system can be uploaded to the application. This IOT module is dependent on the WizNet5100 Ethernet chip datasheet with a system IP stack for handling TCP and UDP packets [16].

The proposed system tries to address over flow and over utilization issues. Incorporating SCADA and PLC into the system has been recommended for future work. Remote access to the monitoring system by use of the IOT module and secure data transmission through use of the different Ethernet protocols is a major contribution of this study though some forms of water theft including meter reversals, meter tampering and meter bypass have not been fully addressed.

A closer approach to that of [14] is the use of PLC and SCADA for monitoring and the integrated system based on flow rate readings captured using the PLC interfaced flow rate sensors, water level readings in the reservoir taken by the level sensor, proximity sensors and pressure transmitters [17]. Water distribution pressure is maintained using the level reading of the reservoir as the reference value and maintaining a water level above the set value. However, the conclusive inference based on pressure differences remains unanswered because in the conventional distribution

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

networks, pressure rise and fall is based on water usage and time of day or rate of water consumption [17]. Therefore, designating a nominal value or range could lead to misleading theft alerts.

Similar approaches have been applied by A. Mancharkar, Rakesh and IShan [18] and A. Gantala and P. Nalajala [19] who interfaced flow rate sensors with microcontroller and using a fixed set value of flow rate, water theft is deducted. Similar work by A. Rahat saw the use of PLC, level sensor, Solenoid valve and time base feature to detect water theft perpetrated by way of pressure manipulation [20]. These works also greatly make use of Solenoid valves for automatic shutdown of water as soon as theft is identified.

H.A Gaikwad and P.V.G Puranik [8], J. Tharanyaa, A. Jagadeesan and A. Lavanya [9] and P.D.B Madihali and P.SS Ittannavar have proposed different applications that make use of an array of flow rate sensors interfaced with microcontrollers as processing units to detect and report theft. Mechanisms for remote alerting have also been proposed in their work [9].

The different research initiatives herein discussed conform to the idea that any form of distribution pressure alteration or interruption in the distribution network can significantly affect water supply. However, it should be noted that not all forms of water theft directly interferes with distribution pressure. The major form of pressure disruption is illegal connection. With other issues such as pipe bursts and leakages, water pressure can still be interfered with and hence reported as theft [21]. With activities such as meter tampering, meter reversals and meter bypass where the water pressure is not affected directly, such systems are rendered inefficient in detecting the problem. This remains a major research gap which this work aims at closing. Using data captured from a distribution network over a period of time, a trained Machine Learning model can be used to predict the existence of theft more accurately.

Till now, no work has been done on piped water theft prediction using Machine Learning models. In this work the latest technology is utilized to efficiently solve the piped water theft problem using server side processing.

A Machine Learning theft prediction model using Random Forest Classier has been proposed. The model was trained and tested on data collected by remote flow rate sensors distributed over the network over a period of 3 hours and its performance evaluated on four indexes, that is; accuracy, precision, recall and F-measure. We further evaluated our model using benchmark methods, that is; Support Vector Machine (SVM), Logistic Regression (LR) and K-Nearest Neighbour (KNN).

A. State of the Art

The existing traditional piped water theft detection systems are not comprehensive in nature; that is, they can only detect theft in cases where pressure is interrupted. In scenarios where the client themselves perpetrate the theft by manipulating the water meters, such methods are unable to detect the theft since no form of pressure disruption may be caused in the distribution network. These methods are also based on hard coded rules running on the microcontrollers which can fail in many cases. Additionally, their inability to adapt to the continuous changes in the water distribution patterns is another drawback. Therefore, in this project, we track the patterns of water distribution over time and train a Random Forest Classifier which can use the received data from flow rate sensors laid along the distribution network to predict changes in flowrate along the network.

III. MATERIALS AND METHODS

In this section, a thorough description of the proposed random forest classifier has been made followed by a discussion of the Benchmark models that is SVM, LR, and KNN.

A. Random Forest

RF is a statistical learning model which works well with small to medium data compared to Neural Network which require large training data [22]. Random Forests can be precisely defined as a combination of tree predictors in which each tree is dependent on values of an independently sampled random vector whose distribution is same for all the trees within the forest.

The traditional classifier Model is vulnerable to over fitting and has quite a limited accuracy. To improve the classification accuracy, multiple models are aggregated; a method referred to as ensemble or classifier combination [23]. This involves integration of several single classifiers also referred to as weak classifiers. The results of classification of all weak classifiers determine the final classification result. This involves an ensemble of trees which are grown and allowed to vote for the majority class. Random vectors governing the tree growth in the ensemble are generated. This is a way of growing the ensembles.

Bagging is an early example whereby to grow each and every tree in the ensemble, random selection is made without replacement from the training set examples. Secondly, the split selection involves selecting the split randomly from among the p best splits and this is done at each node [24]. The third approach involves selecting the training set from a random set of weights on those examples within the training set.

RF has the capability of processing continuous data and data of high-dimensionality and discreteness. The results of random forest are less affected in noisy data. A great number of theories and research experiments have proved random forests to have the ability of generating high accuracy and efficiency. Additionally, random forest is one of the most popular frontier research fields in data mining and bioinformatics, among others [25].

B. Benchmark Methods

i) Support Vector Machine

Support Vector Machine (SVM) is a supervised learning model that performs classification by constructing hyperplane in a high dimensional space that maximizes the margin between classes [26].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

The most important idea here is to obtain a partition hyperplane that has a maximum margin in sample space. The main aim of SVM method is mapping the sample space into a high dimensional feature space or even infinite dimension sometimes referred to as Hilbert space.

This is done through nonlinear mapping techniques such that nonlinear separation problem in original sample space is transformed into a linearly separable problem in the feature space [27].
Given a training dataset $T = \{(x_1, y_1), \ldots, (x_N, y_N)\}$, where $x_i = (x_{i1}, x_{i2}, \ldots, x_{im})T$ is a sample, y_i is the label of the sample, N is the number of samples, and m is the number of attributes. The partition hyperplane can be described as
$f(x) = w^T x + b$, where $w = (w_1, w_2, \ldots, w_m)T$ (1)
is the normal vector and b is the displacement.

ii) Logistic Regression

Logistic Regression is a supervised learning classification algorithm well known to be the gold standard prediction method [28]. LR is a statistical model that describes the relation between predictor variables denoted by $x' = (x_1, x_2, \ldots, x_p)$ and a response variable, which is a two value categorical variable [29], that is, the "normal flow" or "abnormal flow" as investigated in the study. The conditional probability of occurrence of water theft (abnormal flow) can be written as $K(Y = 1|x) = \pi(x)$; then, the LR model for k predictor variables can be written as;

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}} \quad (2)$$

Where $0 \leq \pi(x) \leq 1$
The logit transformation is a useful transformation of logistic regression, defined as below;

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (3)$$

The odds ratio (OR) associated with one unit change in x_j is represented with $e^{(\beta_j)}$

iii) K-Nearest Neighbour

K-Nearest Neighbour (KNN) is a simple instance-based classifier that stores all cases and assigns the document to the majority class of its *k* nearest neighbours [26]. KNN algorithm is sometimes also referred to as; case based reasoning, example based-reasoning, instance based learning, memory-based reasoning and lazy learning. KNN has been used in many applications such as pattern recognition and statistical estimation, among others.
The method used by KNN to classify data is non-parametric in nature that is; does not exploit the underlying probability distribution assumptions of the data set. Two broad categories of KNN are recognized; structure less NN techniques where entire data is grouped into training and test sample data and structure based NN techniques which are based on some of the structures of data such as; orthogonal structure tree (OST), axis tree, ball tree, nearest future line and central line [30]. Distance is evaluated from training point to sample point with the lowest distance point referred to as nearest neighbor.
The commonest application of KNN has been seen in datasets with continuous attribute data. The algorithmic steps of KNN proceed as follows.

1. Finding the *k*-training instances that are closest to an unknown instance
2. Picking the classification that is most commonly occurring for the *k* instances.

Various ways are used to measure similarity between two instances of n attribute values. For each and every measure, the following three requirements apply.
Assume dist (p, q) be the distance measure between two points p, q then;

$$dist(p,q) \geq 0 \; and \; dist(p,q) = 0 \; iff \; p = q \quad (4)$$

$$dist(p,q) = dist(q,p) \quad (5)$$

$$dist(p,r) \leq dist(p,q) + dist(q,r); \quad (6)$$

Equation (6) is referred to as the "triangle in equality". It states that the shortest distance between any two points is a straight line [31]. Z score standardization and minimum-maximum normalization are used for continuous [31] data. KNN suffers from a number of drawbacks. These include; low efficiency and dependency on the selection of good values for *k* [33].

IV. Project Description

A circuit prototype was designed in proteus, with two Arduino microcontrollers designated as slave and master programmed and interfaced with flowrate sensors, GSM (Global System for Mobile Communication) modem, and LCD (Liquid Crystal Display). A brief description of these components is presented in the next section. Experimentation was done over a 3-hour period and water flow rate data collected in 10 seconds interval. This data was tabulated, preprocessed and analyzed. 80% of the data was used for training a Random Forest classifier and 20% was used for testing purposes. The model was evaluated on accuracy, precision, recall and F-measure. Training and Testing was also done on benchmark models that are SVM, LR and KNN and the resulting performance measurements were compared with the proposed model.

A. Design

i) Block Diagram

An experimental prototype was designed as shown in figure 2 and used for data collection. The system adopts a nominal value of flow rate after a delay of about 2 minutes. This is in line with the water supply mechanisms by water

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

distribution companies which strives to supply water in a given pressure zone at a range of predefined pressures. Though some fluctuations may occur, the distribution pressure cannot fall above or below a certain value [32].

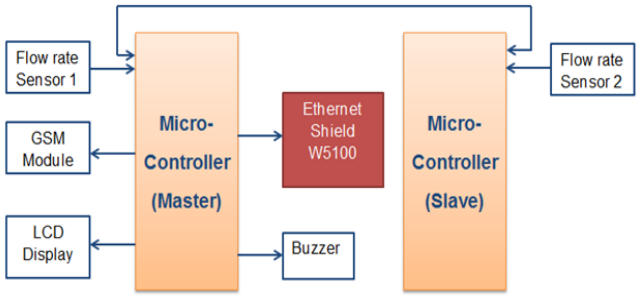


Fig. 2. Block diagram of the designed prototype

ii) Flow chart

The flow of events followed in operation of the prototype for data collection is as shown.

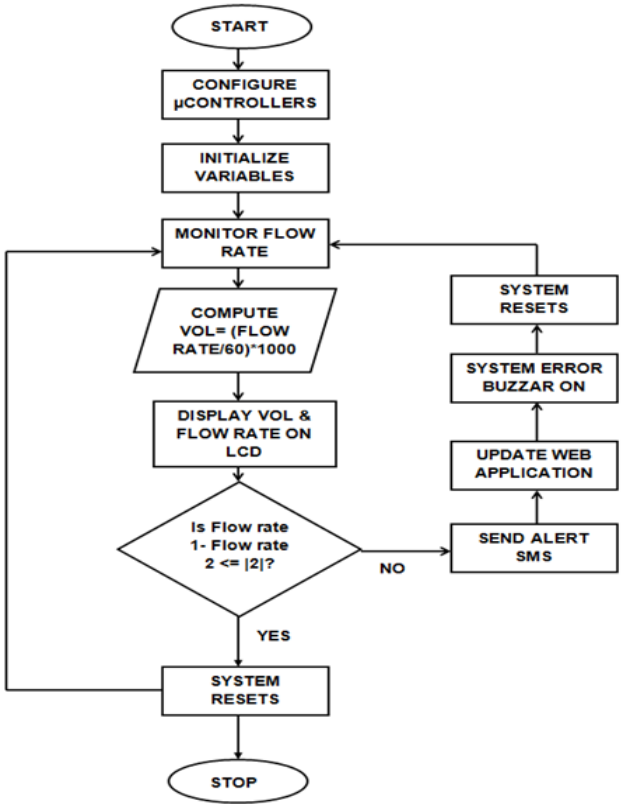


Fig. 3. Flow chart

B. Components Description

i) Arduino Uno

The functionality of Arduino is based on ATmega328 datasheet. It has 14 pins (input or output pins) referred to as digital pins. There are 6 analog input pins, 16 MHz ceramic resonator, power jack, USB connection, an ICSP header, and a reset button. With the use of the Arduino board, the microcontroller can be powered by a computer using a USB cable or directly with a battery or AC to DC adapter. The distinctive feature of the Arduino Uno from other boards that precedes it is that it has the FTDI USB serial driver chip [33].

Fig. 4. Arduino Uno

ii) Flowrate Sensor

The flow rate sensor functions on the principle of Hall effect which states that a voltage difference is induced in a conductor transverse to the electric current and the magnetic field perpendicular to it [34]. This idea is applied in the flow rate sense using a rotor shaped like a propeller. A liquid flowing through the sensor pushes against the fins of the rotor thereby causing a rotation. The rotation of the rotor induces a voltage producing an output of about 4.5 pulses from the sensor for every volume of 1 litre of water passing through the rotor per minute. The magnet attached to the rotor shaft generates a changing magnetic field which induces this voltage [37]. The flow rate sensor's data cable (yellow) is connected on pins D2 and D3 of the Arduino Uno which are also called interrupt pins.

Fig. 5. Flowrate sensor

iii) GSM Modem

SIM900 is a quad band GSM modem. The modem board communicates via UART as it has pins labeled TXD and RXD for UART communication. The board is powered by 12V but it also has a 5V line for Arduino power. It uses AT (attention) commands for its communication control.

Fig. 6. GSM Modem

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

V. CIRCUIT DESIGN AND EXPERIMENTATION

The water theft detection system was designed and simulated in proteus software with two Arduino Uno microcontrollers programmed as Master and Slave because every Arduino Uno can support a single flowrate sensor. The Master controls the entire circuit including the slave microcontroller which is instructed by the master.

The GSM module was interfaced with the Master and uses serial communication to send sms alerts to a Mobile phone. The SMS states the flow rate along a given service line and the volume consumption at that instant.

We further performed system testing on the complete integrated system for validation purposes. To test the compliance of the complete system to the system requirements, we fitted water pipes of 20mm diameter with flow rate sensors connected at branch points as shown in figure 7. Water was then supplied through the system using a water reservoir placed at an elevation from the ground surface.

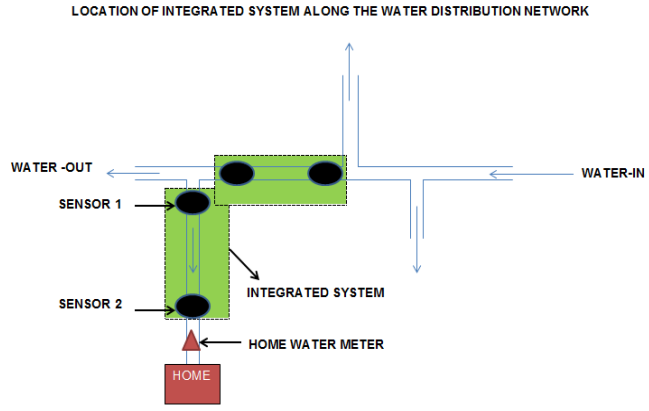


Fig. 7. Conceptual Water Distribution Network

A. *Data Collection*

Experiments were ran to depict the real life scenario of water distribution and flow rate values were read from the LCD in a 10 seconds interval. The water flow rate was varied by connecting a suction pump to the distribution Network and alternating it with a diversion to vary the pressures. The alert messages of abnormal flow were received on the Mobile phone and a record taken accordingly.

After 3 hour period, 1017 values were captured and tabulated. Some variables such as absolute difference between the established norm value at the instant and the recorded flow rate, volume consumption were computed and included in the dataset.

Absolute difference, A_d = |flow rate – established norm value|

Volume flow at the instant = flow rate * time

A total of 6 variables were recorded in our dataset.

B. *Data Analysis and Processing*

We noted that after about 2 minutes of water flow, there was stability in the flow rate values read. Every time the system detected abnormality in the flow, there was reset and therefore the system would adopt a new value of flow rate designated as the normal flow. Water theft was based on these fluctuations; the bigger the fluctuation, the more likely was the conclusion that there was error in the distribution system. Taking the adopted normal values at every system reset point; the water flow pattern shown in figure 8 was observed.

This pattern aligns with the idea that water distribution companies supply water at a given fixed range of water pressures in each and every pressure zone which is the major basis for theft detection [9]. A pressure district is a collection of varying pressure zones and there can be varying distribution pressures within a pressure district, however, the fluctuation does not go outside the normal ranges as depicted in our flowchart presented in figure 3.

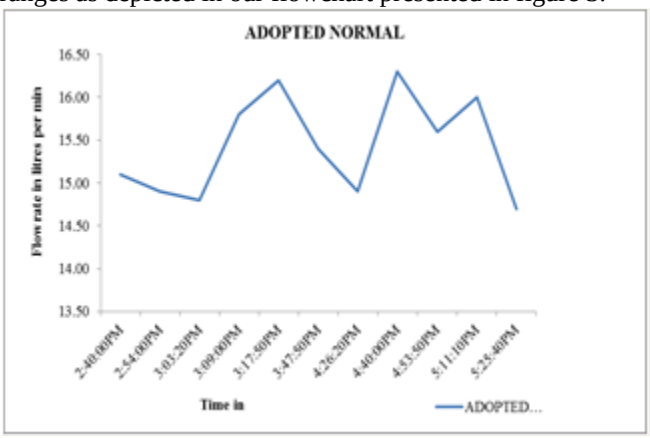


Fig. 8. Water flow pattern

From the graph obtained, it can be observed that the maximum fluctuation of the flow rates throughout the entire experiment was (16.3-14.7= 1.6litres/min).

VI. MODELING AND PREDICTION

The model training and testing took a number of tasks which involved Importation of the necessary libraries that is; numpy, pandas, matplotlib, and seaborn after installing dependencies in anaconda. Jupyter Notebook was used with Python 2. A thoroughly cleaned dataset in csv format was read. Understanding the data patterns is an Important step in data analysis [35] hence data exploration and visualization was done. To improve the performance of the classifiers, some feature engineering was undertaken. Here, domain knowledge and data mining techniques are used to extract features from raw data.

Feature engineering is also considered as some form of applied machine learning. In our experiment we have constructed Random Forest Classifier and other benchmark predictive models which included; SVM, KNN and LR. We used 80% of the data for training purposes and 20% for testing. Finally predictions were drawn from test data and performance measures, that is; accuracy, precision, recall and F-measure were computed from the confusion matrix.

A. *Model Evaluation*

The water theft prediction model performance was evaluated using accuracy, precision, recall and F-measure

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

using the confusion matrix. A preliminary discussion of confusion matrix presented here is necessary in enabling us understand the performance metrics used for performance evaluation in our study.

Table 1 shows a confusion matrix illustrated with four main quadrants, which include; True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). The number of correctly classified positive samples is indicated my TP, the classification model may predict some negative samples incorrectly as positive, these are referred to as FP where as those positive samples that are predicted incorrectly as Negative are represented by FN. However, the numbers of negative correctly predicted samples are represented as TN [36].

Table 1: Confusion Matrix

Confusion Matrix		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

Accuracy refers to the ratio of the number of correctly classified samples to the total number of samples. The harmonic mean of precision and recall of the classification is called F-measure [36]. In this definition, precision is described as the ratio of all positive results correctly classified to the total number of positive results in the classification and recall describes the ratio of correctly classified positive results to the actual number of positive results that would have ideally been returned. These evaluation performance metrics were calculated using the following formulae.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Accuracy = \frac{TP}{TP+FN} \quad (9)$$

$$F-Measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (10)$$

VII. RESULTS AND DISCUSSION

The results of our experiment showed that the proposed classification model achieves the highest accuracy and recall. KNN achieved a similar accuracy though its recall was lower than that of the proposed mode as shown in table II.

Table- II: Performance metrics values

Method	Accuracy	Precision	Recall	F-Score
RF	0.97	0.95	0.98	0.96
SVM	0.96	0.98	0.93	0.95
LR	0.96	0.98	0.93	0.95
KNN	0.97	0.95	0.96	0.95

A visual graphical representation of the variation of performance scores of the four classification models is shown below as a bar graph.

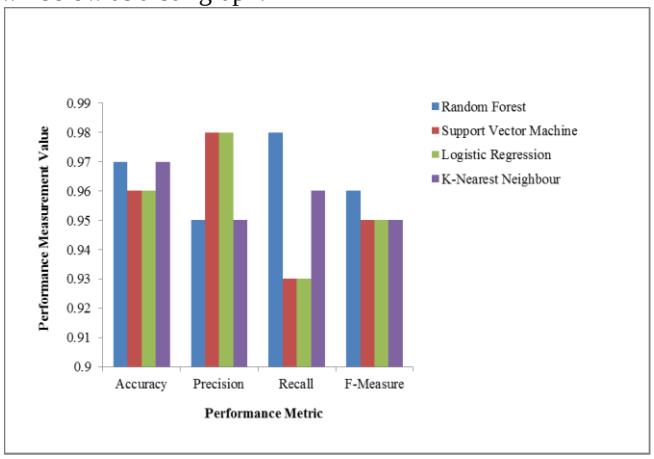


Fig. 8. Performance metric results of the prediction models

VIII. CONCLUSION

The existing research initiatives depend solely on hardware to deduct the occurrence of piped water theft. These works have basically made use of flow rate sensor data collected from water distribution and fed into the microcontroller memory to deduce the state of the network based on hardcoded rules. In this study, we have resolved the issue by integrating hardware and software approaches to intelligently predict theft. The programmed electronic circuit was used as a data collection module for the purposes of training the Random forest classifier. The data collected over a 3 hour period in 10 seconds intervals gives a fair representation of real life scenario of piped water distribution. This data was cleaned and preprocessed which gave a total of five variables used in the study. These were; flowrate, adopted norm value, absolute difference between adopted norm and flowrate value at the instant, volume and status of the distribution system. The data was then divided into training set and testing set which was 80% and 20% of the dataset respectively.

The Random Forest classifier was then constructed and evaluated on four performance metrics, that is; accuracy, precision, recall and F-measure. Comparing the results with three competitive approaches that is, SVM, LR and KNN used in our experiment, no considerable differences were noted other than the recall value which was highest in RF model. Reliable feature importance estimate and estimation of test error with efficiency that no repeated model training is required are two notable advantages of RF compared to the benchmark approaches. Our trained model can be incorporated in existing systems at the server end to intelligently deduce the existence of water theft. For deployment purposes, water flowrate values along distribution lines can be collected over a period of time and used to generate a dataset that can then be applied in the model training. In such a case, unsupervised techniques may be applied.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

FUTURE WORK

The data collected depicted a common pattern of water flow which enabled quick model training. The high accuracy obtained ascertains the efficiency that can be achieved with Machine Learning approaches. Flow rate data from water distribution networks can be collected for a reasonable period of time and be used to model and predict water theft using our model or unsupervised learning techniques. As an initiative to eradicating piped water theft problem, a complete system can be further developed.

REFERENCES

[1] The United Nations World Water development report (UNWWD), water for sustainable world, UNESCO, United Nations Chapter 4, 2015

[2] Pradeep K. Naik, Water crisis in Africa: myth or reality?, International Journal of Water Resources Development, ISSN: 0790-0627 (Print) 1360-0648, 2016 (Online) Journal homepage: http://www.tandfonline.com/loi/cijw20

[3] World Health Organization, WHO *World Water Day Report*, 2015. Retrieved from http://www.who.int/water_sanitation_health/takingcharge.html

[4] W. Gumindoga, G. Takawira, L. Fengting, N. Innocent & M. Clifton, Health Safety of Drinking Water Supplied in Africa: A Closer Look Using Applicable Water Quality Standards as a Measure, Springer Science+Business Media B.V. 2017

[5] Review of Sector Reforms and Investments , Access to Water and Sanitation in Sub-Saharan Africa, Key Findings to Inform Future Support to Sector Development, Synthesis Report, GIZ Competence Center Water, Wastewater, Solid Waste, Eschborn, January 2019

[6] United Nations Department of economic and social affairs (UNDESA), "International Decade for Action 'Water for life'," United Nations, New York, 2005-2015.

[7] Ministry of water and Environment, "Uganda Water and Environment Sector Performance Report," Government of Uganda, Kampala, 2018.

[8] National Water and Sewerage Corporation (NWSC), "National Water and Sewerage Corporation (NWSC)," NWSC, 23 06 2013. [Online]. Available: http:www.nwsc.com [Accessed 7 05 2015].

[9] H. A. Gaikwad & P. V. G. Puranik, "Automated urban water supply system and theft identification," International Journal of Electronics and Communication Engineering & Technology (IJECET), vol. 6, no. 6, pp. 145-156, 2015.

[10] J. Tharanyaa, A.Jagadeesan & A.Lavanya "Theft identification and Automated Water Supply System Using Embedded Technology," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 2, no. 8, pp. 3727-3734, 2013.

[11] P. D. B. Madihalli & P. S. S. Ittannavar, "Smart Water Supply Management," International Journal of Emerging Trends in Electrical and Electronics (IJETEE – ISSN: 2320-9569), vol. 10, no. 9, pp. 77-79, 2014.

[12] V. Felbab-Brown, "Brookings Mountain West Lecture Series, University of Nevada, Las Vegas," University of Nevada, Las Vegas , 20 02 2015.[Online].Available: https://www.unlv.edu/brookingsmtnwest/past-lectures-events. [Accessed 25 10 2015].

[13] National Water and Sewerage Corporation, "National Water and Sewerage Corporation (NWSC)," Littlegate publishing, 01 04 2015. [Online]. Available: http://www.littlegatepublishing.com. [Accessed 11 10 2015].

[14] G. M Tamilselvan and V. Ashishkumar, "IOT Based Automated water distribution system with water theft control and water purchasing system," International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, vol. 7, no. 4, pp. 151-156, 2018.

[15] R. Malhotra, "An empirical framework for defect prediction using Machine Learning techniques with Android software," Applied Soft Computing, Elsevier, vol. 49, pp. 1034-1050, 2016

[16] A. E. shiled, "Arduino Ethernet shield," 24 04 2010. [Online] Available: https://docs.rs-online.com/0084/0900766b80db991d.pdf. [Accessed 27 04 2020].

[17] A. Panchal and K. Degate, "Automated water supply system and water theft identification using PLC and SCADA," International Journal of Engineering Research and Applications ISSN : 2248-9622, vol. 4, no. 4, pp. 67-69, 2014.

[18] A. Mancharkar & K. Rakesh, "Automated water distribution system for smart city using PLC and SCADA," International Journal of Emerging Technologies and Engineering (IJETE), vol. 3, no. 3, 2016

[19] A. Gantala & P. Nalajala "Public water supply monitoring to avoid tampering and water man Fraud using smart water meter system," nternational Journal of Mechanical Engineering and Technology (IJMET) ,vol. 8, no. 8, pp. 1194-1201, 2017.

[20] A. Rahate & G. Ashwini, "Automated water distribution system and theft detection," International Journal of Advance Engineering and Research Development, vol. 4, no. 4, pp. 650-654, 2017.

[21] M. Panwar "Issues, Challenges and Prospects of Water Supply in Urban India," IOSR Journal Of Humanities And Social Science (IOSR-JHSS), vol. 20, no. 5, pp. 68-73, 2015.

[22] R. A. Berk, "Statistical Learning from a Regression Perspective," Springer, Switzerland, 2019

[23] F. K. N. &. Z. J. P. Jian-Bina W U, "A review of Technologies on random Forests," Statistics and Information Forum , vol. 3, pp. 32-38, 2011.

[24] L. Breiman, Random Forests, CA 94720 : Statistics Department University of California Berkeley, 2018.

[25] H. L. L.-W. &. L. M. R P Adams, "A physiological series dynamics based approach to patient monitoring and outcome prediction," EEE journal of Biomedical and Health Informatics,, vol. 19, no. 3, pp. 1068-1076 , 2015.

[26] K. S. R. &. D. M. Harikrishna, "Children's Story Classification in Indian Languages Using Linguistic and Keyword-based Features," ACM Trans. Asian Low-Resour. Lang. Inf. Process, vol. 19, no. 2, p. 22 pages, 2019.

[27] M. Z. &. H. Shi, "A Novel Support Vector Machine Algorithm for Missing Data," ACM, 2018.

[28] T. B. &. D. P. S. Briggs, "Decision trees for predicting risk of mortality using routinely collected data," International Journal of Social and Humanistic Computing, vol. 6, no. 6, p. 303–306, 2012.

[29] B. M. &. K. M. W. K. J Wang, "Modeling and predicting the occurrence of brain metastasis from lung cancer by Bayesian network: A case study of Taiwan," Computers in Biology and Medicine, vol. 47, p. 147–160 , 2014.

[30] N. B. &. Vandana, "Survey on nearest neighbor techniques," International Journal of Computer Science and Information Security (IJCSIS), vol. 80, no. 2, 2010.

[31] M. &. B. D. P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," in International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA), Delhi, 2013.

[32] M. J. P. (. C. University, "Performance assessment System, Basics of Water Supply System, Training Module for Local Water and Sanitation Management," Maharashtra Jeevan Pradhikaran (MJP) CEPT University, Maharashtra, 2012

[33] A.Uno-Farnell, Arduino Datasheets, Ivrea: Italy, https://www.farnell.com/datasheets/1682209.pdf.

[34] Arduino, "How to interface an Arduino Uno with a Flow rate sensor to measure a liquid," Arduino, 23 03 2018. [Online]. Available: ttps://maker.pro/arduino/tutorial/how-to-interface-arduino-with-flow-rate-sensor-to-measure-liquid. [Accessed 16 09 29].

[35] A. Z. &. A. Casari, Feature Engineering for Machine Learning Principles and Techniques for data scientists, Sebastopol, California: O'Reilly Media, April, 2018 First Edition.

[36] Z. X. &. F. Y. S. Yang, "A patient outcome prediction based on Random Forest," ACM ISBN 978-1-4503-7640-2/19/1 , 2019.

[37]"Working with water flow sensors and Arduino," 05 08 2019. [Online].Available: https://www.electroschematics.com/12145/working-with-water-flow-sensors-arduino. [Accessed 23 12 2019].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-08922-6

Movie trailer classification using deer hunting optimization based deep convolutional neural network in video sequences

Prashant Giridhar Shambharkar [1,2] · M. N. Doja [2]

Received: 3 May 2019 / Revised: 24 February 2020 / Accepted: 7 April 2020
Published online: 02 May 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In current situation, video classification is one of the important research domains. Since video is a complex media with various components, classification of video is normally a complex process. This paper presented a human action based movie trailer classification using optimized deep convolutional neural network in video sequences. Initially, images are converted into the grayscale conversion. Using the adaptive median filtering process, the pre-processing stage is accomplished. Threshold based segmentation approach is utilized for subtracting the background from the video frames and to extract the foreground portion. In the feature extraction stage, the visual features (color and texture features) and motion features are extracted from the segmented portion. Finally, the mined features are trained and classified with the help of optimized deep convolutional neural network (DCNN) for the movie trailer classifications. Here, the deer hunting optimization (DHO) is introduced to optimize the weight values of DCNN. The proposed (DCNN-DHO) human action based movie trailer classification is executed in the MATLAB environment. The experimental results are evaluated and compared with the existing methods in terms of accuracy, false alarm rate, sensitivity, specificity, precision, F-measure and false discovery rate. The results of the proposed method are compared with filtering process and without filtering process in which 95.23% of accuracy is achieved for the suggested approach with filtering and 90.91% of accuracy is achieved for the suggested approach without filtering process.

Keywords Adaptive median filtering · Background subtraction · Deep convolution neural network · Movie trailer classification

✉ Prashant Giridhar Shambharkar
prashant.shambharkar@dtu.ac.in

M. N. Doja
mdoja@jmi.ac.in

Extended author information available on the last page of the article

1 Introduction

Human action recognition is the main structure in the different applications of natural human-machine interaction, video surveillance, digital entertainment, multimedia content analysis and retrieval [1]. Using the motion dynamics and visual appearance of the involved human and objects, various human actions can be characterized. According to intrinsic property, different popular datasets are utilized for human action recognition. By distinguish the temporal motion and spatial appearance, computer vision investigators have been emerging enhanced visual features to identify the human actions in video sequences. For expressive the spatially and temporally coupled action patterns, several existing schemes improve the diverse spatiotemporal features. As video sequences can logically be seen as three-dimensional (3D) Spatio-temporal signal [2]. The prevalent convenience of high-speed internet causes the video to become acquainted with information medium on the web. For consumers and producers, the fast-growing rate of supply and demand for video content generates a compelling task. A large number of videos are formed and indexed are increasing at a fast rate. The required amount of time is needed to watch these videos [3]. Automatically analyzing videos can help humans in solving distinct issues. This task is luxurious at present for them to execute alone [4].

Video information is a sequential data and moving content of video provides the absolute data. Several methods are available to obtain temporal information such as dense routes, local motion features and Spatio-temporal volumes [5]. The vast amount of movie information is formed. Then the user wants a few specific tools to recover that data. Within a certain class or genre, users can use their decisions to select a movie [6]. Producing a trailer is one of the video summarization. At first, it analyzes the content of the video, then filters the significant video clips and lastly merges the chosen video clips into a summary video. Video summary method classified into two types such as static and dynamic video summary. Movie trailers are still significantly distinct from other video summarization approaches. Creating a movie trailer process to gathers the features of commercials to promote to attract an audience. In the special highlights, the trailer is more essential to express the film and generate an artistic atmosphere. Generally, the high-quality trailer includes the five attributes such as, dialogue scenes, prominent storylines, the entire artistic atmosphere, making suspense by eliminating the story ending and key objects [7].

The manual manufacture of a movie trailer is a difficult and time-consuming task. As stated by the independent article, the invention period of the movie trailer could be accomplished within three to four months. Generally, the production companies are trying to identify some attractive keywords for creating a movie trailer, it will be guiding a user to obtain a ticket and watch their movie [8]. In the fast-developing database, automatically forecasting the semantic content of videos for searching and indexing become extremely applicable. To accomplish the action recognition in such videos, distinct algorithms are utilized that are both simple and fast to train at a similar period. Also, the approaches are robust to noise in the real-world nature of such videos [9]. The goal of action recognition is to identify the classes of the actions present in a video, independently of the background [10]. For action recognition, the modeling latent temporal structures of actions are mostly utilized [11]. The large diversity of viewpoints of the captured human action information is a major difficult task for human action recognition [12].

The methodologies of human action recognition can be characterized into two types such as detection recognition and classification of video. The motion of a person's action is detected in the process of detection recognition. Based on the action, the videos are categorized in the methods of classification. By newly advanced datasets with videos, these methodologies are estimated. These datasets are either collected from the web, YouTube or video clips that were noted in realistic situations where lighting conditions are not controlled [13]. In recent years, many researchers using a deep learning scheme such as CNN, RNN, and LSTM. Deep methods naturally need massive samples and time to learn the constraints of models [14]. Recently, DCNNs has solved the different problems in the video analysis that are the best standard schemes. These initial outcomes are stimulating and pave the route for many applications to be entirely explored [15].

Contribution This paper presented the movie trailer classification in video sequences. The main objective of this work is to give an accurate classification. At first, pre-processing is performed using the adaptive median filtering method. Then the threshold based segmentation approach is utilized for background subtraction. Different features like color, texture and motion features are extracted from the frames individually. All the mined features are given to the optimized DCNN method. To optimize the weight values, the DHO algorithm is used in the DCNN based classification. It is trained based on the features and classified the movie trailers in terms of action, comedy, horror, sci-fi and crime. DCNN-DHO is accurately classified the movie trailers. The proposed method compared with filter and without filtering process. Also, the proposed (DCNN-DHO) scheme compared to the existing methods. The several performances are evaluated and compared to the existing methods. Motivation behind to this work are this approaches described in earlier indicated an effort to comprehend the textual content, audio-video contents, or both together. Still, movies and trailers exists in a variety of forms due to the different cultural environments. Therefore, the proposed method make movies' trailers should be adaptive to suit the different cultural environments.

Novel optimized classifier (DCNN-DHO) concept minimizes the error during classification. This optimized classifier identifies the weights and biases arbitrarily between input and hidden neurons. Ordinary deep learning algorithms learns only these weights by providing the solution to the linear model. Nevertheless, this new concept, delivers improved performance in simplification, and its training time is extremely effective. Optimized deep learning is depending on empirical risk reduction principle. In some conditions, such as inadequate samples, the risk reduction principle is substandard from a statistics point of view. Also it provides added robustness, and accomplishes improved results for linearly non-separable samples. Thus novel hybrid classifier necessitates human involvement and might create an impact over the classification accuracy.

It have more applications in areas such as intelligent space, human-machine interaction, robotics, elderly care, virtual reality, etc. Also, it has a key role in different computer vision tasks include video retrieval, video annotation, etc. A large number of research possibilities such as action recognition, context analysis, and many other tasks are exposed in video analysis.

Organization Rest of this paper is arranged as follows: Section 2 includes the related works, Section 3 gives the problem definition, Section 4 explains the proposed methodology, Section 5 provides the experimental analysis, and Section 6 contains the conclusion of our work.

2 Related works

The challenges associated with the analysis of emotional content of the videos which were user generated and edited professionally were addressed by Esra et al. [16]. Both the modelling aspects and representation were concentrated by the authors. Using the audio and visual features in the mid-level the videos were represented. Using the convolutional neural networks, more specifically from the raw data the audio and visual features were learned automatically. Furthermore, SentiBank domain-specific and dense trajectory based motion features were included. The videos were categorized into one of the predefined emotion classes through the fusion methods and ensemble learning. Two datasets include subset of DEAP dataset and Video Emotion dataset were used for the performance evaluation. The experimental outcomes by using both the datasets showed that, for the video affective content analysis the ensemble learning was superior to the SVMs.

On the basis of affective impact and genre of movie to predict the interestingness of video shot, a multimodal framework was proposed by Olfa Ben-Ahmed and Benoit Huet [17]. They assumed that, within the complete media the influence and emotional characteristic of a video conclude the genre is one of the reasons for recognizing the interestingness of the specific video. In a representation of mid-level the multimodal content was evaluated over different genres such as sci-fi, romance, horror, drama and action by explaining every audio-video shot as a distribution. Some shots were more interesting than others which included neutral genre content because of more characteristic of the media. They considered a dataset named as MediaEval2017 Media Interestingness Prediction Task for the evaluation of their developed method. They proved that on this dataset, their proposed method outperformed the other existing methods with respect to the mean average position.

Considered collaboratively the electroencephalogram signals and the features of video with bidirectional Long Short-Term Memory network (Bi-LSTM) which was a multimodal, Takahiro et al. [18] have developed an approach for the video classification based on the information of what a user likes (WUL). On the basis of WUL it was difficult to categorize the videos because the user defined labels of the videos are small in number. Hence, along with the video classification based on WUL they applied a new transfer learning. And also to show the psychophysiological signal representation evaluated from the Bi-LSTM they conducted a user study which was very efficient for the classification based on WUL.

In the context of the classification of movie trailers genre the convolutional neural networks are explored by Gabriel et al. [19]. First they have created a new dataset for the movie trailers. The developed dataset have above 3500 movie trailers with the known genres and also the dataset was made as public for the interested readers. Then they provide a detailed new classification approach which encapsulated the CNN framework for the classification of movie trailer genre. This method was named as CNN-Motion and the results were evaluated and compared with the other advanced approaches. The results from the developed approach proved that it outperformed the other compared approaches.

To classify the multi-label movie trailer genre on the basis of CNNs a new deep neural approach was developed by Jonatas Wehrmann and Rodrigo C. Barros [20]. With the residual connections an ultra-deep convolutional network was encapsulated. This approach used a convolution layer which was special to extract the features based on images contain sequential information to map the genres and movie trailers. Their proposed approach was compared with the other advanced approaches which were developed for the classification of movie genre employed with handcrafted features which were in low level and common image descriptors.

The results proved that, for all the movie trailer genres the classification performance of their developed method outperformed the other approaches.

A deep network was intended by Joe et al. [29] for the classification of video. The numerous deep neural network (DNN) structures were utilized to merge the image data across a video over a longer period than early endeavored. Two approaches were employed that can handle the full-length videos. The first technique determines the several convolutional temporal feature pooling structures and investigating the different design selections after familiarizing the CNN for this task. The second scheme represents the video as a well-ordered sequence of frames. For this reason, RNN (recurrent neural network) is utilized with LSTM cells that are linked to the primary CNN output.

A DCNN architecture was introduced by Rigoll et al. [30] for background subtraction. The recommended scheme was presented with a new background subtraction system with DCNN to achieve the process of segmentation. Here, the approach avoids parameter tuning and feature engineering. By training a single CNN, the system constraints could be learned from data that can handle different video scenes. Besides, the recommended method to evaluate the background model from the video. Generally, 5% of video frames were utilized randomly for the training of CNN. Besides, spatial-median filtering is used as the post-processing of the system outputs.

A recurrent convolutional network was developed by Paul et al. [31] for video-based person re-identification. This work presented a new RNN design for person re-identification. Using CNN, different features are mined from every frame given the video sequence of a person. CNN contains a recurrent final layer that permits data to flow amongst time-steps. For the complete arrangement, mined features are combined by temporal pooling to give an overall appearance. For video-based re-identification, recurrent layer, convolutional system, and temporal pooling layer are equally trained using a Siamese network design to act as a feature extractor.

An automatic emotion-based music classification was presented by Jun and Seo [32] for supportive intelligent IoT applications. To categorize the music with high accuracy according to the sensitive range of people, an emotion-based automatic music classification scheme is utilized. Based on the sentiment of humans, the recommended method is constructed to categorize them automatically after new songs are added to music related IoT applications. This is a major problem for emergent applications. For gathering sentimental information, a review was accompanied depends on the emotional model. In a small and medium-sized enterprise, music features are derived by conferring with the functioning group. Using a support vector machine (SVM) and multiple regression investigation, classification of emotion is accomplished.

A 3D-based DCNN was introduced by Liu et al. [33] for action recognition with depth sequences. This work investigates the deep learning-based action recognition through depth sequences and the equivalent skeleton joint data. From raw depth sequences, to directly study the Spatio-temporal features, a 3D based DCNN method is utilized. For every sequence, a joint vector is calculated based on the simple position and angle information among skeleton joints. At last, the results of SVM is obtained from 3D based DCNN features and joint vector are combined for action recognition.

Learning a deep compact image representation was presented by Wang and Yeung [34] for visual tracking. This work investigates the stimulating issue of tracking the route of a moving object in a video. Compared to earlier methods, the recommended scheme just learn the appearance of the tracked object online. Several methods are stimulated by recent

improvements in deep learning structures to solve the feature learning issue. A stacked denoising autoencoder is trained using auxiliary natural images to learn generic image features. A knowledge transfer approach is utilized in the online tracking procedure. Online tracking contains a neural network as a feature extractor that is made from the encoder portion of the trained autoencoder.

Using DCNN, an action recognition from depth maps was presented by Wang et al. [35]. For human action recognition, three-channel DCNN and weighted hierarchical depth motion maps (WHDMM) are utilized on small training datasets. In mining discriminative features, three methods are established to control the ability of ConvNets for recognition. By spinning the 3D points of the captured depth maps, several viewpoints are mimicked. To encode the spatiotemporal motion patterns of actions into 2D spatial designs, WHDMMs are made at different temporal scales. By changing the WHDMMs into pseudocolor images, the 2D spatial designs are further developed for recognition.

3 Problem definition

Automatic analysis of videos via image-processing and machine learning techniques is the major problems identified in much-studied research theme. Automatic movie trailer genre classification is the one possible task performed in video analysis. Video sequences provides action details in deep rather than still image. In the areas of computer vision and pattern recognition, the amount of interest has increased by the detection of actions in the videos which are induced by the humans.

4 Proposed methodology

We proposed a DCNN-DHO approach for the classification of movie trailer and related event analysis. This proposed method consists of four phases which include pre-processing, segmentation, feature extraction and classification.

In fig. 1 architecture of the proposed method is given. Here in the first stage, the videos in the database are divided into frames. Then pre-processing is done to clear the noise. The pre-processing stage uses an adaptive median filter with grayscale conversion scheme to remove the noise. Then, the pre-processed image is segmented using threshold. Motion features, texture features and color features are extracted from the segmented image in the feature extraction process. The machine learning approaches receives the input data and construct a model of the input in the training phase. A hypothesis function is provided and it is employed for predicting the future data. Hence, we can use the DCNN-DHO to train and classify the movie trailers.

4.1 Adaptive median filtering based pre-processing

Pre-processing is a main process to enhance the image so that it increases the quality for further processing. In pre-processing, initially the video frames are resized from the original size. Original frames have $360 \times 640 \times 3$ and this is resized into $256 \times 256 \times 3$. Then they are converted into grayscale images. Then the adaptive median filter is applied for removing the noise present in the video frames. For removing the impulsive noise,

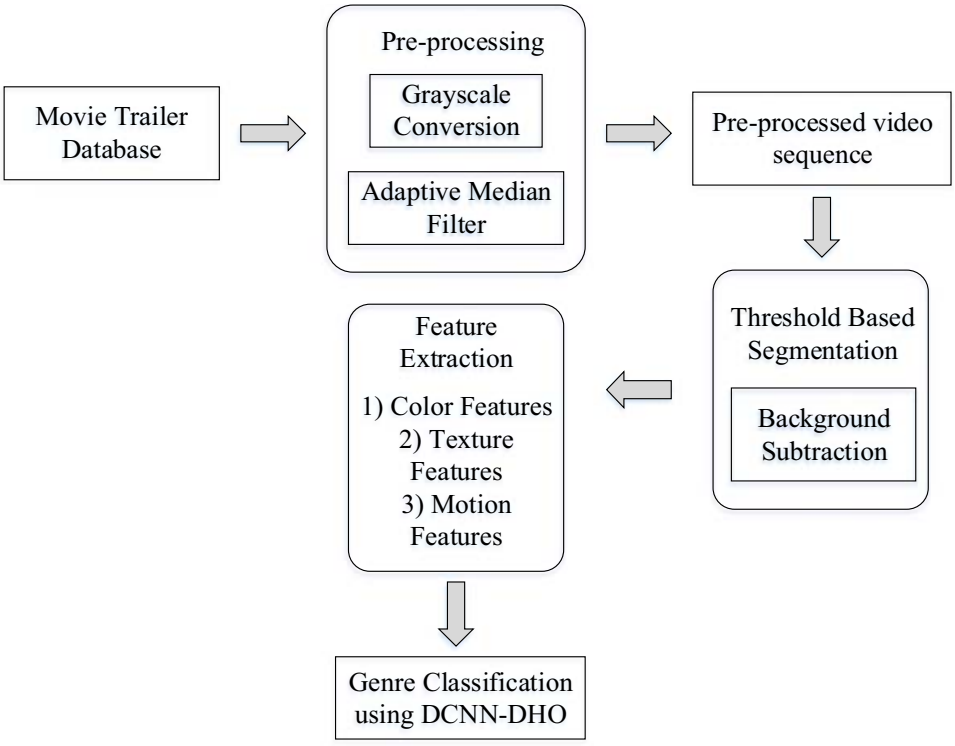


Fig. 1 Proposed Architecture

the adaptive median filtering method [21] applies filtering and noise detection algorithms. For filtering the image pixels, the window size is adaptive in nature which means if the condition doesn't meet then the size of the window is increased. By using the median of the window the pixel is filtered only if the condition is met. The output of this pre-processing is noise removed video frame.

Consider, I_{ij} represents the image pixels, I_{max} represents the maximum pixel value, I_{min} denotes the minimum pixel value, W_{max} is the maximum window size that can be extended, W denotes the current window size, and I_{med} is the allotted window's median. Thus, this filtering method can be developed in two levels as demonstrated below.

4.1.1 Level 1

a) The value of median is not an impulse, if the median is greater than minimum pixel but less than maximum pixel. So this approach proceeds to the second level for checking whether the current pixel is impulse or not.
b) Otherwise the window size is incremented and the first level is again repeated up to the median value becomes not an impulse so as to reach the second level. Or else, the window size is reached as maximum when the median is fixed as the pixel value of filtered image.

4.1.2 Level 2

a) The pixel of filtered image is unchanged if the image pixel is greater than the minimum pixel but less than the maximum pixel because the current pixel is not an impulse.
b) Otherwise, the pixel of the filtered image is considered as the median from the first level when the image pixel is equal to the minimum pixel or maximum pixel

Resized frame is shown in fig. 2(a), then the resized image is converted into grayscale image is shown in fig. 2(b) and the noise removing image is shown in fig. 2(c). Original frames have $360 \times 640 \times 3$ and this is resized into $256 \times 256 \times 3$. At last in the pre-processing, adaptive median filtering is used for denoising.

4.2 Threshold based segmentation

In the image segmentation process, the background subtraction is carried out based on the threshold value. Automatic threshold method [22] is applied to extract the foreground. A thresholding approach automatically calculates the threshold value and so it is called automatic thresholding. Maximum classes square error belongs to Otsu method, which employs a criterion for maximizing the between-class variance of pixel intensity and can detect proper threshold to segment picture automatically [27].

Based on the threshold the background and the foreground pixels are assigned to 0 and 1 respectively. Pixel values which are greater than the threshold are considered as foreground and the pixel values which are lower than or equal to the threshold are considered as the background. The threshold condition is given as follows.

$$\rho(x,y) = \begin{cases} 1 & if\ \rho(x,y) > \tau \\ 0 & otherwise \end{cases} \tag{1}$$

Where, $\rho(x,y)$ denotes the pixel value and τ represents the threshold value.

Following are the steps used for the segmentation of video frames.

- Estimate the background from the video frame.
- Automatic threshold is applied to subtract the background.
- Based on the threshold value foreground portion is extracted.

Figure 3 shows the background subtracted image. For every image, automatically selected threshold value by the system without human invention is termed an automatic threshold method. This is requirement the knowledge about the sizes of the objects, intensity features of the objects, fractions of the image employed by the objects and the number of different types of objects appearing in the image.

Fig. 2 a) Resized input frame b) Grayscale converted image c) Denoised image

Fig. 3 Background subtracted image

4.3 Feature extraction

In this work, from the video input, the motion and the visual features are considered. Texture features, color features and motion features are extracted for the classification process in which the color and the texture features are included in the visual features. Color histogram, HSV conversion, LAB conversion and YCBCR features are extracted as color features. Mean, standard deviation, skewness, wavelet transform and edge detection are extracted as the texture features and the motion features include optical flow, corner point, motion vector and shape are also extracted [23] [24] [25]. These feature values are given as input to train the DCNN classifier. The outputs of the feature extraction process are given in Figs. 4 , 5,6, 7 to Fig. 8 and 9.

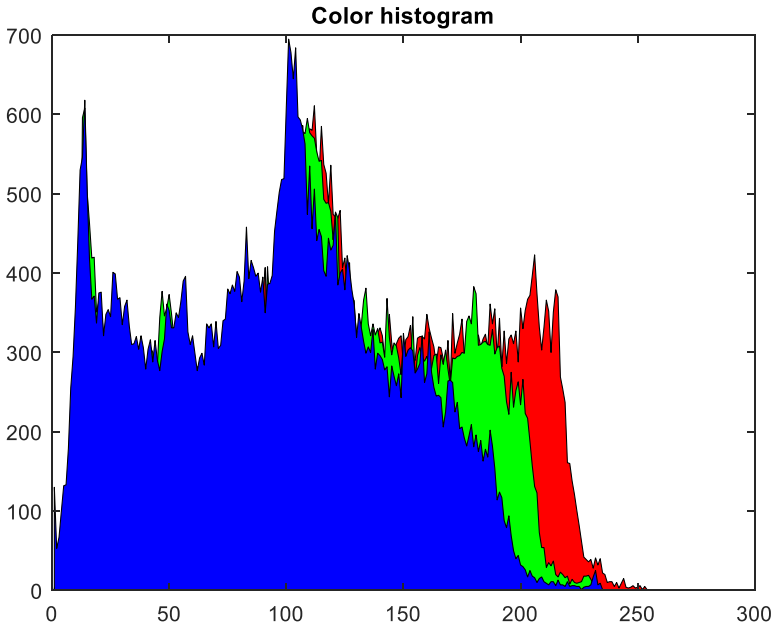


Fig. 4 Color Histogram of the input frame

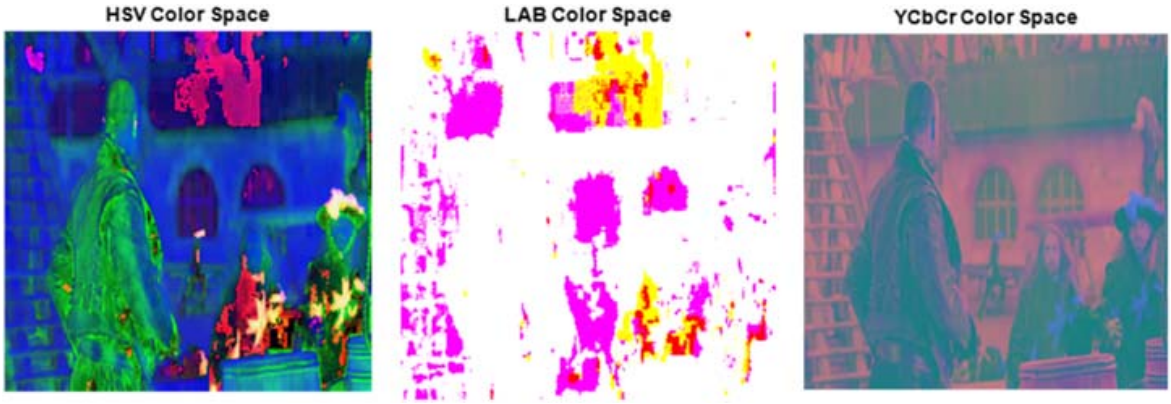


Fig. 5 Color features of the input frame

The color histogram is defined as, $\text{Color}_h(R, G, B) = n \times P(R, G, B)$ where n represents the count of pixels, R, G and B are the color channels. HSV color space conversion is performed by using Eq. (2) and YCbCr conversion is performed by using Eq. (3).

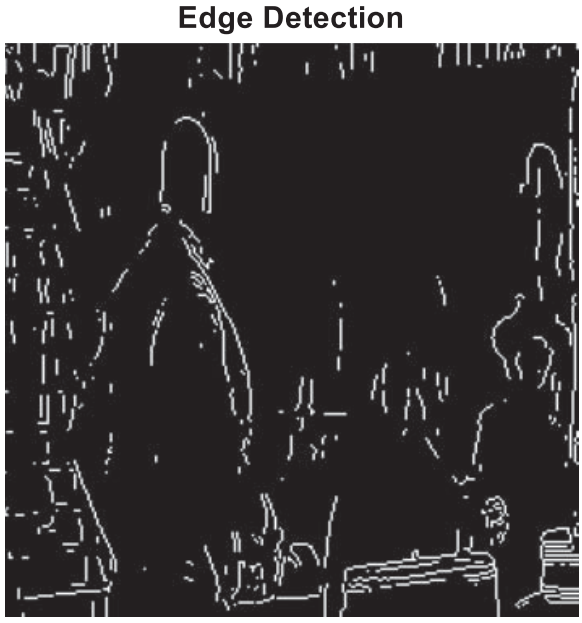
$$H = \cos^{-1}\left(\frac{\frac{1}{2}(A+B)}{\sqrt{A^2 + \text{BC}}}\right), S = 1 - \left(\frac{3}{\text{Sum of } R, G, B}\right), V = \frac{1}{3}(\text{Sum of } R, G, B) \tag{2}$$

Where, $A = R - G$, $B = R - B$ and $C = G - B$.

$$Y = 16 + \frac{65.738R + 129.057G + 25.064B}{256}, \text{Cb} = 128 - \frac{37.945R - 74.494G + 112.439}{256}, \\ \text{Cr} = 128 + \frac{112.439R - 94.154G - 18.285B}{256} \tag{3}$$

The direct conversion of RGB to LAB [27] is not possible. For this reason, initially the images are transfer from RGB to XYZ using eq. (4).

Fig. 6 Edge detection



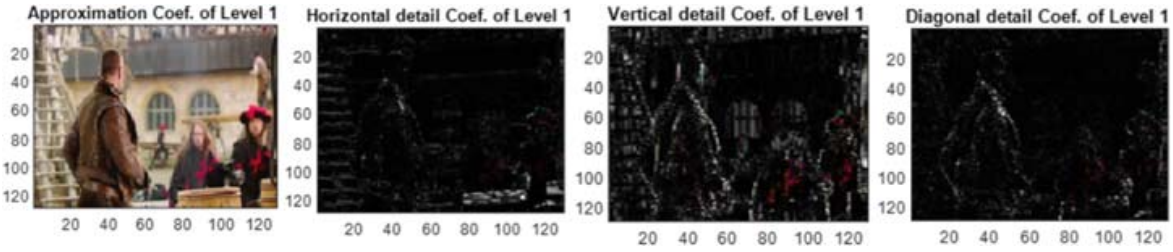


Fig. 7 Wavelet transform

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4}$$

Here, RGB defines the red, green and blue values individually. In XYZ color space, the X and Z are the color value of the image and Y is the brightness of the image. The values of L, a, and b components are calculated by eqs. (5), (6), and (7).

$$\begin{cases} L^* = 116\left(\dfrac{Y}{Y_n}\right)^{1/3} - 16 & for\ \dfrac{Y}{Y_n} > 0.008856 \\ L^* = 903.3\left(\dfrac{Y}{Y_n}\right)^{1/3} & for\ \dfrac{Y}{Y_n} \le 0.008856 \end{cases} \tag{5}$$

$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{6}$$

$$b^* = 200\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{7}$$

Here, X_n, Y_n, *and* Z_n are the constant values $X_n = 0.950450$, $Y_n = 1.000000$, $Z_n = 1.088754$ and $f(u)$are computed by eq. (8).

$$\begin{cases} f(u) = u^{1/3} & for\ u > 0.008856 \\ f(u) = 7.787u + \dfrac{16}{116} & for\ u > 0.008856 \end{cases} \tag{8}$$

In LAB color space, a is the red-green value, b is the blue-yellow value, and L is the image brightness.

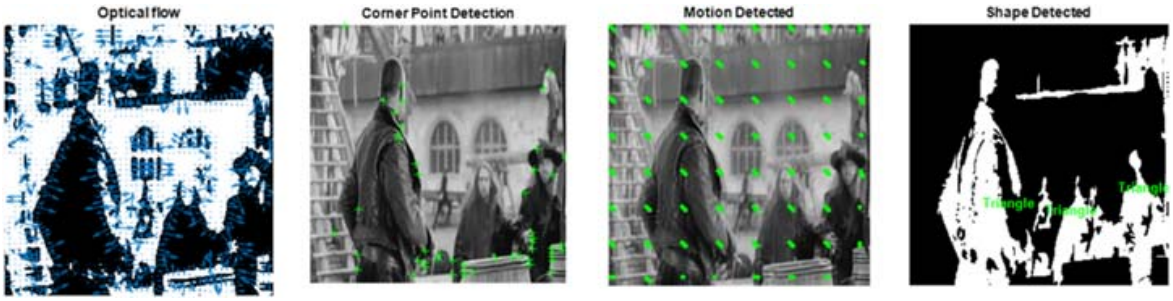


Fig. 8 Outputs of motion features

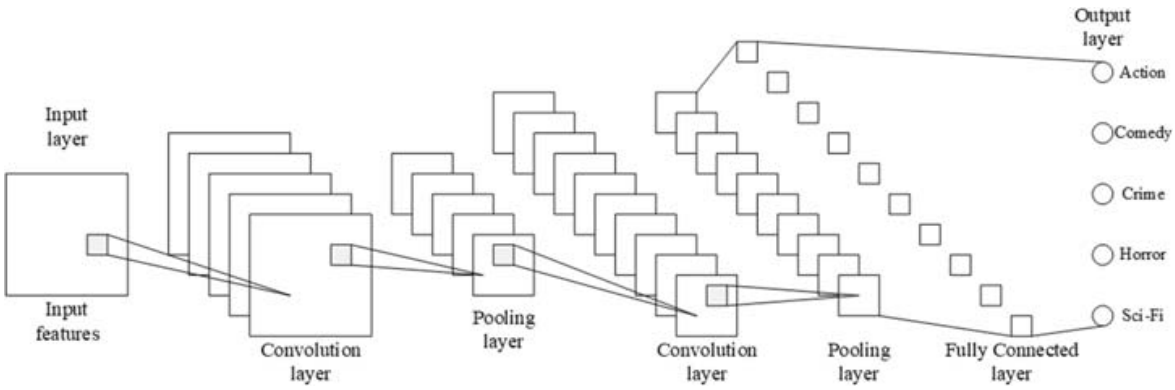


Fig. 9 Deep Convolution Neural Network Architecture

Mean of the image is evaluated by Eq. (9), standard deviation is calculated by Eq. (10) and skewness is evaluated by Eq. (11).

$$\mu = \frac{\text{Sum of pixel values}}{n} \tag{9}$$

$$\sigma = \left(\frac{1}{n}\sum_{i=1}^{n}(P_i - \mu)^2\right)^{\frac{1}{2}} \tag{10}$$

Where, n represents the count of total pixels, P_i denotes the value of i^{th} pixel.

$$\text{Skewness} = 3\ (\mu - M)/\sigma \tag{11}$$

Where, μ is the mean, M is the median and σ denotes the standard deviation.

The wavelet transform can be applied by using Eq. (12), the edge detection is estimated by using the Sobel filter and it can be performed by using Eq. (13).

$$X(i,j) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} \overline{\psi\left(\frac{t-j}{i}\right)}\ x(t)\mathrm{dt} \tag{12}$$

Where, i is the scaling, j is the shift and t defines the time.

$$G = \left(g_x^2 + g_y^2\right)^{\frac{1}{2}} \tag{13}$$

Where, G is the gradient magnitude, g_x and g_y are the gradient approximations.

The motion feature optical flow can be obtained by using Eq. (14), the corner point can be detected by using Harris corner detector and the equation is represented as Eq. (15), the motion vector is calculated by using Eq. (16) and the shape detection is performed by using Eq. (16).

$$O_f = X_t + X_x u + X_y v \tag{14}$$

Where, X_t is the time derivative of image, X_x and X_y are the space image derivatives and u, v are the point coordinates.

$$C(u,v) = \sum_{x,y}\omega(x,y)\left[X(x+u,y+v) - X\left(x,y\right)\right]^2 \tag{15}$$

Where, $\omega(x, y)$ is the square window function, $X(x+u, y+v)$ denotes the shifted intensity and $X(x, y)$ represents the intensity.

$$\overrightarrow{M} = \text{argmin}\|X(x,y,t) - X(x - v_1, y - v_2, t-1)\| \tag{16}$$

$$\alpha = W/H, \beta = \left[\left(W - W'\right)^2 + \left(H - H'\right)^2\right]^{\frac{1}{2}} \tag{17}$$

Where, W and H are the width and height of the n^{th} frame, W' and H' are the width and height of the $(n+1)^{\text{th}}$ frame respectively.

The block-based motion vectors can be estimated by using block matching, which minimizes a measure of matching error. The matching error between the block at position (x, y)in the current image, I_t, and the candidate block at position $(x+u, y+v)$in the reference image, I_{t-1}, is usually defined as the sum of absolute difference (SAD). Here, $SAD_{(x,y)}(u, v) = \sum_{j=0}^{B-1} \sum_{i=0}^{B-1} |I_t(x+i, y+i) - I_{t-1}(x+u+i, y+v+j)|$ and the block size is$B \times B$. The estimate of the motion vector, $(\widehat{u,\ v})$is defined to be the (u, v) which minimizes$SAD_{(x,y)}(u, v)$. The motion vector is calculated by using Eq. (18).

$$\left(\widehat{u}, \widehat{v}\right) = \arg \min_{(u,v) \in S} SAD_{(x,y)}(u, v) \tag{18}$$

Where $S = \{(u, v)| - R \le u, v \le R \text{ and } (x+u, y+v) \text{is a valid position in } I_{t-1}\}$ is the search set. R is an integer which defines the search range.

The shape detection is performed by using Eq. (19).

$$\text{Shape factor}(SF) = \frac{area}{(diameter)^2} \tag{19}$$

Here the *diameter* denotes the maximum distance between two pixels in any labelled area. The *area* defines the labelled area or region. The shape factor defines the different type of shape as shown in Table 1.

4.4 DCNN-DHO based training and classification of movie trailer

The objective of DCNN model to learn filter kernel by delivering a more abstract depiction of the data in every layer. In vision system, DCNN is presently the most dominant tool. It contains three layers such as, convolutional layers, pooling layers and fully connected layers. The DCNN model gets input samples at the input layer. The output layer is treated individually as a special layer. Through convolving the kernel with input feature maps, every convolutional layer provides feature maps (this feature map are taken from extracted features). Then the height and width of each feature map is 3×3. So the shape is square and the number of channel is three. Using convolutional layers, a pooling layer is constructed to provide the down sample feature maps which is frequently established by determining the local maxima in a neighbourhood. In addition, pooling provides translational invariance and it decreases the number of neurons to be processed in future layers. Every neuron has a deeper connection in fully connected layers than convolutional layers. Convolutions are used in a minimum of one

Table 1 SF values for different shape

Shape type	Shape factor (SF)
Circle	$0.7 \leq SF \leq 0.8$
Square	$0.48 \leq SF \leq 0.55$
Rectangle	$0.2 \leq SF \leq 0.3$
Triangle	$0.44 \leq SF \leq 0.483$
Oval	$0.32 \leq SF \leq 0.34$
Diamond	$0.36 \leq SF \leq 0.38$

layer of DCNNs [26] neural networks rather than regular matrix multiplications. Convolution is a mathematical operation between the two functions which provides a modified version of the functions. Convolution of pixel is defined in eq. 20:

$$v_{ij}^{xy} = \mathrm{ReLU}\left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{ijm}^{pq} v_{(i-1)m}^{(x+p)(y+q)}\right) \tag{20}$$

Where ReLU is the activation function and b is the bias. *w and v* is the weight sharing and weight dying respectively. The basic elements of the video emerge the intuition related to the application of DCNNs for the classification based on a video. Movie trailers are considered as a group of arranged frames and every frame will possibly indicate the provided genre by concerning the specific shot that it describes. Anyhow, some of the frames are essential in representing the actual movie trailer genre than the others. There may be generic or genre-free frames that are presented in any case of the movie at hand. In the perfect DCNN framework, every frame contains a single instance labelled based on the whole video genre. Every stage of a DCNN contains more than one convolutional layer with the same number of pooling layers and it is continued by more than one fully connected layer. All the convolution layers are succeeded by ReLU non-linear normalization.

In this work, the DCNN contains one input layer, one output layer, two pooling layers, two convolution layers and one fully connected layer. DHO [28] is used to represent the weights of DCNN classifier that initializes the weight and this process reduces an estimated error. The main objective of the DHO is to find the optimal position for the human to hunt the deer. DHO is inspired by the hunting behavior of humans toward deer. Even though the activities of the hunters differ, the way of attacking the buck/deer is based on the hunting strategy they develop. The hunting strategy depends on the movement of two hunters in their best positions, termed as leader and successor. Accordingly, each hunter updates his position until they reach the buck. Initialize the population size, the wind angle, the deer's position angle, the number of iterations and the end conditions. The position of the leader is updated by:

$$Y_{i+1} = Y^{lead} - p\left|\cos(v) \times Y^{lead} - Y_i\right| \tag{21}$$

Here, the position at the current iteration is denoted as Y_i, the position at the next iteration is represented as Y_{i+1} and p is a random number whose value ranges from 0 to 2. Y^{lead} is the leader location of the search agent from a present population. In the exploration phase, the same idea in encircling behavior can be adopted by adjusting the vector L. Since we assume a random search initially, the value of the vector L considered is less than 1. Therefore, the position update is based on the successor position rather than the first best solution obtained. This allows a global search as represented by the following equation,

$$Y_{i+1} = Y^{successor} - X.p.|L \times Y^{successor} - Y_i| \tag{22}$$

Where $Y^{successor}$is the successor position of the search agent from the present population. X and L are coefficient factors. The coefficient factors are measured as,

$$X = {}^{1}/_{4} log\left(i + \frac{1}{i_{max}}\right) b \tag{23}$$

$$L = 2.c \tag{24}$$

Where the maximum iteration is defined as i_{max}, b is a parameter that lies between −1 and 1, and c is an arbitrary number in the range of [0, 1]. Finally, the average value of the leader and successor positions are considered for weight updation of DCNN. The output layer contains five classes which include action, comedy, crime, horror, and sci-fi. The inputs we have given in this network are the 98×12 data. This is the data including 98 videos and corresponding extracted features of the movie trailer from the previous step. Using these twelve features the DCNN network classifies the movie trailer whether it is action or comedy or crime or horror or sci-fi. An output array is a result with 5 probabilities; one for each movie trailer genre.

5 Experimental analysis

Our proposed (DCNN-DHO) method is employed in MATLAB 2016a platform and the outcomes are studied and compared with the existing works. The performances are evaluated in terms of accuracy, false alarm rate, sensitivity, specificity, precision, F-measure and false discovery rate. All these performance metrics are evaluated for the proposed method with filtering process and without filtering process.

5.1 Dataset description

5.1.1 Dataset 1

The movie trailers are gathered from the source https://github.com/tadarsh/movie-trailers-dataset/. There are 474 YouTube trailers of American movies released in 2010–2014. Totally, in the source there are ten classes which include action, adventure, crime, comedy, horror, drama, romance, mystery, thriller and sci-fi. Some of the trailers belong to many classes. From that we have considered only five classes which comprise comedy, action, horror, crime and sci-fi. Because of using this five classes we can reduce the computational complexity.

5.1.2 Dataset 2

LMTD (Labelled Movie Trailer Dataset) is a larger movie trailer dataset which obtained from the source https://github.com/jwehrmann/lmtd. It totally contains 10,000 movie trailers from 22 genres. It includes movie trailers from 9 movie genres, namely comedy, crime, action, drama, horror, romance, adventure, thriller and sci-fi.

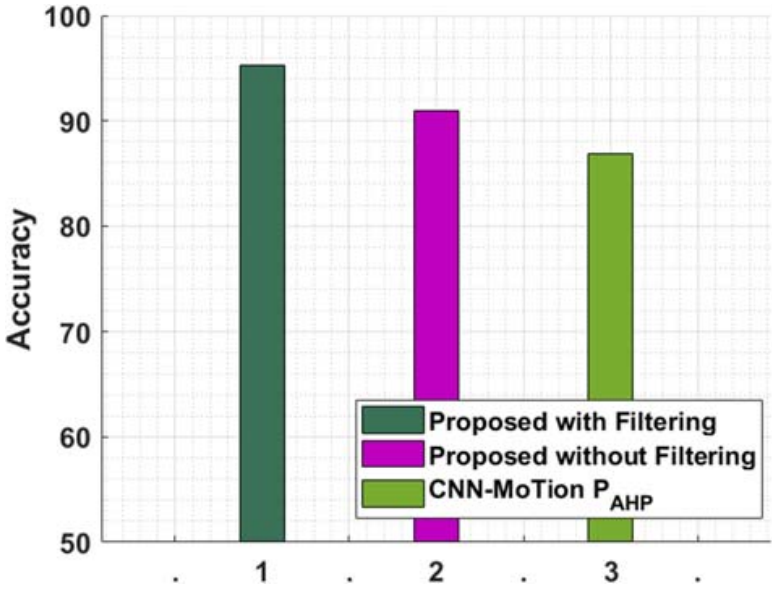


Fig. 10 Comparison of accuracy

5.2 Performance metrics

The true positive, true negative, false positive, and false negative values are calculated and using these values, the performance metrics are evaluated which include accuracy, false alarm rate, sensitivity, specificity, precision, F1 score and false discovery rate. These performances are calculated by using the following formulas.

$$\text{Accuracy} = \frac{a+b}{a+b+c+d} \tag{25}$$

$$\text{Sensitivity} = \frac{a}{a+d} \tag{26}$$

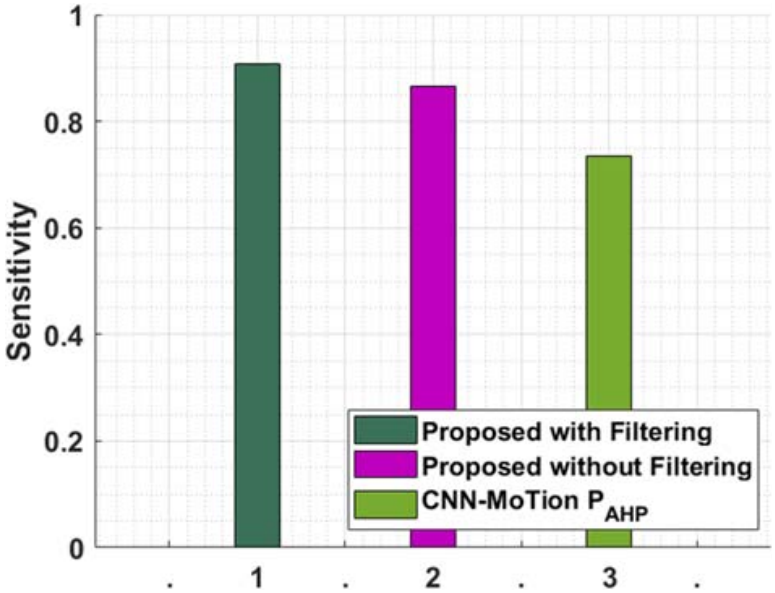


Fig. 11 Comparison of sensitivity

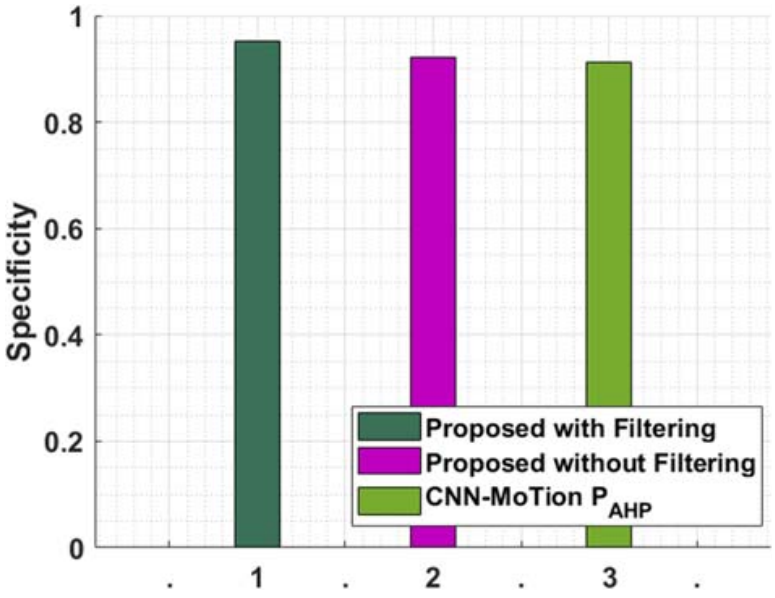


Fig. 12 Comparison of specificity

$$\text{Specificity} = \frac{b}{b+c} \tag{27}$$

$$\text{Precision} = \frac{a}{a+c} \tag{28}$$

$$F1\ \text{score} = \frac{2a}{2a+c+d} \tag{29}$$

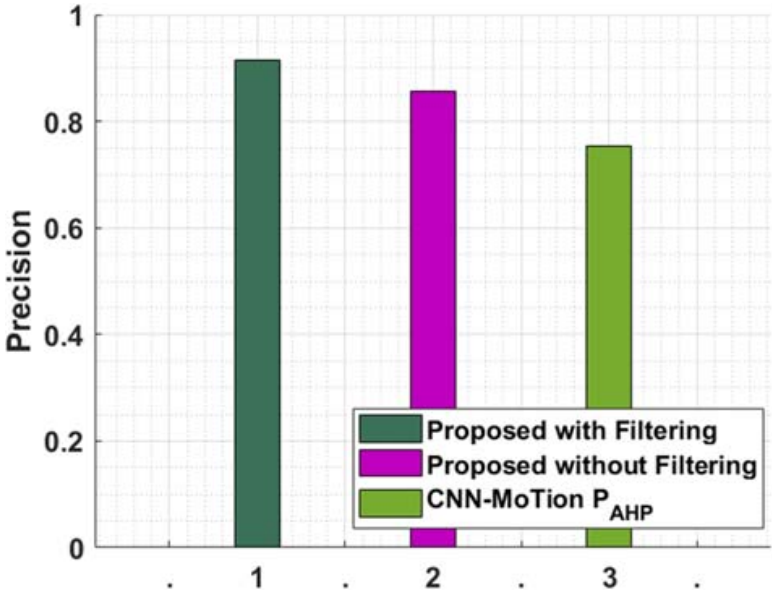


Fig. 13 Comparison of precision

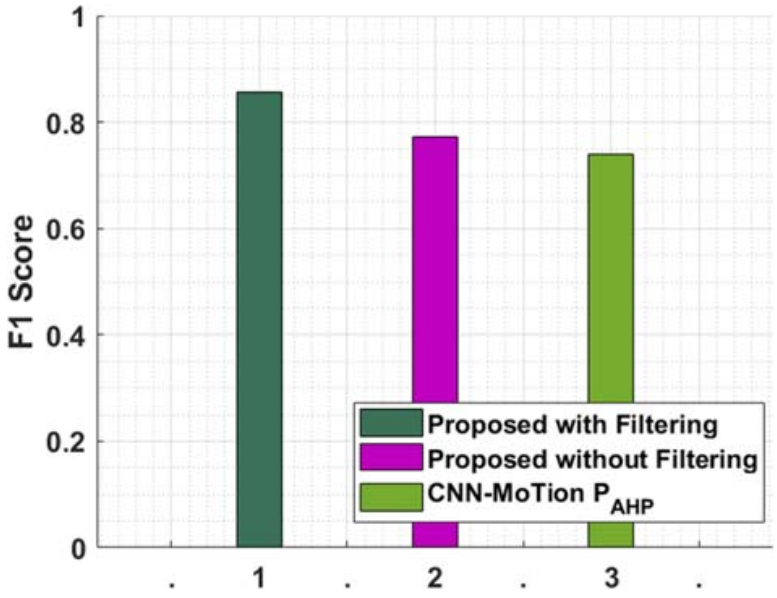


Fig. 14 Comparison of F1 score

$$\text{False Discovery Rate} = \frac{c}{a+c} \tag{30}$$

$$\text{False Alarm Rate} = \frac{c}{b+c} \tag{31}$$

Where, a denotes the true positive, b represents the true negative, c indicates the false positive and d represents the false negative.

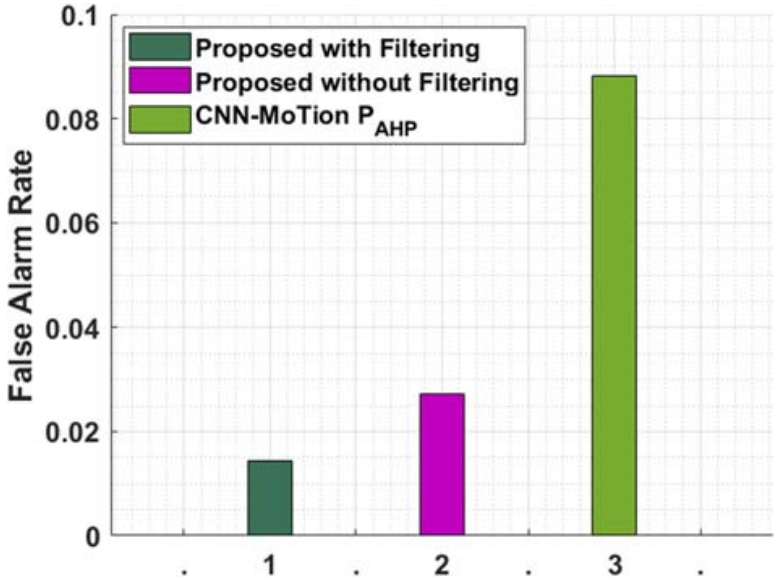


Fig. 15 False Alarm Rate

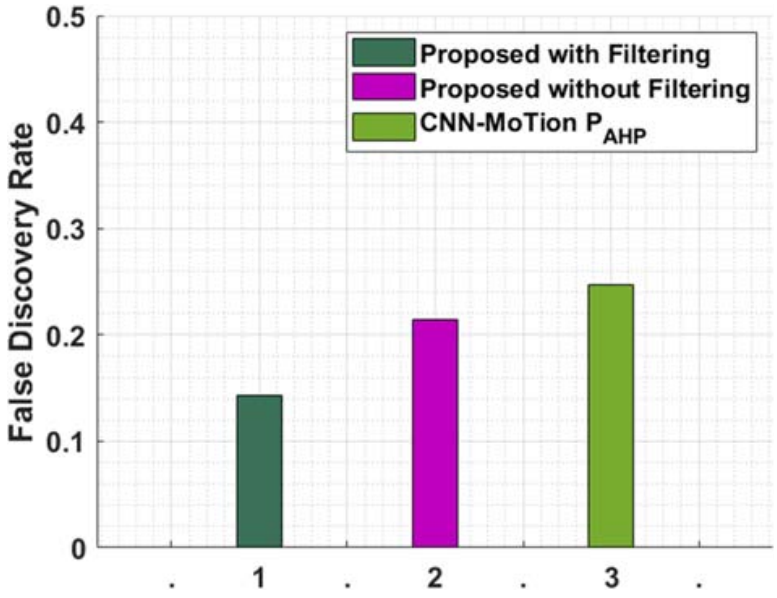


Fig. 16 Comparison of False Discovery Rate

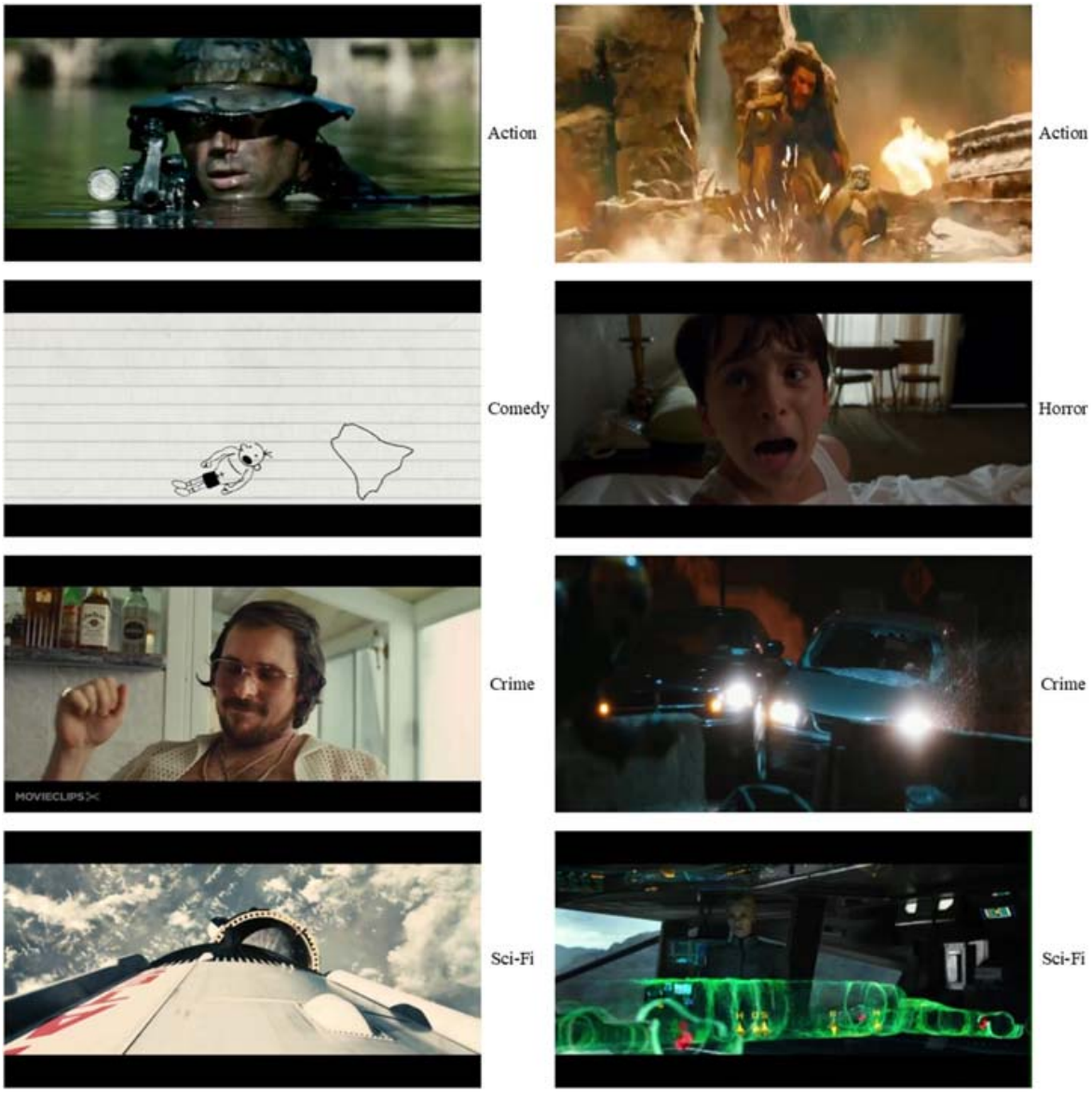


Fig. 17 Correctly classified movie trailers using the proposed

Fig. 18 Misclassified movie trailers using the proposed

5.3 Performance analysis

The results of our proposed (DCNN-DHO) method with filtering process and without filtering process are compared with the existing method CNN_MoTion P_{AHP} [19].

Figure 10 displays the comparison graph of accuracy of our proposed (DCNN-DHO) approach with filtering, proposed method without filtering and the existing method CNN_MoTion $P_{AHP.}$ The above graph represents that the proposed approach with filtering has achieved the highest accuracy when comparing with CNN_MoTion P_{AHP} and the proposed approach without filtering process. The accuracy of the proposed with filtering method is 95.23%, the proposed without filtering method has achieved the accuracy of 90.91% and the CNN_MoTion P_{AHP} method has attained the accuracy of 86.87%.

Figure 11 represents the comparison graph of sensitivity of our proposed approach with filtering, proposed method without filtering and the existing method CNN_MoTion $P_{AHP.}$ The above graph represents that the proposed method with filtering process has achieved the highest sensitivity when comparing with CNN_MoTion P_{AHP} and the proposed approach without filtering process. The sensitivity value of the proposed method with filtering is 0.90, the proposed without filtering method has achieved the sensitivity of 0.86 and the CNN_MoTion P_{AHP} method has attained the sensitivity of 0.73.

Figure 12 displays the comparison graph of specificity of our proposed approach with filtering, proposed method without filtering and the CNN_MoTion P_{AHP} method. The above graph represents that the proposed method with filtering process has achieved the highest specificity when compared with CNN_MoTion P_{AHP} and the proposed approach without

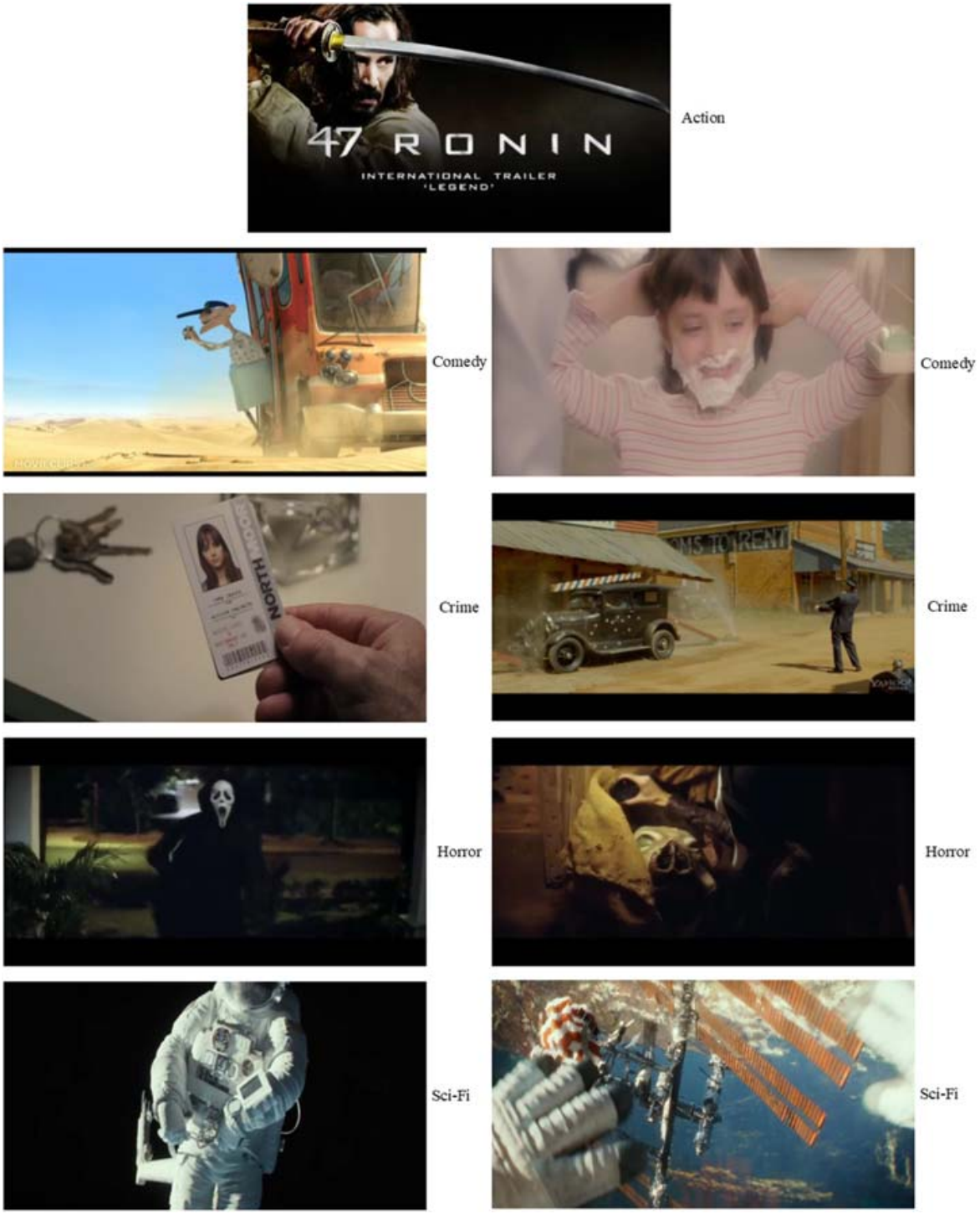


Fig. 19 Correctly classified movie trailers using the proposed system with filtering

filtering process. The specificity value of the proposed with filtering method is 0.95, the proposed without filtering method has achieved the specificity of 0.92 and the CNN_MoTion P_{AHP} method has attained the specificity of 0.91.

Figure 13 represents the comparison graph of precision of our proposed approach with filtering, proposed method without filtering and the CNN_MoTion P_{AHP} method. The above graph represents that the developed method with filtering has achieved the highest precision on comparing with CNN_MoTion P_{AHP} and the proposed approach without filtering process. The precision value of the proposed with filtering method is 0.91, the proposed without filtering

Fig. 20 Misclassified movie trailers using the

method has achieved the precision of 0.85 and the CNN_MoTion P_{AHP} method has attained the precision of 0.75.

Figure 14 indicates the comparison graph of F1 score of our proposed approach with filtering, proposed method without filtering and the CNN_MoTion P_{AHP} method. The above graph represents that the proposed method with filtering process has achieved the highest F1 score value on comparing with CNN_MoTion P_{AHP} and the proposed approach without filtering process. The F1 score value of the proposed with filtering method is 0.85, the proposed without filtering method has achieved the F1 score value as 0.77 and the CNN_MoTion P_{AHP} method has attained the F1 score value as 0.73.

Figure 15 represents the false alarm rate comparison of our proposed method with filtering, proposed method without filtering and the CNN_MoTion P_{AHP} method. The above graph represents that the proposed method with filtering process has achieved the lowest false alarm rate when compared with CNN_MoTion P_{AHP} and the proposed approach without filtering process. The false alarm rate value of the developed method with filtering is 0.014, the proposed without filtering method has achieved the false alarm rate value as 0.027 and the CNN_MoTion P_{AHP} method has attained the false alarm rate value as 0.088.

Figure 16 shows the comparison of false discovery rate of our proposed method with filtering, proposed method without filtering and the CNN_MoTion P_{AHP} method. The above graph represents that, with filtering process the developed method has achieved the lowest false discovery rate when compared with CNN_MoTion P_{AHP} and the proposed approach without filtering process. The false discovery rate value of the proposed with filtering method is 0.142, the proposed without filtering method has achieved the false discovery rate value as 0.214 and the CNN_MoTion P_{AHP} method has attained the false discovery rate as 0.246.

System without filtering.

System without filtering.

Figures 17 and 18 show the output classification of movie trailers using the proposed system without performing the filtering process. Ten movie trailers' classification outputs are represented in Fig. 17 and Fig. 18. Fig. 17 represents that out of the ten trailers eight are classified correctly. Fig. 18 represents that two trailers are classified incorrectly in which a comedy movie trailer is classified as sci-fi and a horror movie trailer also classified as a sci-Fi class.

Proposed system with filtering.

Figures 19 and 20 show the output classification of movie trailers using the proposed (DCNN-DHO) system with the filtering process. Ten movie trailers' classification output are represented in Fig. 19 and Fig. 20. Fig. 19 represents that out of the ten trailers nine are classified correctly by performing the filtering process before segmentation. Figure 18 represents that only one trailer is classified incorrectly in which an action movie trailer is classified as sci-fi class.

6 Conclusion

An approach for human action based movie trailer classification using DCNN-DHO in video sequences is proposed. In video processing normally the videos are converted into frames. In this work also the videos are converted into frames. The frames are first converted into grayscale and pre-processing is performed by employing adaptive median filter to remove the noise from the image frames. Background subtraction is performed by using the automatic threshold method to extract the foreground. Feature extraction is carried out in which the color features, the motion features and the texture features are extracted to train the classifier. For training and classification of movie trailers a DCNN-DHO is utilized. The experimental results are evaluated and compared in terms of accuracy, precision, sensitivity, specificity, F-measure, false alarm rate and false discovery rate. The experimental outcomes proved that the proposed (DCNN-DHO) method attained high accuracy of 95.23% with filtering process and 90.91% of accuracy is achieved without filtering process.

References

1. Liu A-A, Su Y-T, Nie W-Z, Kankanhalli M (2016) Hierarchical clustering multi-task learning for joint human action grouping and recognition. IEEE Trans Pattern Anal Mach Intell 39(1):102–114
2. Kumar V, Namboodiri A, Jawahar CV (2018) Semi-supervised annotation of faces in image collection, signal. Image and Video Processing 12(1):141–149
3. Ou W, Luan X, Gou J, Zhou Q, Xiao W, Xiong X, Zeng W (2018) Robust discriminative nonnegative dictionary learning for occluded face recognition. Pattern Recogn Lett 107:41.49
4. Qin Z, Shelton CR (2017) Event detection in continuous video: An inference in point process approach. IEEE Trans Image Process 26(12):5680–5691
5. Choros K (2018) Video genre classification based on length analysis of temporally aggregated video shots, computational collective intelligence, 509-518
6. Chu W-T and Guo H-J (2017) Movie genre classification based on poster images with deep neural networks. In Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, 39–45
7. Singhal A, Kumar P, Saini R, Roy PP, Dogra DP, Kim BG (2018) Summarization of videos by analysing affective state of the user through crowdsource. Cogn Syst Res 52:917–930
8. Singh J, Goyal G, Gupta S (2018) FADU-EV an automated framework for pre-release emotive analysis of theatrical trailers, multimedia tools and applications, 1-8
9. Kar A, Rai N, Sikka K and Sharma G (2017) Adascan: adaptive scan pooling in deep convolutional neural networks for human action recognition in videos." in proceedings of the IEEE conference on computer vision and pattern recognition, 3376-3385.
10. Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW (2017) Action recognition in video sequences using deep bi-directional LSTM with CNN features. IEEE Access 6:1155–1166
11. Wei P, Sun H, Zheng N (2019) Learning composite latent structures for 3D human action representation and recognition. IEEE Transactions on Multimedia 21:2195–2208
12. Zhang P, Lan C, Xing J, Zeng W, Xue J and Zheng N (2019) View adaptive neural networks for high performance skeleton-based human action recognition." IEEE transactions on pattern analysis and machine intelligence

13. Li F, Shuang F, Liu Z, Qian X (2018) A cost-constrained video quality satisfaction study on mobile devices. IEEE Transactions on Multimedia 20(5):1154–1168
14. Tian Y, Kong Y, Ruan Q, An G, Fu Y (2017) Hierarchical and spatio-temporal sparse representation for human action recognition. IEEE Trans Image Process 27(4):1748–1762
15. De Amorim MN, Saleme EB, de Assis Neto FR, Santos CA, Ghinea G (2019) Crowdsourcing authoring of sensory effects on videos. Multimed Tools Appl 78(14):19201–19227
16. Acar E, Hopfgartner F, Albayrak S (2016) A comprehensive study on mid-level representation and ensemble learning for emotional analysis of video material. Multimed Tools Appl 76(9):11809–11837
17. Ben-Ahmed O and Huet B (2018) Deep multimodal features for movie genre and interestingness prediction, 2018 international conference on content-based multimedia indexing (CBMI)
18. Ogawa T, Sasaka Y, Maeda K, Haseyama M (2018) Favourite video classification based on multimodal bidirectional LSTM. IEEE Access 6:61401–61409
19. Simoes GS, Wehrmann J, Barros RC, Ruiz DD (2016) Movie genre classification with convolutional neural networks. InNeural networks (IJCNN), 2016 international joint conference on IEEE 259-266
20. Wehrmann J, Barros R (2017) Movie genre classification: a multi-label approach based on convolutions through time. Appl Soft Comput 61:973–982
21. Shrestha S (2014) Image denoising using new adaptive based median filters. arXiv preprint arXiv:1410.2175
22. Al-Amri SS and Kalyankar NV (2010) Image segmentation by using threshold techniques. arXiv preprint arXiv:1005.4020
23. Malakar A and Mukherjee J (2013) Image clustering using color moments, histogram, edge and K-means clustering. International journal of science and research (IJSR), India online ISSN 2319-7064
24. Dandotiya Y, Atre A (2017) Image retrieval using edge detection, RLBP, color moment method for YCbCr and HSV color space. Communication and Aerospace Technology, IEEE 2:662–668
25. Yoo J, Lee G-c (2019) Moving object detection using an object motion reflection model of motion vectors. Symmetry 11(1):34
26. Zhang W, Xu L, Li Z, Lu Q, Liu Y (2016) A deep-intelligence framework for online video processing. IEEE Softw 33(2):44–51
27. Ye Z, Hu Z, Wang H, Chen H (2011 May 28) Automatic threshold selection based on artificial bee colony algorithm. In 2011 3rd international workshop on intelligent systems and applications, IEEE 1-4
28. Brammya G, Praveena S, Ninu Preetha NS, Ramya R, Rajakumar BR, Binu D (2019 May 24) Deer hunting optimization algorithm: a new nature-inspired meta-heuristic paradigm. Comput J
29. Yue-Hei NJ, Hausknecht M, Vijayanarasimhan S, Vinyals O, Monga R and Toderici G (2015) Beyond short snippets: deep networks for video classification. In proceedings of the IEEE conference on computer vision and pattern recognition, 4694-4702
30. Babaee M, Dinh DT, Rigoll G (2018) A deep convolutional neural network for video sequence background subtraction. Pattern Recogn 76:635–649
31. McLaughlin N, del Rincon JM and Miller P (2016) Recurrent convolutional network for video-based person re-identification. In proceedings of the IEEE conference on computer vision and pattern recognition, 1325-1334
32. Seo Y-S, Huh J-H (2019) Automatic emotion-based music classification for supporting intelligent IoT applications. Electronics 8(2):164
33. Liu Z, Zhang C, Tian Y (2016) 3D-based deep convolutional neural network for action recognition with depth sequences. Image Vis Comput 55:93–100
34. Wang N, Yeung D-Y (2013) Learning a deep compact image representation for visual tracking. In Advances in neural information processing systems:809–817
35. Wang P, Li W, Gao Z, Zhang J, Tang C, Ogunbona PO (2015) Action recognition from depth maps using deep convolutional neural networks. IEEE Transactions on Human-Machine Systems 46(4):498–509

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

Prashant Giridhar Shambharkar B.E. From Amravati University, Done M.Tech From RGPV, Bhopal, Working as an Assistant Professor in Department of Computer Science & Engineering having 13+ years teaching experience of various computer science and IT related subjects, worked in various committee at responsible position, member of ISO 9001-2008 member at institution level, worked as exam coordinator, counsellor, mentor, time table in-charge, assistant centre superintendent in university exam, centre controller in UPSEE exam, Alumni Incharge, Organized Literary Activities, etc. Currently pursuing Ph.D from Jamia Millia Islamia, New Delhi and area of interest includes Data mining, Real Time Systems. Guided many Graduate level projects and supervised M.Tech theses.

Mohammad Najmud Doja Dr. M.N. Doja is Professor in the Department of Computer Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia, New Delhi-110,025. He was the founder head of the Department of Computer Engineering for six years from its inception and has established the innovative and upcoming labs in the department. He started B.Tech. (Computer Engineering) and B.E. (Computer Engineering) as day and evening course respectively. He started Ph.D programme in the Department of Computer Engineering. He has more than two decades of academic, research, training and administrative experiences in the field of Computer Engineering and Information Technology.He received his B.Sc. (Engg), M.Tech. and Ph.D. degrees from B.I.T, I.I.T. Delhi and Jamia Millia Islamia, New Delhi respectively. His areas of research are Software Engineering, Networks, Security, Simulation, Operating System and Soft Computing. Prof. Doja is an active researcher. He has more than 100 publications in referred journals and conferences of international and national repute. He has been referee for a number of journals in the area of Computer Engineering and Information Technology. Several students have been awarded Ph.D under his able guidance and several students are working under his supervision. He is also author of a number of books in the area of Information Technology. He has successfully completed several projects including AICTE sponsored projects. He has chaired various sessions of various reputed conferences and delivered lectures on recent and burning topics in Computer Engineering and

Information Technology as an invited speaker at a number of places across India and world. Prof. Doja is a member of Academic Council, Board of Research Studies and Board of Studies of a number of universities including Ambedkar University Lucknow, NSIT New Delhi, A.M.U. Aligarh, Guru Gobind Singh Indraprastha University Delhi, NIT Jalandhar, Hamdard University New Delhi etc. He has been a member of a number of committee for various universities in various capacities. He has been expert member/member of various committee constituted by UGC and AICTE. He has also been associated with the affiliation work of various universities like Guru Gobind Singh Indraprastha University Delhi, M.D. University Rohtak, U.P. Technical University Lucknow and Uttrakhand Technical University.

Affiliations

Prashant Giridhar Shambharkar[1,2] • **M. N. Doja**[2]

[1] Delhi Technological University, Delhi, India

[2] Jamia Millia Islamia, New Delhi, India

arXiv:2004.13819v1 [cs.CL] 19 Apr 2020

Neural Machine Translation for Low-Resourced Indian Languages

Himanshu Choudhary, Shivansh Rao, Rajesh Rohilla
Delhi Technological University (Formerly Delhi college of Engineering)
himanshu.dce12@gmail.com, rao.shivansh570@gmail.com, rajesh@dce.ac.in

Abstract

A large number of significant assets are available online in English, which is frequently translated into native languages to ease the information sharing among local people who are not much familiar with English. However, manual translation is a very tedious, costly, and time-taking process. To this end, machine translation is an effective approach to convert text to a different language without any human involvement. Neural machine translation (NMT) is one of the most proficient translation techniques amongst all existing machine translation systems. In this paper, we have applied NMT on two of the most morphological rich Indian languages, i.e. English-Tamil and English-Malayalam. We proposed a novel NMT model using Multihead self-attention along with pre-trained Byte-Pair-Encoded (BPE) and MultiBPE embeddings to develop an efficient translation system that overcomes the OOV (Out Of Vocabulary) problem for low resourced morphological rich Indian languages which do not have much translation available online. We also collected corpus from different sources, addressed the issues with these publicly available data and refined them for further uses. We used the BLEU score for evaluating our system performance. Experimental results and survey confirmed that our proposed translator (24.34 and 9.78 BLEU score) outperforms Google translator (9.40 and 5.94 BLEU score) respectively.

Keywords: Multihead self-attention, Byte-Pair-Encodding, MultiBPE, low-resourced, Morphology, Indian Languages

1. Introduction

Many populated countries such as India and China have several languages which change region by region. for example, India has 23 constitutionally recognized official languages (*e.g.*, Hindi, Malayalam, Telugu, Tamil, and Punjabi) and numerous unofficial local languages. Not only big countries, even small countries also rich in language diversity. There are 851 languages spoken in Papua New Guinea, which is one of the smallest populated regions. In India, the population is about three billion, but only about 10% of them can speak English[1]. Some studies say that out of those 10% English speakers only 2% can talk, write, and examine English well, and rest 8% can merely recognize simple English and talk with a variety of accents. Thinking about a large number of valuable sources is available on the web in English and most people in India can not understand it well, it becomes important to translate such content into neighborhood languages to facilitate people. Sharing pieces of information between human beings is important not only for business purposes but also for sharing their emotions, reviews, and acts. For this, translation plays an essential role in minimizing the communication hole between different peoples. considering the vast amount of text, it is not viable to translate them manually. Hence, it becomes crucial to translate text from one language (say, English) to other languages (say, Tamil, Malayalam) automatically. This technique is also referred to as *machine translation*.

English to Indian language translation poses the challenge of morphological and structural divergence. For instance, (i) the number of parallel corpora and (ii) differences between languages, mainly the morphological richness and variation in word order due to syntactical divergence. Indian languages (IL) suffers from both of these problems, especially when they are being translated from English. Moreover, Indian languages such as Malayalam and Tamil differ not only in word order but are also more agglutinative as compared to English which is fusional. For instance, English has Subject-Verb-Object (SVO) whereas Tamil and Malayalam have Subject-Object-Verb (SOV). While syntactic differences contribute to difficulties of translation models, morphological differences contribute to data sparsity. We attempt to overcome both issues in this paper.

There are various papers on machine translation, but apart from foreign languages most of the works on Indian languages are limited to Hindi and on conventional machine translation techniques such as (Patel et al., 2018) and (Raju and Raju, 2016). Most of the previous work is focused on separating the words in suffix and prefix based on some rules and then applying translation techniques. We addressed this issue with BPE to make this whole process more efficient and reliable. Moreover, We observed that very less work is being done on low resourced Indian languages and techniques such as Byte-pair-encoding (BPEmb), MultiBPEmb, word-embedding, and self-attention are still unexplored which have shown a significant improvement in Natural Language Processing. Though unsupervised machine translation (Artetxe et al., 2017) is also in the focus of many researchers, still it is not as precise as supervised learning. We, also addressed that there is no trustworthy Public data available for the translation of such languages. Thus, in this paper, we have applied a neural machine translation technique with Multihead-self attention along with word embeddings and Pre-Trained Byte-Pair-Encoding. We worked on English-Tamil and English-Malayalam language pairs as it is one of the most difficult languages pair (ZdenekŽabokrtskỳ, 2012) to translate due to morphological richness of Tamil and Malayalam language. A similar approach can be applied

[1] https://www.bbc.com/news/magazine-20500312

to other languages as well. We obtained the data from En-Tamv2.0, Opus and UMC005, preprocessed them and evaluated our result using the evaluation matric BLEU. We used OpenNMT-py for the implementation of our models [2]. Experimental results, as well as the survey by native peoples, confirms that our result is far better than conventional translation techniques on Indian languages.
The Main contributions of our work are as follows:

- This is the first work to apply pre-trained BPE and MultiBPE embeddings on Indian language pairs (English-Tamil, English-Malayalam) along with Multihead self-attention technique.
- We achieved good accuracy with a relatively simpler model and in less training time rather than training on a complex neural network which requires much resources and time to train.
- We have addressed the issues with data preprocessing of Indian languages and shown why it is a crucial step in neural machine translation.
- We made our preprocessed data publicaly available, which by our knowledge contains the largest number of a parallel corpus for the languages (English-Tamil, English-Malayalam, English-Telugu, English-Bengali, English-Urdu)
- Our model outperforms Google translator with a margin of 3.36 and an 18.07 BLEU score.

The paper is organized as follows. Sections Background and Approach describe the related work and the method that we used for our translator, respectively. Section experiments and Results show data preprocessing and results and analysis of our model. Finally, Section 5. concludes the paper and future work.

2. Background

A large amount of work has been reported on machine translation (MT) in the last few decades, the first one in the 1950s (Booth, 1955). Various approaches is used by researchers, such as rule-based (Ghosh et al., 2014), corpus-based (Wong et al., 2006), and hybrid-based approach (Salunkhe et al., 2016). Each approach has its own flaws and strength. Rule-based machine translation (RBMT) is MT systems based on the linguistic information about the source and target languages which is retrieved from (multilingual, bilingual or monolingual) dictionaries and grammars covering the main syntactic, semantic and morphological regularities. It is further divided into transfer-based approach (TBA)(Shilon, 2011) and inter-lingual based approach (IBA). In the Corpus-based approach, we use a large-sized parallel corpus as raw data. This raw data contains ground truth translation for the desired languages. These corpora are used to train the model for translation. A corpus-based approach further classified in (i) statistical machine translation (SMT) (Patel et al., 2018) and (ii) example-based machine translation (EBMT) (Somers, 2003). SMT is the combination of decoding algorithms and basic statistical language models.EBMT, on the other hand, uses the translation examples and generates the new translation accordingly. It is done by finding the examples which are matching with the input. The alignment has to be performed after that to find out the parts of translation that can be reused. Hybrid-base machine translation combines any corpus-based approach and transfer approach in order to overcome their limitations. According to the recent research (Khan et al., 2017) the machine translation performance of Indian languages such as (*e.g.*, Hindi, Bengali, Tamil, Punjabi, Gujarati, and Urdu) is of an average of 10% accuracy. This demands the necessity of building better translation systems for Indian languages.
Unsupervised machine translation is further a new way of translation without using the parallel corpus, but the results are still not remarkable. On the other hand, NMT is an emerging technique and shown significant improvement in the translation results. In this paper (Hans and Milton, 2016) phrase-based hierarchical model is used and trained after morphological preprocessing. (Patel et al., 2017) trained their model after compound splitting and suffix separation. Many researchers also tried the same way and achieved a decent result on their respective datasets (Pathak and Pakray,). We observed that morphological pre-processing, compound splitting and suffix or prefix separation can be overcome by using Byte-Pair-Encoding and produce similar or even better translation results without making the model complex.

3. Approach

In this paper, we present a neural machine translation technique using Multihead self-attention and word-embedding along with pre-trained Byte-Pair-Encoding (BPE) on our preprocessed dataset of Indian languages. We developed an efficient translation system, that overcomes the OOV (Out Of Vocabulary) and morphological analysis problem for Indian languages which do not have many translations available on the web. first, we provide an overview of NMT, Multi-head self-attention, word embedding, and Byte Pair Encoding. Next, we describe the framework of our translation model.

3.1. Neural Machine Translation Overview

Neural Machine translation is a powerful algorithm based on neural networks and uses the conditional probability of translated sentences to predict the target sentences of given source language (Revanuru et al., 2017a). When coupled with the power of attention mechanisms, this architecture can achieve impressive results with different variations. The following sub-sections provide an overview of basic sequence to sequence architecture, self-attention and other techniques that are used in our proposed translator.

3.1.1. Sequence to sequence architecture

Sequence to sequence architecture is used for response generation whereas in Machine Translation systems it is used to find the relations between two language pairs. It consists of two important parts, an encoder, and a decoder. The encoder takes the input from the source language and the

[2]http://opennmt.net/OpenNMT-py/

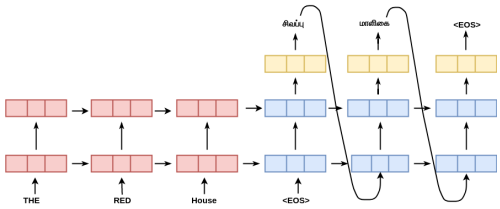


Figure 1: Seq2Seq architecture for English-Tamil

decoder leads to the output based on hidden layers and previously generated vectors. Let A be the source and B be a target sentence. The encoding part converts the source sentence $a_1, a_2, a_3..., a_n$ into the vector of fixed dimensions and the decoder part gives the word by word output using conditional probability. Here, $A_1, A_2, ..., A_M$ in the equation are the fixed size encoding vectors. Using chain rule, the Eq. 1 is transformed to the Eq. 2.

$$P(B/A) = P(B|A_1, A_2, A_3, ..., A_M) \tag{1}$$

$$P(B|A) = P(b_i|b_0, b_1, b_2, ..., b_{i-1}; \\ a_1, a_2, a_3, ..., a_m) \tag{2}$$

The decoder generates output using previously predicted word vectors and source sentence vectors in Eq. 1.

3.1.2. Attention Model

In a basic encoder-decoder architecture, encoder memorizes the whole sentence in terms of vector, and store it in the final activation layer, then the decoder uses that vector to generates the target sentence. This architecture works quite well for small sentences, but for larger sentences, maybe longer than 30 or 40 words, the performance degrades. To overcome this problem attention mechanisms play an important role. The basic idea behind this is that each time, when the model predicts an output word, it only uses the parts of input where the most relevant information is concentrated instead of the whole sentence. In other words, it only pays attention to some weighted words. Many types of attention mechanisms are used in order to improvise the translation accuracy, but the multi-head self-attention overcomes most of the problems.

Self-attention In self-attention architecture (Vaswani et al., 2017) at every time step of an RNN, a weighted average of all the previous states will be used as an extra input to the function that computes the next state. With the self-attentive mechanism, the network can decide to attend to a state produced many time steps earlier. This means that the latest state does not need to store all the information. The mechanism also makes it easier for the gradient to flow more easily to all previous states, which can help against the vanishing gradient problem.

Multi-Head Attention When we have multiple queries q, we can combine them in a matrix Q. If we compute alignment using dot-product attention, the set of equations that are used to calculate context vectors can be reduced as shown in figure 3. Q, K, and V are mapped into lower-dimensional vector spaces using weight matrices and the results are used to compute attention (which we call a Head).

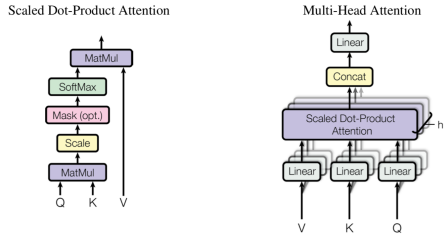


Figure 2: Attention model

In Muti-Head Attention we have h such sets of weight matrices which give us h Heads.

$$A(q, K, V) = \sum_i \frac{e^{q \cdot k_i}}{\sum_j e^{q \cdot k_j}} v_i$$

$$A(Q, K, V) = softmax(QK^T)V$$

Figure 3: Multi-Head Attention

3.1.3. Word Embedding

Word embedding is a unique way of representing the word in a vector space such that we can capture the semantic similarity of each word. Each word is represented in hundreds of dimensions. Generally, pre-trained embeddings are used trained on the larger data sets, and with the help of transfer learning, we convert the words from vocabulary to vector. (Cho et al., 2014).

3.1.4. Byte Pair Encoding

BPE (Gage, 1994) is a data compression technique that replaces the most frequent pair of bytes in a sequence. We use this algorithm for word segmentation, and by merging frequent pairs of charters or character sequences we can get the vocabulary of desired size (Sennrich et al., 2015). BPE helps in the suffix, prefix separation, and compound splitting which in our case used for creating new and complex words of Malayalam and Tamil language by interpreting them as sub-words units. We used BPE along with pre-trained fast-text word embeddings [3] (Heinzerling and Strube, 2018) for both the languages with the variation in the vocabulary size. In our model, we got the best results with vocabulary size 25000 and dimension 300.

MultiBPEmb MultiBPEmb is a collection of multiple languages subword segmentation models and pre-trained subword embeddings trained on Wikipedia data similar to monolingual BPE. On the contrary, instead of training one segmentation model for each language, here we train a single model and a single embedding for all the languages. We can also create a vocabulary of only two languages, source, and target. It deals with the mixed language sentences (Native language along with English) which are being popular nowadays on social media. Since our sentences were

[3] https://github.com/bheinzerling/bpemb

ID	Language	Train	Test	Dev
1	Tamil	183451	2000	1000
2	Malayalam	548000	3660	3000
3	Telugu	75000	3897	3000
4	Bengali	658000	3255	3500
5	Urdu	36000	2454	2000

Table 1: Dataset for Indian Languages

clean it almost produced similar results, with variation in the BLEU score by 0.60 in Tamil and 1.15 in Malayalam.

4. Experimentation and Results

4.1. Evaluation Metric

BLEU score is a method to measure the difference between machine translation and human translation (Papineni et al., 2002). The approach works by matching n-grams in result translation to n-grams in the reference text, where unigram is a unique token, bigram is a word pair and so on. A perfect match results in a score of 1.0 or 100%.

4.2. Dataset

We obtained the data from different resources such as EnTamV2.0 (Ramasamy et al., 2012), Opus (Tiedemann, 2012) and UMC005(Jawaid and Zeman, 2011) .The sentences are of domain news, cinema, bible and movie subtitles. We combined and preprocessed the data of Tamil, Malayalam, Telugu, Bengali, and Urdu. After preprocessing (as described below) and cleaning, the dataset is split into train, test, and validation. Our final dataset is described in table 1. In our knowledge this is the largest, clean and preprocessed public dataset [4] available on the web for general purpose uses. As there is no publicly available dataset to compare various approaches on Indian languages, our datasets can be used to set baseline results to compare with.

4.3. Data Pre-processing

In the Research works (Hans and Milton, 2016) (Ramesh and Sankaranarayanan, 2018) EnTamV2.0 dataset is used. Also, the Opus dataset is a much widely used parallel corpus resource in various researcher's works. However, we observed that in both of these well-known parallel resources there are many repeated sentences, which may results into the wrong results (can be higher or lower) after dividing into train, validation, and test sets, as many of the sentences, occur both in train and test sets. In most of the work, the focus relies on the models without interpreting the data which performs much better on our own test set rather than on general translated sentences. Thus, it is essential to analyses, correct and cleans the data before using it for the experiments. Researchers should also provide a detailed source of the corpus otherwise results can be misleading such as in paper (Revanuru et al., 2017b). We observed the following four important issues in the online available corpus.

- Sentence repetition with the same source and target.
- Different translations by the same source.
- Same translated sentences by different source sentences.
- Indian language tokenization.

To overcome the first issue, we took unique pairs from all the parallel sentences and removed the repeating ones. To tackle the second and third case we removed sentence pairs which were repeated more than twice and the difference between their length are in the window of 5 words. It is because for both of these cases we cannot identify that which source is correct for the same translation and which translated sentence is comes from the same source. We observed that there were some sentences, which were repeating even more than 20 times in the Opus dataset. This confuses the model to learn, identify and capture different features and overfits the model. Though data-augmentation (Fadaee et al., 2017) can improve the translation results, but in that case, the original data should be pre-processed, otherwise many augmented sentences may appear in both train and test data which leads to higher but wrong BLEU score as it will not work efficiently on new sentences.

For the tokenization of the English language, there are many libraries and frameworks such as (*e.g.*, Perl tokenizer) but these do not work well on the Indian languages, due to the difference between morphological symbols. The word-formation of Indian languages is quite different which we believed can only be handled by either special library for that particular language or by Byte-Pair-Encoding. In the case of BPE, we don't need to tokenize the words which generally leads to better translation results.

After working on all these minor, but effective pre-processing we got our final dataset. While extracting the data from the web, we also removed sentences with a length greater than 50, known translated words in target sentences, noisy translations, and unwanted punctuations. For the reliability of data, we also took the help of native speakers of these languages.

4.4. Translator

We tried various new techniques as described above to get a better intuition of the effects on these two Indian language pairs. Our first model consists of 4 layer Bi-directional LSTM encoder and a decoder with 500 dimensions each along with a vocabulary size of 50,004 words for both source and target. First, we used Bahdanau's attention and Adam optimizer with the dropout (regularization) of 0.3 and the learning rate 0.001. Here we used the 300 dimensional Pre-trained fast text [5] word embeddings for both the languages. Secondly, we used Pre-trained fast text Byte-Pair-Encoding [6] with the same attention. In the third model, we changed the attention to multi-head with 8 heads and 6 encoding and decoding layers. It shows an improvement of 1.2 and 6.18 BLEU scores for Tamil and Malayalam respectively. For the final model we used Multilingual fast text pre-trained Byte-pair-Encoddings [7] and got our final

[4] https://github.com/himanshudce/Indian-Language-Dataset

[5] https://fasttext.cc/docs/en/crawl-vectors.html

[6] https://github.com/bheinzerling/bpemb

[7] https://nlp.h-its.org/bpemb/multi/

best results of 9.67 and 25.36 respectively as shown in table 2 and table 3.

4.5. Result

Our results is shown in table 2 and table 3. For Google translate we used Python API to translate the English sentences and compared the results with our various models. From the test results, It is observed that our model overcomes the OOV (Out of Vocabulary) problem in some cases, and is handy enough to be used in day to day life and official work.

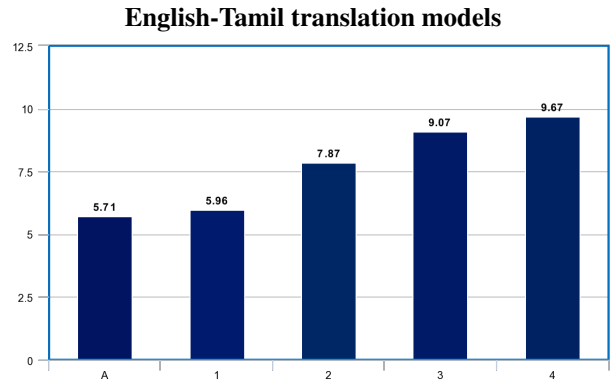


Figure 4: English-Tamil model comparison with Google Translator Table2

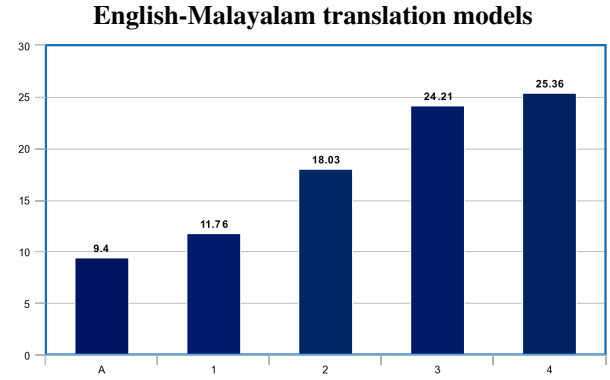


Figure 5: English-Malayalam model comparison with Google Translator Table 2

4.6. Analysis

We conducted a survey with ten random sentences from our test data and accumulated the reviews of native Tamil speaking peoples. On comparing the reviews of Google translator and our translator, it was found, that our translation results were better in 60% cases than the Google translator. The visualization of an Attention can be seen in 6 of one of the sample sentences from our test data.

5. Conclusion

In this paper, we applied Neural Machine Translation (NMT) on two of the most difficult Indian language pairs (English-Tamil, English-Malayalam). We addressed the issues of data pre-processing and tokenization. To handle morphology and word complexities of Indian languages we applied pre-trained fast text BPEmb, MultiBPEmb embeddings along with multi-head self-attention which outperformed Google translator with a margin of 3.96 and 15.96 BLEU points respectively. The same approach can be ap-

ID	Model	BLEU Score
A	Google Translator	5.71
1	Bi-LSTM(4-Layers)+ A+ Bahdanau Attention + WE	5.96
2	Bi-LSTM(4-Layers)+ A + Bahdanau Attention + Pre-BPE(25000)	7.87
3	Bi-LSTM(6-Layer)+ A + Multi-Head Attention+ Pre-BPE(25000)	9.07
4	Bi-LSTM(6-Layer)+ A + Multi-Head self Attention + Pre-MultiBPE(100000)	9.67

Table 2: English-Tamil model comparison with Google Translator (A=Adam, WE=Word Embeddings)

ID	Model	BLEU Score
A	Google Translator	9.40
1	Bi-LSTM(4-Layers)+ A+ Bahdanau Attention + WE	11.76
2	Bi-LSTM(4-Layers)+ A + Bahdanau Attention + Pre-BPE(25000)	18.03
3	Bi-LSTM(6-Layer)+ A + Multi-Head Attention+ Pre-BPE(25000)	24.21
4	Bi-LSTM(6-Layer)+ A + Multi-Head self Attention + Pre-MultiBPE(100000)	25.36

Table 3: BLEU Score of English-Malayalam translated system. (A=Adam, B= Bahdanau, WE=Word Embedding)

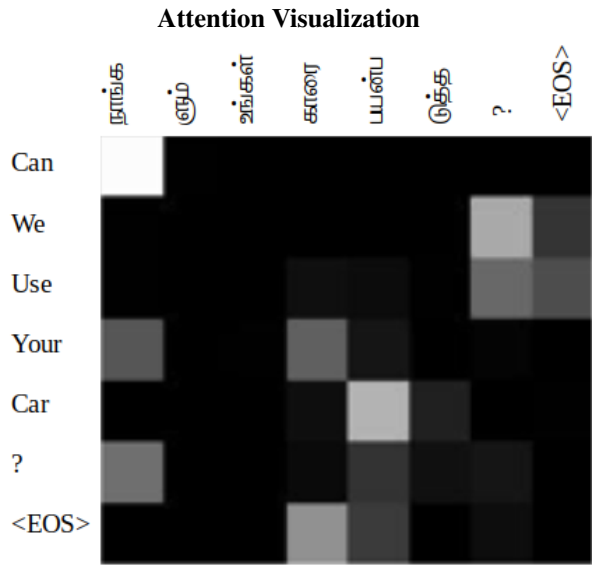


Figure 6: Attention visualization of English-Tamil sentence pair from our test data

plied to other Indian languages as well. Since the accuracy of our model was fairly good, it can be used for creating English-Malayalam and English-Tamil translation applications that will be very useful in domains like tourism, education and corporate. In the future, we can also explore the possibility to improve the translation results for code-switched languages using MultiBPE and other variations.

6. Bibliographical References

Artetxe, M., Labaka, G., Agirre, E., and Cho, K. (2017). Unsupervised neural machine translation. *CoRR*, abs/1710.11041.

Booth, A. D. (1955). Machine translation of languages, fourteen essays.

Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Fadaee, M., Bisazza, A., and Monz, C. (2017). Data augmentation for low-resource neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 567–573, Vancouver, Canada, July. Association for Computational Linguistics.

Gage, P. (1994). A new algorithm for data compression. *The C Users Journal*, 12(2):23–38.

Ghosh, S., Thamke, S., et al. (2014). Translation of telugu-marathi and vice-versa using rule based machine translation. *arXiv preprint arXiv:1406.3969*.

Hans, K. and Milton, R. (2016). Improving the performance of neural machine translation involving morphologically rich languages. *arXiv preprint arXiv:1612.02482*.

Heinzerling, B. and Strube, M. (2018). BPEmb: Tokenization-free Pre-trained Subword Embeddings in 275 Languages. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).

Jawaid, B. and Zeman, D. (2011). Word-order issues in English-to-Urdu statistical machine translation.

Khan, N. J., Anwar, W., and Durrani, N. (2017). Machine translation approaches and survey for indian languages. *arXiv preprint arXiv:1701.04290*.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Patel, R. N., Pimpale, P. B., et al. (2017). Mtil17: English to indian langauge statistical machine translation. *arXiv preprint arXiv:1708.07950*.

Patel, R., Pimpale, P., and Mukundan, S. (2018). Machine translation in indian languages: Challenges and resolution. *Journal of Intelligent Systems*, 09.

Pathak, A. and Pakray, P.). Neural machine translation for indian languages. *Journal of Intelligent Systems*.

Raju, B. N. V. N. and Raju, M. S. V. S. B. (2016). Statistical machine translation system for indian languages. In *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, pages 174–177, Feb.

Ramasamy, L., Bojar, O., and Žabokrtský, Z. (2012). Morphological processing for english-tamil statistical machine translation. In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages (MTPIL-2012)*, pages 113–122.

Ramesh, S. H. and Sankaranarayanan, K. P. (2018). Neural machine translation for low resource languages using bilingual lexicon induced from comparable corpora. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 112–119.

Revanuru, K., Turlapaty, K., and Rao, S. (2017a). Neural machine translation of indian languages. In *Proceedings of the 10th Annual ACM India Compute Conference on ZZZ*, pages 11–20. ACM.

Revanuru, K., Turlapaty, K., and Rao, S. (2017b). Neural machine translation of indian languages. 11.

Salunkhe, P., Kadam, A. D., Joshi, S., Patil, S., Thakore, D., and Jadhav, S. (2016). Hybrid machine translation for english to marathi: A research evaluation in machine translation:(hybrid translator). In *Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on*, pages 924–931. IEEE.

Sennrich, R., Haddow, B., and Birch, A. (2015). Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.

Shilon, R. (2011). *Transfer-based Machine Translation between morphologically-rich and resource-poor languages: The case of Hebrew and Arabic*. Ph.D. thesis, Citeseer.

Somers, H. (2003). An overview of ebmt. In *Recent advances in example-bas ed machine translation*, pages 3–57. Springer.

Tiedemann, J. (2012). Parallel data, tools and interfaces in opus. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In I. Guyon, et al., editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

Wong, F., Dong, M., and Hu, D. (2006). Machine translation using constraint-based synchronous grammar. *Tsinghua Science and Technology*, 11(3):295–306.

ZdenekŽabokrtskỳ, L. O. (2012). Morphological processing for english-tamil statistical machine translation. In *24th International Conference on Computational Linguistics*, page 113.

arXiv:2004.12286v1 [physics.soc-ph] 26 Apr 2020

Nutritional Profile Estimation in Cooking Recipes

1st Jushaan Kalra
Dept. of Computer Engineering
Delhi Technological University
New Delhi, India
jushaan18@gmail.com

1st Devansh Batra
Dept. of Information Technology
NSUT
New Delhi, India
devanshb.it.17@nsit.net.in

2nd Nirav Diwan
Dept. of Computer Science and Engineering
IIIT-Delhi
New Delhi, India
nirav17072@iiitd.ac.in

Ganesh Bagler
Center for Computational Biology
IIIT-Delhi
New Delhi, India
bagler@iiitd.ac.in

***Abstract*—The availability of an accurate nutrition profile of recipes is an important feature for food databases with several applications including nutritional assistance, recommendation systems, and dietary analytics. Often in online databases, recipes are obtained from diverse sources in an attempt to maximize the number of recipes and variety of the dataset. This leads to an incomplete and often unreliable set of nutritional details. We propose a scalable method for nutritional profile estimation of recipes from their ingredients section using a standard reliable database for the nutritional values. Previous studies have testified the efficiency of string-matching methods on small datasets. To demonstrate the effectiveness of our procedure, we apply the proposed method on a large dataset, RecipeDB [1], which contains recipes from multiple data sources, using the United States Department of Agriculture Standard Reference (USDA-SR) Database as a reference for computing nutritional profiles. We evaluate our method by calculating the average error across our database of recipes (36 calories per serving) which is well within the range of errors attributable to physical variations.**

***Index Terms*—Nutrition, Named Entity Recognition, Recipe Dataset, Nutrition Composition Tables, USDA, SR**

I. Introduction

Estimating the nutritional profile of a cooking recipe is a challenging problem. While there is no dearth of web-based services that provide recipes, their cooking instructions along with ingredient details, pertaining to a wide range of cuisines across the world, their nutritional profiles are not easily available. Here, we propose a Named Entity Recognition(NER)-based strategy for extracting different elements of recipes and to compute the nutritional profile of a recipe by mapping them to their USDA nutritional description.

Several methods for the calculation of nutritional values of a cooked meal have been proposed. The most accurate method [1] for this calculation employs chemical analysis. Since this method is applied on the cooked meal, it does not lead to any untoward errors. However, this analysis is not feasible for large datasets of recipes from online resources, since user-uploaded recipes tend to be extremely noisy and without a standard format for storing data. Furthermore, it is not practical to conduct chemical analysis on every recipe, since they may number in hundreds of thousands. Through the course of our research, we collected more than 100,000 recipes from one source alone and hence we sought for more scalable methods.

An alternative approach is mentioned in [2] where food images are used to calculate calorie contents. Such methods do not provide accurate results suitable for academic research. Since these methods also look for the presence of particular ingredients within food images which are themselves available more accurately in the recipe text, we focus on methods that use the text content itself.

The approach we adopted is aligned with the one mentioned in [3] which assumes that the sum total of nutrition of ingredients in a particular recipe can be approximated for the nutritional profile of the recipe. This simplifies our problem statement since we can now calculate the nutritional value of ingredients from nutritional composition tables, and their sum would give us our required nutritional values.

It has been observed [4] that more accurate results would be obtained if nutritional yield due to cooking is taken into account, but, there is no such consolidated resource for yield values as they differ with ingredient, cooking time and other variable features. Without the knowledge of these variables, it is difficult to estimate the nutritional profile of the recipe with the above method.

Hence the task is to form a complete nutritional composition table for all ingredients in our recipes since the nutritional value of a recipe is the sum total of the nutritional value of its constituent ingredients. To calculate this, we propose a three-step approach– Ingredient Data Mining (Section IIA), Ingredient Name Matching (Section IIB) and Unit Matching (Section IIC) which together give us the required nutritional profile. Refer to Figure 1 for an overview of the system architecture used.

TABLE I. INGREDIENT TAGS EXTRACTION

Ingredient Phrase	Name	State	Quantity	Unit	Temperature	Dry/Fresh	Size
1/2 lb lean ground beef	beef	ground lean	1/2	lb			
1 small onion , finely chopped	onion	chopped	1				small
1 hard-cooked egg , finely chopped	egg	hard-cooked chopped	1				
1 tablespoon fresh dill weed	dill weed		1	tablespoon		fresh	
1/2 teaspoon salt ,freshly ground	salt		1/2	teaspoon			
1/8 teaspoon black pepper,minced	black pepper		1/8	teaspoon			
3/4 cup butter or 3/4 cup margarine , softened	butter	softened	3/4	cup			
2 cups all-purpose flour	purpose flour		2	cups			
1 teaspoon salt	salt		1	teaspoon			
1/2 cup low-fat sour cream	cream	sour low fat	1/2	cup			
1 egg yolk	egg yolk		1				
1 tablespoon cold water	cold water		1	tablespoon	cold		

TABLE II. EXAMPLES OF FOOD DESCRIPTION IN USDA-SR DATABASE

S.No	Description
1	Butter, salted
2	Butter, whipped, with salt
3	Butter, without salt
4	Cheese, blue
5	Cheese, cottage, creamed, large or small curd
6	Cheese, mozzarella, whole milk
7	Milk, reduced fat, fluid, 2% milkfat, with added vitamin A and vitamin D
8	Milk, reduced fat, fluid, 2% milkfat, with added nonfat milk solids and vitamin A and vitamin D
9	Milk, reduced fat, fluid, 2% milkfat, protein fortified, with added vitamin A and vitamin D
10	Milk, indian buffalo, fluid
11	Milk shakes, thick chocolate
12	Milk shakes, thick vanilla
13	Yogurt, plain, whole milk, 8 grams protein per 8 ounce
14	Yogurt, vanilla, low fat, 11 grams protein per 8 ounce
15	Egg, whole, raw, fresh
16	Egg, white, raw, fresh
17	Egg, yolk, raw, fresh
18	Apples, raw, with skin
19	Apples, raw, without skin

II. CALCULATING NUTRITIONAL VALUE OF RECIPES

A. *Ingredient Data Mining*

We utilize the data available from RecipeDB [1] [1] which contains 118,071 recipes from AllRecipes [2] and FOOD.com [3]. In order to estimate the nutritional profile of a recipe, we need to obtain all the ingredients used in a recipe and their corresponding quantities, units and/or size and other useful information such as processing state (ground, thawed, etc.), temperature and dryness.

Consider the recipe - Piroszhki (Little Russian Pastries) [4]. The Table I shows the outputs of our Named Entity Recognition approach on twelve ingredient phrases. We note that for example, in the table, “1 small onion, finely chopped” contains the entire information that we require to calculate the ingredient’s nutritional value, we only need the data in a structured format in order to estimate the nutritional value of the recipe.

We propose a Named Entity Recognition System to train the model to infer the following tags– NAME, STATE, UNIT, QUANTITY, TEMP, DRY/FRESH, SIZE. We manually tagged a corpus of 6612 ingredient phrases and tested the model on a test set of size 2188 ingredient phrases. In order to include ingredient phrases of large diversity in our training and testing set, we utilized Parts of Speech Tagging to form vectors representing each ingredient phrases. A vector representing an ingredient phrase would be defined by the frequency of the tag in the ingredient phrase. We then proceeded to cluster the obtained vectors. The ingredient phrases were chosen for the training and testing set by selecting a subset of ingredient phrases from each cluster. We trained our model using Stanford Named Entity Recognition Model [5]. The

[1] https://cosylab.iiitd.edu.in/recipedb
[2] https://www.allrecipes.com
[3] https://www.food.com
[4] https://www.food.com/recipe/piroszhki-little-russian-pastries-486254

TABLE III. COMPARISON OF INFERENCES FROM MODIFIED JACCARD INDEX WITH VANILLA JACARD INDEX

Ingredient Phrase	Ingredient Name	Food Desc. Inferred W/ Modified JI	Food Desc. Inferred W/ Vanilla JI
1 cup red lentil	red lentils	Lentils, pink or red, raw	Cherries, sour, red, raw
1 roma tomato , quartered	roma tomato	Soup, tomato beef with noodle, canned, condensed	Soup, tomato, canned condensed
1/4 teaspoon ground coriander	coriander	Coriander (cilantro) leaves, raw	Spices, coriander leaf, dried
2 tablespoons tomato paste	tomato paste	Tomato products, canned, paste, without salt added	Soup, tomato, canned, condensed
1 1/4 cups vegetable broth	vegetable broth	Soup, vegetable with beef broth, canned, condensed	Soup, vegetable broth, ready to serve
1 can fava beans	fava beans	Broadbeans (fava beans), mature seeds, raw	Beans, fava, in pod, raw
1 teaspoon ground cayenne pepper	cayenne pepper	Spices, pepper, red or cayenne	Spices, pepper, black
1 whole chicken with giblets patted dry and quartered	chicken with giblets	Chicken, broilers or fryers, meat and skin and giblets and neck, raw	Fast foods, quesadilla, with chicken
2 tablespoons sesame seeds	sesame seeds	Salad dressing, sesame seed dressing, regular	Seeds, sesame seeds, whole, dried
1 teaspoon ground coriander	coriander	Coriander (cilantro) leaves, raw	Spices, coriander leaf, dried

model obtained an F1 score of 0.95 on the test set validated by 5-fold cross validation.

The next two sections explain how ingredient descriptions from the recipes are mapped to ingredient description from the nutritional database (Section IIB) and secondly, how the units of the ingredients are mapped to one of the available units present in the nutritional database (Section IIC).

B. Closest Description Annotation Using String Similarity Matching

In order to accurately map ingredient names to food descriptions in the Standard Reference database, we carefully looked for patterns in food description strings that might help us select the best possible description.

(a) It can be observed that the descriptions in the USDA-SR database are comma-separated terms with a decreasing degree of importance associated with each consequent term. Consider all descriptions from the food description column of Table II. The first term is significant for matching. Hence, Butter, Cheese, Milk, Milk shakes, Yogurt, Egg and Apples occupy the highest priority for finding a match within the ingredient description.

(b) The high priority terms include both singular and plural forms of nouns. They must be lemmatized before matching. For this purpose, we used the NLTK library's WordNet Lemmatizer [6]. Stemmers, although computationally less expensive, were not found to be useful for this purpose because of their high aggression.

(c) The Ingredient Name "Egg whites" best matches with the description "Egg, white, raw, fresh" whereas "Whole eggs" best match "Egg, whole, raw, fresh". The sequence of terms may be different in both the strings being considered. To tackle this, we use a modified form of Jaccard Index [7] as the metric for the similarity between the two strings. The modified Jaccard distance has been explained in **(e)**.

(d) Another observation is that the comma-separated terms in later portions of the food description are more likely to match with the State, Temperature and Freshness of the ingredient. Therefore, we match the whole description along with the State, Temperature and Freshness entities derived from our NER pipeline.

(e) We would like to prioritize the mapping of maximum terms from the Ingredient phrase rather than the food description using a vanilla Jaccard Index. This is because a lot of food descriptions include additional details unspecified in the ingredient description. Assume A and B are the set of words formed after preprocessing the Ingredient Phrase and Food Description respectively by lemmatization, stop-word removal and uniform casing, and |A| and |B| are the number of words in these sets. Similarly $| A \cup B |$ gives the number of words in the union of the sets of words in strings A and B and $| A \cap B |$ gives the number of words in the intersection of the sets of words in strings A and B.

$$J(A,B) = \frac{| A \cap B |}{| A \cup B |}$$

For a lot of descriptions, |B| is extremely large, consider food description for serial numbers 7, 8, 9, 13, 14. The denominator in Jaccard distance increases with an increase in |B|. This leads to a bias against large strings. However, it is only essential to match the maximum number of terms from the Ingredient Phrase. So, we use |A| as the denominator for

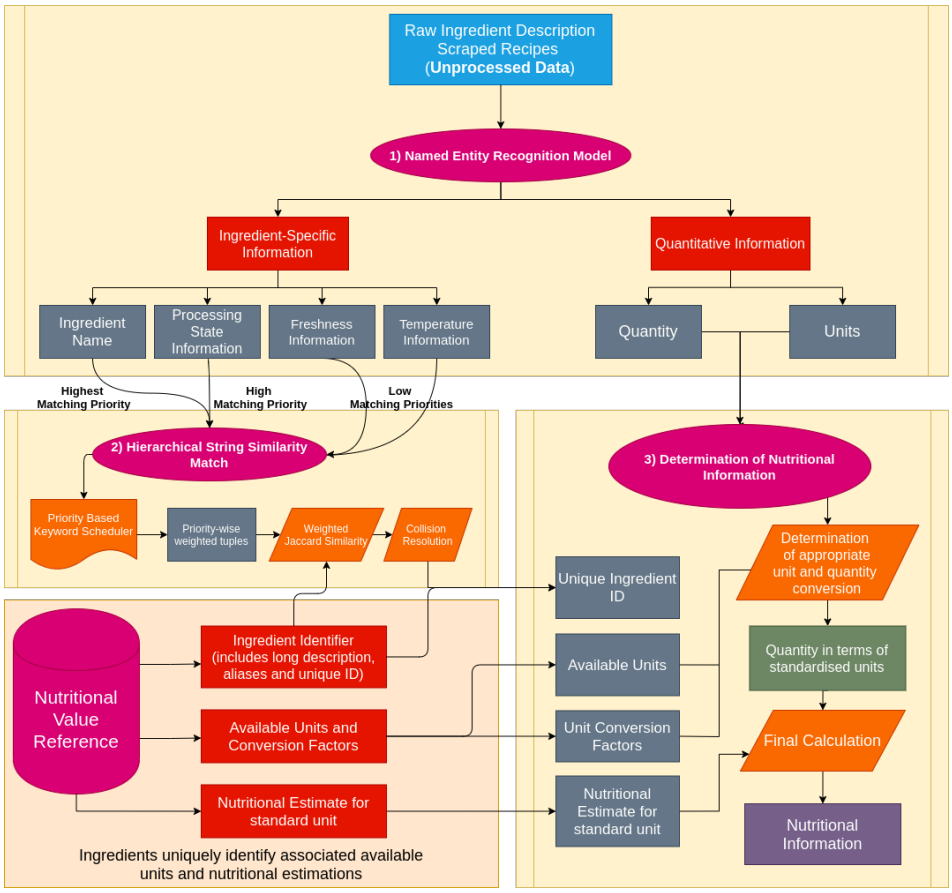


Fig. 1. System Architecture

our modified Jaccard Matching Index.

$$J^*(A,B) = \frac{|\,A \cap B\,|}{|\,A\,|}$$

This modification removes bias against large strings of detailed food descriptions without which "skimmed milk" would match with "Milk Shakes, thick chocolate" completely disregarding more accurate "Milk, reduced fat, fluid, 2% milkfat, protein fortified, with added vitamin A and vitamin D". This bias was found to be highly significant with 227 out of 1000 randomly sampled ingredient phrases from RecipeDB having a different match. Some observed inferences from the two metrics have been shown in Table III. It can clearly be seen that the vanilla Jaccard Index is biased towards shorter, less detailed food descriptions, proving that the Modified Jaccard Index is more suitable for the task.

(f) We also need to account for the presence of negation terms. "unsalted butter" should match with "Butter, without salt". To achieve this, we replaced all negation terms and prefixes (like "un" in unsalted) to "not". Following this, the preprocessing mentioned above is applied and the Ingredient Phrase and Description respectively become "not salt butter" and "butter not salt" leading to a perfect Jaccard Match.

(g) Many food descriptions include "raw" which is analogous to uncooked state or cases where State is not mentioned. So we include a provision to match an additional word when "raw" occurs in the description and no State is identified. This way, "apple" matches with "Apples, raw, with skin".

(h) We also store the sequence numbers (priority) of terms along with the Modified Jaccard Index score. This helps us in resolving collisions such as matching apple with Apples, raw, with skin and Babyfood, apples, dices, toddler. The first description is chosen because it contains the matching word Apples in the first term (having a higher priority than occurring at later indices). This follows the line of reasoning as explained in **(a)**

(i) In case collisions cannot be resolved through Jaccard Index and Sequential Priority, we simply take the first match. This is useful because of the way the descriptions have been indexed within USDA-SR Database. "Apple" matches with the more desirable "Apples, raw, with skin" as opposed to "Apples, raw, without skin" and "Eggs" match with "Egg, whole, raw, fresh" rather than the other variants which are more suitable for "egg whites" and "egg yolk" even though they both return the same Jaccard Index and priority score with "Egg".

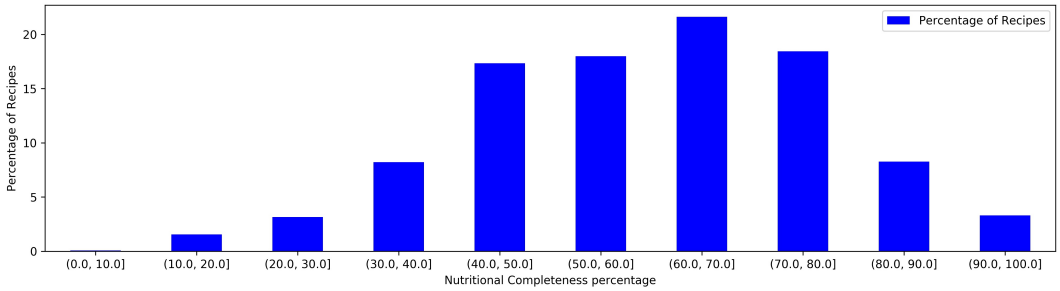


Fig. 2. Percentage mapping of recipes to their nutritional profile

C. Units Matching and subsequent Nutrition calculation

Once an ingredient from a recipe has been matched to its corresponding food description in the nutrition database, we match the unit corresponding to this ingredient from our nutritional database. Unfortunately, in this case string matching techniques would not be satisfactory since we have a fixed number of possible units with a fixed format and applying heuristics similar to Section II B might give unwanted results due to incorrect matching of strings. Furthermore, the units provided in the nutritional database may not be enough. For example - our dish requires "1 teaspoon of butter" whereas, in the USDA database, we do not have teaspoon in available units for butter. On top of that some of the units are not clean, e.g. 'pat (1" sq, 1/3" high)' was one of the units in the USDA-SR Database. See Table IV

Similar problems exist in the units used for the dish. Adding to that, we may have different aliases referencing the same unit in our data, e.g. 'tablespoon' and 'tbsp' refer to the same unit and so do 'pound' and 'lb'.

To circumvent these problems, we applied WordNet Lemmatization using NLTK library [6] on all the units present in our recipes and USDA-SR database then took the first word and applied Regular Expression(regex) to obtain a cleaner version containing only alphabets (this helps us to ignore noise and keep relevant part like taking pat out of 'pat (1" sq, 1/3" high)'). Furthermore, standard units were defined for units where aliases were present, for example, tbsp and tablespoon both now represent the standard unit tablespoon. To deal with the case where a unit could not be found, measurement conversion tables were created with detailed conversions between units on the basis of volume using measurements mentioned in [8]. These tables were used to check for the missing units. The tables mention conversions such as '1 cup' is equivalent to '16 tbsp' and '48 tbsp' and so on.

Table IV shows the data after lemmatizing and data cleaning. For butter, the units 'cup' and 'tablespoon' are present, but 'teaspoon' is not. Hence, we can add teaspoon as a unit since the ratio of volume of a cup and a teaspoon is constant. The sizes small, medium and large were mentioned in both, the nutritional database and recipes as units. All 3 were considered equivalent because of ambiguity between sizes.

TABLE IV. INGREDIENT AND UNIT RELATIONS

ingredient	seq	amount	unit	grams	gram_per_amount
Butter,salted	1	1.0	pat	5.0	5.0
Butter,salted	2	1.0	tbsp	14.2	14.2
Butter,salted	3	1.0	cup	227.0	227.0
Butter,salted	4	1.0	stick	113.0	113.0

After cleaning and lemmatizing the ingredient unit and quantity, we calculate the nutrition profile of each ingredient by merging the recipe data and nutrition data on the unit and multiplying the nutrition profile by the quantity of the ingredient. Quantities were preprocessed to match a specific numerical value. '2-4' was averaged to 3, '2 1/2' was converted to 2.5 and so on.

In certain cases NER did not detect units, in that scenario we searched the ingredient phrase for known units and if found they were updated. Certain ingredient phrases included statements such as '500 g or 1 cup' which the NER wrongly detected as '500 cups'. This was dealt in a semi-automated manner by putting a threshold on the quantity per unit. Finally, wherever a unit was still not present, the most frequent unit for that particular ingredient was used. This works well to maintain consistency in the data since we have a lot of units corresponding to each ingredient, but only a few of them are dominant, e.g. for garlic, if the unit was not detected, it would most probably be clove which we obtained using the above-mentioned heuristic.

III. Results

Using heuristics mentioned in II-B we were able to match 94.49% of the unique ingredients from the recipes, with the rest remaining unmapped from the USDA dataset. To assess the validity of the jaccard matching, the 5000 most frequent ingredients+states were manually matched with the USDA dataset, out of which 3580 were deemed to be correct matches, the rest had a better match available in the dataset (accuracy of 71.6%). It is important to note here that USDA has a lot

of similar ingredients with little variation as is evident from Table II, so while jaccard similarity does not always give the best match, it almost always gives one of the suitable matches from our database.

To further probe how many ingredients along with their units could be mapped to the USDA dataset, we analyzed percentage mapping of recipes to their nutritional profile in terms of the percentage of ingredients in a recipe getting mapped to their USDA nutritional profiles (Figure 2). It indicates that the protocol implemented could successfully map a significant proportion of ingredients to their nutritional profiles thereby contributing to the accuracy of the estimated nutritional profiles of recipes. The figure also indicates that the main problem lies in matching the units of ingredients to appropriate units in the USDA dataset, especially when some units are not mentioned in the nutritional database itself.

Calorie information from All Recipes was extracted and used as a baseline to evaluate our results. The nutritional profiling of recipes at AllRecipes was done by outsourcing it to a reliable third-party. We consider this as the gold standard for the evaluation of our estimated nutritional profiles. We selected data for which we had 100% mapping of ingredients with their nutritional values, and had clean, well-defined servings. This resulted in 2482 recipes. This was done because while our recipe dataset has a global coverage, spanning 26 regional cuisines, the sample food composition table that was used mostly contained details of ingredients used in the United States. For e.g. ‘garam masala’- a spice used in Indian dishes is not an ingredient present in the dataset. Because of these region-centric ingredients, some ingredients were not mapped. Incorporation of other data as mentioned in Food and Agricultural Organisation of the United Nation[5] would help in improving the results.

For the chosen 2482 recipes we observed an average per serving error of 36.42 calories. To put this in the context in terms of the ingredient used in Table IV, (butter, salted), 1 teaspoon of it is equivalent to 35 calories. This is well within our scope of error since some calorie content would differ based on the user, cooking time and utensils used in cooking the dish.

IV. Conclusions

We use NER with Jaccard Similarity and Unit Mapping on a large database containing more than 118,000 recipes to provide accurate estimates of nutritional profiles despite extremely noisy and varied data. We show that the proposed protocol is robust, compatible with any nutritional database, easily replicable and solves one of the foremost problems with dietary analysis and food recommendation systems. We provide the code on Github[6]. We would like to highlight that our system provides a good ‘estimate’ for the nutritional value of food and as nutritional composition tables get updated, our heuristics will give better results without any changes.

[5]http://www.fao.org/infoods/infoods/tables-and-databases/en/

[6]https://github.com/cosylabiiit/Nutritional-Estimation-In-Recipes

Acknowledgment

J.K. and D.B. contributed equally towards the the work done for this research article. J.K., D.B. and N.D. thank IIIT-Delhi and Complex Systems Laboratory for the Summer Research Internship. G.B. thanks IIIT-Delhi for the computational facilities. The authors thank Rudraksh Tuwani for suggesting the use of NER as part of computational protocol.

References

[1] Batra, Devansh, Nirav Diwan, Utkarsh Upadhyay, Jushaan Singh Kalra, Tript Sharma et al. ”RecipeDB: A Resource for Exploring Recipes.”, Preprint.

[2] Al-Maghrabi, Rana. Measuring Food Volume and Nutritional Values from Food Images. Diss. Universit d’Ottawa/University of Ottawa, 2013.

[3] Schakel et al, Procedures for estimating nutrient values for food composition databases,J. Food Compos. Anal, 1997, pp. 102-114

[4] Bognr, A., and J. Piekarski. ”Guidelines for recipe information and calculation of nutrient composition of prepared foods (dishes).” Journal of food composition and analysis 13.4, 2000, pp. 391-410.

[5] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics, 2005, pp. 363-370.

[6] Loper, Edward, and Steven Bird. ”NLTK: the natural language toolkit.” arXiv preprint cs/0205028, 2002.

[7] Niwattanakul, Suphakit, et al. ”Using of Jaccard coefficient for keywords similarity.” Proceedings of the international multiconference of engineers and computer scientists. Vol. 1. No. 6. 2013.

[8] Francis T. Lynch.The Book of Yields: Accuracy in Food Costing and Purchasing, 8th Edition,2010

Astrophys Space Sci (2020) 365:84
https://doi.org/10.1007/s10509-020-03799-1

ORIGINAL ARTICLE

Observational constraints on holographic dark energy model with matter creation

Ajay Kumar[1] · **C.P. Singh**[1]

Received: 8 March 2020 / Accepted: 11 May 2020
© Springer Nature B.V. 2020

Abstract In this paper, we consider a holographic dark energy model driven by the gravitationally induced adiabatic matter creation with constant creation rate. It is assumed that the present accelerating stage is a consequence of the negative pressure describing the irreversible process of gravitational matter creation. Exact solutions are obtained to discuss the evolution of the universe. We obtain the best-fit values of the model parameters through MCMC method by the use of EMCEE python package on latest observational data. With the best-fit values, it is found that the universe shows transition from decelerated phase to accelerated phase. We carry out the diagnostic analysis to distinguish our model with other standard dark energy models. We also discuss the cosmography analysis of our model. We analyse the model by applying Akaike information criterion (AIC) and Bayesian information criterion (BIC) based on the penalization associated to the number of parameters. The results show that the model behaves quintessence like and approaches to ΛCDM model in late time.

Keywords Cosmology · Holographic dark energy model · Matter creation

1 Introduction

The observational data from supernova provides a strong evidence for the accelerated expansion of the universe (Riess et al. 1998; Perlmutter et al. 1999). It has also been supported by the observations of the Cosmic Microwave Background (CMB) (Komatsu et al. 2011), the Baryon Acoustic Oscillations (BAO) (Sanchez et al. 2012) and Planck data (Ade et al. 2014, 2016). According to observations, the present epoch of evolution is dominated by an exotic energy known as 'dark energy' (DE), which has negative pressure and occupies $\approx 68.3\%$ of the total energy density of the universe. The rest matter of the universe is assumed to be dark matter and the baryon matter. In the literature, many theories have been proposed to understand the mechanism of the dark energy. Among the many possible proposals, possibly a cosmological constant (Λ) is the simplest answer to explain the late time cosmic acceleration (Peebles and Ratra 2003). However, this also brings some theoretical difficulties, which are how to explain its origin, right magnitude and why it comes to dominate just now (Weinberg 1989; Sahni and Starobinsky 2000; Padmanabhan 2003). Therefore, it is natural to look for alternative models which can explain the accelerated expansion of the universe.

A new DE candidate, based on the holographic principle, was proposed by 't Hooft (1993). The principle states that the number of degrees of freedom which is finite in a bounded system has relation with the area of its boundary (Susskind 1995; Cohen et al. 1999). In cosmology, one can obtain the upper bound of the entropy contained in the universe. Cohen et al. (1999) stated that for a system with size L, the total energy should not exceed the mass of black hole of the same size for the quantum zero-point energy density associated with the ultraviolet cut-off, thus $L^3\rho_d \leq LM_p^2$, where ρ_d is the vacuum energy density caused by UV cut-off, M_p (here $M_p^2 \sim \frac{1}{8\pi G}$) is the reduced Planck mass and L is the infrared cut-off. In cosmological context, while taking the largest possible L to saturate this inequality, we get the energy density related to holographic principle and is de-

✉ C.P. Singh
cpsphd@rediffmail.com

A. Kumar
maths.ajy@gmail.com

[1] Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi 110 042, India

Published online: 19 May 2020

fined as holographic dark energy (HDE) which is given by $\rho_d = 3b^2 M_p^2 L^{-2}$, where b is a numerical constant and coefficient 3 is for mathematical convenience. A duality between UV and IR cut-off means that L in both cases are related to the vacuum energy and the large scale of the universe respectively. The large scale of the universe can be taken as, for example Hubble horizon, event horizon or particle horizon (Hsu 2004; Li 2004; Huang and Li 2004; Zhang and Wu 2005; Chen et al. 2007; Feng et al. 2007; Li et al. 2009; Luongo 2017; Malekjani et al. 2018).

Just like the early-time cosmic acceleration associated with inflation, a negative pressure can be seen as a possible driving mechanism for the late-time accelerated expansion of the universe as well. One of the earliest alternatives that could provide a mechanism producing such accelerating phase is through a negative pressure produced by viscous or particle production effects. Adiabatic matter creation is a non-equilibrium thermodynamical process which makes use of ideas of the thermodynamics of open systems in the context of cosmology. The merits of the matter creation with respect to usual DE ideology are (a) the former has a strong physical basis namely non-equilibrium thermodynamics, while the latter (DE) has not and (b) the matter creation mechanism unities the dark sector (DE and DM), since a single dark component (the dark matter) needs to be introduced into the cosmic fluid and thus it contains only on free parameter.

Gunzig et al. (1983) and Prigogine et al. (1988, 1989) discussed the theoretical aspects of matter creation in the framework of cosmology and showed that we can modify the second law of thermodynamics to accommodate the flow of energy from gravitational field to the matter field which results in particle creation and consequently entropy. Many authors (Lima et al. 1991; Calvao et al. 1992; Lima and Germano 1992; Lima et al. 1996; Gunzig et al. 1998; Alcaniz and Lima 1999) have explored the scenarios of matter creation in cosmology, but here we are particularly interested in the gravitationally induced particle creation scenario in which a specific choice of the particle creation rate is assumed. Some authors (Lima et al. 2008; Steigman et al. 2009; Cardenas 2012; Lima et al. 2014; Ramos et al. 2014; Nunes and Pavon 2015; Pan et al. 2018) have discussed the matter creation cosmology with different forms of matter creation rate. In these works, it has been shown that phenomenological particle creation can explain not only the present era of cosmic acceleration but also provide a viable alternative to the concordance ΛCDM model. Recently, Singh and Kaur (2019) have studied matter creation cosmology in Brans-Dicke theory and constrained the model with combined set of observational data.

In this paper, we consider the cosmological consequences due to adiabatic matter creation mechanism in HDE model with Hubble horizon as an IR cut-off. It has been observed that the HDE model with Hubble horizon as an IR cut-off is not consistent with the current observations. Therefore, instead of assuming other IR cut-off, we include the matter creation mechanism in HDE model with the same Hubble horizon as an IR cut-off with the possibility that this model would be consistent with the current observations. We consider constant particle creation rate and solve the field equations to obtain the exact solution. We discuss the evolution and dynamics of the model by constraining the model parameters through combined observational data of type Ia Supernova, observational Hubble data and combined data of baryon acoustic oscillations and cosmic microwave background. We plot the trajectories of the scale factor, matter density parameter and deceleration parameter for best-fit values of model parameters. We compare our model with the standard ΛCDM model. We also plot the trajectory of Hubble evolution of our model and ΛCDM with error bar from Hubble data. As a comparison, the model with matter creation is examined with the selection criteria of AIC and BIC.

The paper is organized as follows: In Sect. 2 we present the HDE model with matter creation. Section 3 deals with the solution of the field equations with constant creation rate. In Sect. 4, we discuss the observational analysis by implementing the MCMC method by the use of EMCEE package on publicly available observational data sets. Section 5 is further divided into four Sects. 5.1-5.4. In Sect. 5.1, we discuss the evolution by plotting the trajectories of the scale factor, matter density parameter, deceleration parameter, effective equation of state parameter, Hubble error bar plot and the age of the universe. In Sect. 5.2, we discuss statefinder and $Om(z)$ analysis using best-fit values of model parameters. In Sect. 5.3, we extend the geometric investigation of the model by finding various cosmography parameters. Further, we discuss model selection criteria using AIC and BIC for the model in Sect. 5.4. Finally, we present the summary of our work in Sect. 6.

The units used throughout the paper are: $c = 8\pi G = \hbar = 1$.

2 Model and field equations

We consider a spatially flat, homogeneous, and isotropic universe described by the Friedmann-Robertson-Walker (FRW) line element

$$ds^2 = -dt^2 + a^2(t)\left[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right], \qquad (1)$$

and filled with a perfect fluid with energy density ρ and isotropic dynamical pressure P described by the energy-momentum tensor

$$T_{\mu\nu} = (\rho + P)u_\mu u_\nu + P\, g_{\mu\nu}, \qquad (2)$$

where u^μ is the fluid four-velocity (with normalization $u^\mu u_\mu = -1$).

By introducing holographic dark energy and gravitationally induced matter creation into the energy-momentum tensor, the total energy density $\rho = \rho_m + \rho_d$ and total effective pressure $P = p_m + p_c + p_d$. Here, ρ_m is the energy density of dark matter, ρ_d the energy density of HDE, p_m the pressure of dark matter, p_c the pressure due to the matter creation and p_d is the pressure of HDE.

In this background, the nontrivial Einstein's field equations are given by

$$3H^2 = \rho = \rho_m + \rho_d, \tag{3}$$

$$2\dot{H} + 3H^2 = -P = -(p_m + p_c + p_d), \tag{4}$$

where an over dot represents the derivative with respect to cosmic time t and $H = \dot{a}/a$ is the Hubble parameter. As a consequence of (3) and (4), the energy conservation equation is given by

$$\dot{\rho} + 3H(\rho + P) = 0. \tag{5}$$

Prigogine and his collaborators (Prigogine et al. 1988, 1989) studied how to insert the creation of matter consistently in Einstein's field equations. This was achieved by introducing the usual balance equation for the number density of particles, $(nu^\mu)_{;\mu} = 0$, a source term on the right-hand side to account for production of particles, namely,

$$(nu^\mu)_{;\mu} = n\Gamma, \tag{6}$$

where Γ denotes the particle production rate. Using the FRW metric we obtain a new balance equation

$$\dot{n} + 3Hn = n\Gamma. \tag{7}$$

The above equation, when combined with the second law of thermodynamics, naturally leads to the appearance of a negative pressure directly associated to the rate Γ, the creation pressure p_c, which adds to the other pressures (i.e., of radiation, baryons, dark matter and vacuum pressure) in the total energy-momentum tensor. For a homogeneous and isotropic universe, in which there is an adiabatic process of particle creation, it is easily found that

$$p_c = -\frac{(\rho_m + p_m)}{3H}\Gamma. \tag{8}$$

Therefore, since p_c is negative, it can help to drive the era of accelerated cosmic expansion. Since the production of ordinary particles is much limited by tight constraints imposed by local gravity measurements, and radiation has a negligible impact on the recent cosmic dynamics, for sake of simplicity, we will assume that the produced particles are just dark matter particles. We consider that there is no interaction between the two fluids. Therefore, the energy conservation equation (5) (with pressureless dark matter, i.e., $p_m = 0$) for dark matter and HDE conserves separately which are given by

$$\dot{\rho_m} + 3\rho_m H = \rho_m \Gamma, \tag{9}$$

$$\dot{\rho_d} + 3(\rho_d + p_d)H = 0. \tag{10}$$

In this paper, we consider the IR cut-off as the inverse of Hubble scale ($L = H^{-1}$) in the relation $\rho_d = 3b^2 M_p^2 L^{-2}$ so that the corresponding energy density is given by

$$\rho_d = 3b^2 H^2. \tag{11}$$

Cosmic inflation and the accelerated expansion of the universe can be characterized by the equation of state (EoS) of dark energy. Since we consider the universe to be filled with perfect fluid, therefore, we have a linear EoS, $p_d = \omega_d \rho_d$ for the dark energy, where ω_d is the EoS parameter. The condition for accelerated expansion $\ddot{a} > 0 \Rightarrow p < (-1/3)\rho$. It has been observed that the linear dark energy EoS reveals by Supernova (Salzano et al. 2014).

Therefore, assuming the linear EoS for HDE, and using (8), (9) (10) and (11), Eq. (3) gives the following evolution equation:

$$\dot{H} + \frac{3}{2}\left[(1-b^2)\left(1 - \frac{\Gamma}{3H}\right) + (1+\omega_d)b^2\right]H^2 = 0. \tag{12}$$

Let us discuss how the matter creation rate, Γ modifies the evolution of the HDE model with matter creation as compared to the case to the HDE model without matter creation.

As already mentioned, the HDE model has been proposed on the basis of duality between ultraviolet (UV) cut-off and infrared cut-off. The UV cut-off is related to the vacuum energy, and IR cut-off is related to the large scale of the universe. Hsu (2004) studied HDE model with Hubble horizon as an IR cut-off and found that the HDE model with Hubble horizon as an IR cut-off is not able to explain the current accelerated expansion. He found the dark energy EoS parameter $\omega_d = 0$, which is not consistent with the current observations. In what follows, let us review the HDE model with Hubble scale without matter creation.

In the absence of matter creation, i.e., taking $\Gamma = 0$, Eq. (12) gives the solution of the Hubble parameter,

$$H = H_0\left(\frac{a_0}{a}\right)^{\frac{3}{2}(1+b^2\omega_d)} = H_0(1+z)^{\frac{3}{2}(1+b^2\omega_d)}, \tag{13}$$

where a_0 and H_0 are the present values of the scale factor and the Hubble parameter, respectively and $z = (a_0/a) - 1$ is the redshift. Throughout this paper we use the subscript "0" to denote the present value of quantities.

From the above equation the scale factor can be obtained as

$$a = a_0 \left[1 + \frac{3}{2}(1 + b^2\omega_d)H_0(t - t_0)\right]^{\frac{2}{3(1+b^2\omega_d)}}, \tag{14}$$

which shows that the expansion of the universe is of power-law form. The deceleration parameter, which is defined as $q = -a\ddot{a}/\dot{a}^2$, gives

$$q = -1 + \frac{3}{2}(1 + b^2\omega_d), \tag{15}$$

which is constant. Therefore, no transition redshift in the universe will be achieved when the Hubble horizon is taken as the IR cut-off. The evolution of the scale factor is shown in Fig. 3, which shows decelerated expansion through out the evolution.

To overcome such problems, many authors (Li et al. 2009; Chen et al. 2007; Zhang and Wu 2005) proposed different IR cut-off like particle horizon, event horizon, Ricci scalar, etc., to discuss the accelerating universe. However, the Hubble horizon is a natural choice of cosmological length scale. Therefore, instead of considering any other IR cut-off or interaction between the fluids as did by many authors, we restrict ourselves to Hubble horizon as an infrared cut-off but in the presence of gravitationally induced matter creation to realize an accelerated expansion universe, a transition from the past decelerated expansion to recent accelerated expansion and cosmic observational compatible model. The main advantage of considering HDE with matter creation is the introduction of another free parameter, thus providing an extra degree of freedom in the process of constructing a physical interpretation of the solution. The evolution of matter creation model has been tested by latest observational data sets.

3 Solution of HDE model with matter creation

In the presence of matter creation, Eq. (12) can be fully solved once the particle creation rate Γ is specified. Due to the absence of a rigorous quantum field theory in curved space-time, including (non-equilibrium) back reaction, form which the expression of Γ should be calculated, a number of phenomenological parametrization of Γ have been proposed, e.g., $\Gamma = constant$ (de Haro and Pan 2016), $\Gamma \propto H$ (Pan and Chakraborty 2015), $\Gamma \propto H^2$ (Abramo and Lima 1996; Gunzig et al. 1998) and a linear combination of these (Pan et al. 2016).

In this paper, we consider HDE model driven by some constant particle creation rate, $\Gamma = 3\alpha H_0$, where α is order unity dimensionless constant and H_0 is the present value of the Hubble parameter.

Taking this assumption of Γ in (12) and using $\frac{d}{dt} = \frac{\dot{a}}{a}\frac{d}{d\,ln\,a}$, we obtain the following dimensionless equation:

$$h' + \frac{3}{2}(1 + b^2\omega_d)h = \frac{3}{2}\alpha(1 - b^2), \tag{16}$$

where $h = \frac{H}{H_0}$ is the dimensionless Hubble parameter and a prime denotes $\frac{d}{d\,ln\,a}$.

Considering the relation $h(a_0) = 1$, we can solve Eq. (16) analytically to obtain

$$h(a) = \frac{(1 - b^2)\alpha}{(1 + b^2\omega_d)} + \left\{1 - \frac{(1 - b^2)\alpha}{(1 + b^2\omega_d)}\right\}\left(\frac{a}{a_0}\right)^{-\frac{3(1+b^2\omega_d)}{2}} \tag{17}$$

where $b^2 \neq -1/\omega_d$. Integrating (17), we obtain the scale factor $a(t)$ as

$$\frac{a(t)}{a_0} = \left[1 + \frac{(1 + b^2\omega_d)}{(1 - b^2)\alpha}\left\{e^{\frac{3}{2}(1-b^2)\alpha H_0(t-t_0)} - 1\right\}\right]^{\frac{2}{3(1+b^2\omega_d)}} \tag{18}$$

Now onwards, we will consider $a_0 = 1$. From the above equation, we can see that at the early time (when $t \to 0$), the scale factor

$$a(t) \to \left[1 + \frac{3}{2}(1 + b^2\omega_d)H_0(t - t_0)\right]^{\frac{2}{3(1+b^2\omega_d)}}$$

which is of power-law form. In the late time (when $t \to \infty$), the scale factor $a(t) \to e^{\frac{(1-b^2)\alpha}{(1+b^2\omega_d)}H_0(t-t_0)}$ which corresponds to the de Sitter universe. Thus, the HDE model can explain early deceleration and late time accelerating phase with constant matter creation rate.

The Hubble parameter in terms of redshift z can be given as

$$\frac{H(z)}{H_0} = \left[\frac{(1 - b^2)\alpha}{(1 + b^2\omega_d)} + \left\{1 - \frac{(1 - b^2)\alpha}{(1 + b^2\omega_d)}\right\}(1 + z)^{\frac{3}{2}(1+b^2\omega_d)}\right]. \tag{19}$$

To study the decelerated or accelerated epochs and the transitions between them, we assume $H_0(t - t_0) = y$. The second order derivative of the scale factor $a(t)$ with respect to y gives

$$\frac{d^2a}{dy^2} \propto \left[2 - 3(1 - \alpha + \alpha b^2 + b^2\omega_d)e^{\frac{3}{2}(b^2-1)\alpha y}\right] \tag{20}$$

From (20), we can check the behavior of the scale factor so that the universe shows decelerated expansion in the early

time followed by a transition to accelerated expansion in the late time. The conditions satisfying above description are

$$(b^2-1)\alpha>0 \quad \text{and} \quad (1-3\alpha+3\alpha b^2+3b^2\omega_d)<0. \quad (21)$$

These conditions will be used to find the best fit values of model parameters.

In this model, the deceleration parameters can be obtained as

$$q=-1+\frac{3}{2}[(1+b^2\omega_d)-(1-b^2)\alpha]e^{\frac{-3}{2}(1-b^2)\alpha H_0(t-t_0)} \quad (22)$$

In terms of redshift z, q is given by

$$q(z)=-1+\frac{3}{2}\frac{(1+b^2\omega_d)-(1-b^2)\alpha}{\left[1+\frac{(1-b^2)\alpha}{(1+b^2\omega_d)}\left\{(1+z)^{\frac{-3}{2}(1+b^2\omega_d)}-1\right\}\right]}. \quad (23)$$

It can be seen that $q(z)\to -1$ as $z\to -1$. In the absence of matter creation ($\alpha=0$), this expression yields $q=(1+3b^2\omega_d)/2$, which is same as given in Eq. (15). While, in the absence of dark energy matter ($b=0$), we find

$$q(z)=-1+\frac{3}{2}\frac{(1-\alpha)}{\left[1+\alpha\left\{(1+z)^{-\frac{3}{2}}-1\right\}\right]}. \quad (24)$$

Thus, we find that the existence of a transition redshift at late times depends exclusively on the α parameter, i.e., on the matter creation. We can find the current value of q by putting $z=0$ in (23) as

$$q_0=\frac{(1-3\alpha+3\alpha b^2+3b^2\omega_d)}{2}. \quad (25)$$

Let us find the redshift, z_{tr}, at which the universe transits from the decelerating phase and enters into the current accelerating phase, i.e., $q(z)=0$. It is given by

$$1+z_{tr}=\left[1+\frac{(1+b^2\omega_d)(1-3\alpha+3\alpha b^2+3b^2\omega_d)}{2\alpha(1-b^2)}\right]^{\frac{-2}{3(1+b^2\omega_d)}} \quad (26)$$

The cosmic time of transition "t_{trans}" between the decelerated to the accelerated epochs can be given by

$$t_{trans}=t_0+\frac{2}{3(b^2-1)\alpha H_0}\,ln\left(\frac{2}{3(1-\alpha+\alpha b^2+b^2\omega_d)}\right). \quad (27)$$

The effective equation of state (EoS) parameter, ω_{eff} using the standard relation $\omega_{\text{eff}}=-1-\frac{1}{3}\frac{2a}{h}\frac{dh}{da}$, is given by

$$\omega_{\text{eff}}(z)=-1+\frac{(1+b^2\omega_d)+(b^2-1)\alpha}{\left[1+\frac{(1-b^2)\alpha}{(1+b^2\omega_d)}\left\{(1+z)^{-\frac{3}{2}(1+b^2\omega_d)}-1\right\}\right]}. \quad (28)$$

The condition of EoS parameter, $\omega<-1/3$ for acceleration of the present universe can be obtained by taking $z=0$ in (28), which is given by

$$3\omega_{\text{eff}}+1=(1-3\alpha+3\alpha b^2+3b^2\omega_d)<0. \quad (29)$$

4 Observational tests: SNe, OHD, BAO/CMB data

In this section, we briefly present the statistical analysis to constraint the matter creation HDE model presented in the previous section. First of all, we consider the Type Ia supernova (SNe) data. We use the latest Joint Light Curve analysis (JLA) sample comprising 31 data points in the redshift range $z\in[0.01, 1.3]$ (Betoule et al. 2014). The χ^2 is defined by

$$\chi^2_{SNe}=r^t C_b^{-1} r, \quad (30)$$

where $r=\mu_b-M-5\,logd_L$. Here μ_b is the observational distance modulus for a supernova at redshift z, M is a free normalization parameter and C_b is the covariance matrix of μ_b for which see Table F.2 in Betoule et al. (2014). In the above equation d_L is the luminosity distance which is defined by

$$d_L=\frac{c(1+z)}{H_0}\int_0^z\frac{dz'}{H(z',\theta)} \quad (31)$$

where c is the speed of light and θ is the representation of model parameters α and ω_d.

On the other hand, For observational Hubble data (OHD), we use the 43 measurements of the Hubble parameter values spread in the interval $0<z<2.5$, see (Cao et al. 2018). The χ^2 function is defined as

$$\chi^2_{OHD}=\sum_{i=1}^{n}\frac{[H(z_i)-H_{obs}(z_i,\theta)]^2}{\sigma_i^2} \quad (32)$$

where $H(z_i)$ and $H_{obs}(z_i,\theta)$ represents the theoretical and observed values of Hubble parameter, respectively. The standard deviation in the observed value is denoted by σ_i.

Finally, we consider the combined baryon acoustic oscillation (BAO) and cosmic microwave background (CMB) data, i.e., BAO/CMB data from different observational missions (dos Santos et al. 2016). We have taken the sample of BAO distances measurements from SDSS(R) (Padmanabhan et al. 2012), the 6dF Galaxy survey (Beutler et al. 2011), BOSS CMASS (Anderson et al. 2014) and three parallel measurements from WiggleZ survey (Blake et al. 2012). We combine these results with the Planck 2015 (Ade et al. 2016).

In the context of BAO, the angular diameter, $d_A(z,\theta)$ is defined as

$$d_A(z_*,\theta) = c\int_0^{z_*}\frac{dz'}{H(z',\theta)} \tag{33}$$

where z_* indicates the photons decoupling redshift and hold the value $z_* = 1090$ as per the Planck 2015 results (Ade et al. 2016). $D_v(z,\theta)$ represents the dilation scale which is given by $D_v(z,\theta) = \left(\frac{d_A^2(z,\theta)cz}{H(z,\theta)}\right)^{1/3}$.

The distant redshift d_z is given by

$$d_z = \frac{r_s(z_*)}{D_v(z)}, \tag{34}$$

where $r_s(z_*)$ is defined as the co-moving sound horizon at the time when photons decouple, which is assumed to be the same as it is considered in Ref. (dos Santos et al. 2016).

With this information, the χ^2 function for BAO/CMB can be written as (dos Santos et al. 2016)

$$\chi^2_{BAO/CMB} = A^T C^{-1} A \tag{35}$$

where A is the matrix

$$A = \begin{bmatrix} \frac{d_A(z_*,\theta)}{D_v(0.106,\theta)} - 30.84 \\ \frac{d_A(z_*,\theta)}{D_v(0.35,\theta)} - 10.33 \\ \frac{d_A(z_*,\theta)}{D_v(0.57,\theta)} - 6.72 \\ \frac{d_A(z_*,\theta)}{D_v(0.44,\theta)} - 8.41 \\ \frac{d_A(z_*,\theta)}{D_v(0.6,\theta)} - 6.66 \\ \frac{d_A(z_*,\theta)}{D_v(0.73,\theta)} - 5.43 \end{bmatrix}$$

and C^{-1} is the inverse of the covariance matrix (dos Santos et al. 2016). We have taken the correlation coefficient from Ref. (Hinshaw et al. 2013).

In order to extract the observational constraints and incorporate the datasets statistically, we use the Markov chain Monte Carlo (MCMC) package EMCEE to perform the numerical analysis of our models (Foreman-Mackey et al. 2013). In our analysis, we have considered the value of the Hubble constant as $H_0 = 67.8$ *Km/s/Mpc* (Ade et al. 2016). In this paper, we also use the value of constant $b = 0.50$ which is in the range $b = 0.495 \pm 0.039$ obtained by Li et al. (2013) from the Planck constraints on HDE model.

We can combine the above probes by using a joint likelihood analysis: $\mathcal{L}_{tot} = \mathcal{L}_{SNe} \times \mathcal{L}_{OHD} \times \mathcal{L}_{BAO/CMB}$ or $\chi^2_{tot} = \chi^2_{SNe} + \chi^2_{OHD} + \chi^2_{BAO/CMB}$. The parameters space for our model is $\{\alpha, \omega_d, M\}$. Here, M is a free normalization parameter used in SNe Ia observation. We minimize χ^2_{tot} to obtain the best-fit values of the free parameters. We have presented the observational summary in Table 1 and one of the contour plots is shown in Fig. 1 with $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level.

Table 1 The data fitting results of the free parameters in the HDE model with matter creation

Parameter	[a]SOB	[b]SB	[c]SO
α	$0.673^{0.010}_{-0.010}$	$0.584^{0.019}_{-0.021}$	$0.750^{0.023}_{-0.038}$
ω_d	$-0.002^{0.002}_{-0.003}$	$-0.010^{0.007}_{-0.015}$	$-0.051^{0.040}_{-0.085}$
M	$24.994^{0.018}_{-0.018}$	$25.020^{0.017}_{-0.019}$	$24.963^{0.019}_{-0.018}$

[a]SOB=SNe+OHD+BAO/CMB

[b]SB=SNe+BAO/CMB

[c]SO=SNe+OHD

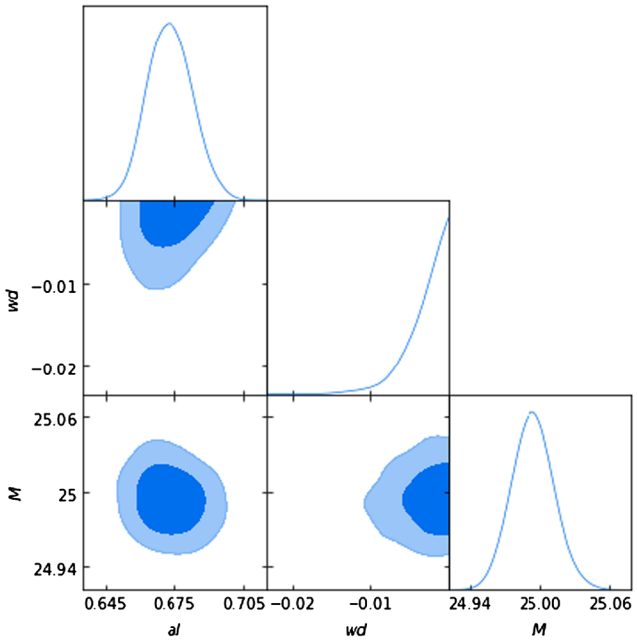


Fig. 1 1σ and 2σ confidence level contour plots for the model parameters using the observational data SNe+OHD+BAO/CMB. The labels al and wd denote α and ω_d parameters, respectively

5 Behavior of the model

In this section, we discuss the evolution, statefinder parameters, $Om(z)$ analysis, cosmography analysis and model selection of HDE model with matter creation by considering the best-fit values obtained in Sect. 4.

5.1 Evolution of the model

We plot the evolution of the scale factor in Fig. 2 for the best-fit model parameters listed in Table 1 obtained from different observational data sets. From the figure it can be observed that the expansion starts with accelerated rate in early time followed by deceleration and then the exponential acceleration at the late time. In Fig. 2, the dot on the trajectory denotes the transition point whose values are listed in Table 2.

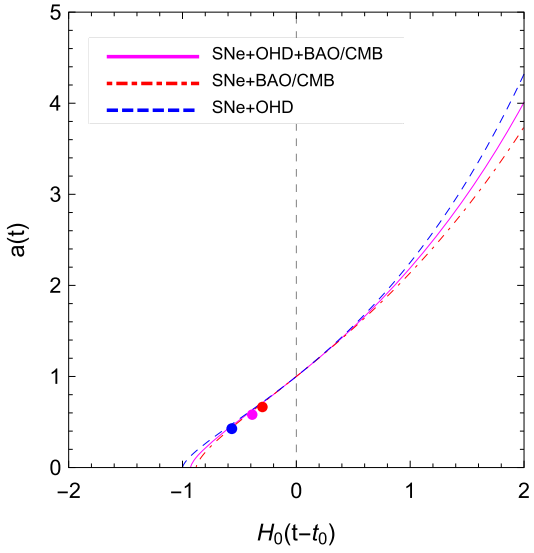


Fig. 2 Evolution of the scale factor as a function of time for different combinations of data sets in HDE model with matter creation. A dot on each trajectory denotes the transition scale factor, a_{tr}

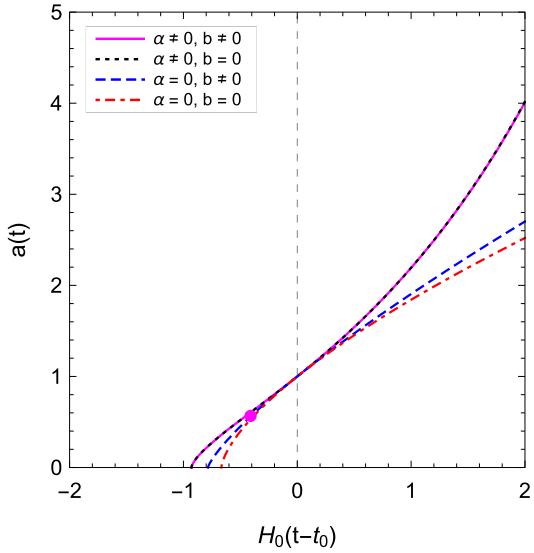


Fig. 3 Evolution of the scale factor as a function of time for different models where the combined observational data of $SNe + OHD + BAO/CMB$ is used. A dot on the trajectory denotes the transition scale factor, a_{tr}

Table 2 The numerical values of a_{tr}, z_{tr}, q_0, $\omega_{eff}(z=0)$ and $t_0(Gyr)$ using the best-fit result of model parameters

Data	a_{tr}	z_{tr}	q_0	$\omega_{eff}(z=0)$	t_0
SOB	0.620	0.611	−0.258	−0.505	13.393
SB	0.737	0.356	−0.161	−0.440	12.691
SO	0.504	0.982	−0.363	−0.575	14.414

In Fig. 3, we plot the evolution of the scale factor for different combination of models to compare their behaviors. The evolutions of the HDE with matter creation ($\alpha \neq 0$ and $b \neq 0$) and matter-dominated model with matter creation ($\alpha \neq 0$ and $b = 0$) approximately coincide to each other. These two trajectories show that the both the models evolve with early acceleration, medieval deceleration and then late time acceleration. The other two evolutions, the HDE model without matter creation ($\alpha = 0$ and $b \neq 0$) and matter-dominated model $\alpha = 0$ and $b = 0$ decelerate through out their evolutions.

We also introduce the matter density parameter Ω_m, which is defined by $\Omega_m = \rho_m/\rho_{cr}$, where $\rho_{cr} = 3H_0^2$ is the critical energy density. Using (3) and (19), we obtain

$$\Omega_m(a) = (1-b^2) \times \left[\frac{(1-b^2)\alpha}{(1+b^2\omega_d)} + \left\{1 - \frac{(1-b^2)\alpha}{(1+b^2\omega_d)}\right\}\left(\frac{a}{a_0}\right)^{\frac{-3}{2}(1+b^2\omega_d)}\right]^2, \tag{36}$$

where $\alpha \neq 0$ and $b \neq 0$. From the above equation, it is to be noted that for $\alpha = 0$ and $b = 0$, i.e., in the absence of both HDE and matter creation, the matter density parameter follows $\Omega_m \sim (a/a_0)^{-3}$, which corresponds to the matter dominated phase. However, in the absence of matter creation only, i.e. putting $\alpha = 0$ and $b \neq 0$, $\Omega_m \sim (1-b^2)\left(\frac{a}{a_0}\right)^{\frac{-3}{2}(1+b^2\omega_d)}$. In the presence of matter creation, i.e., $\alpha \neq 0$ and $b = 0$, we have $\Omega_m \sim [\alpha + (1-\alpha)(a/a_0)^{-3}]$. The evolution of Ω_m with the scale factor for the best fit values are shown in Fig. 4 which shows that how does the matter creation influence the behavior of Ω_m. In each case, as $a \to 0$, the matter density diverges which indicates the existence of the Big-Bang at the origin of the universe.

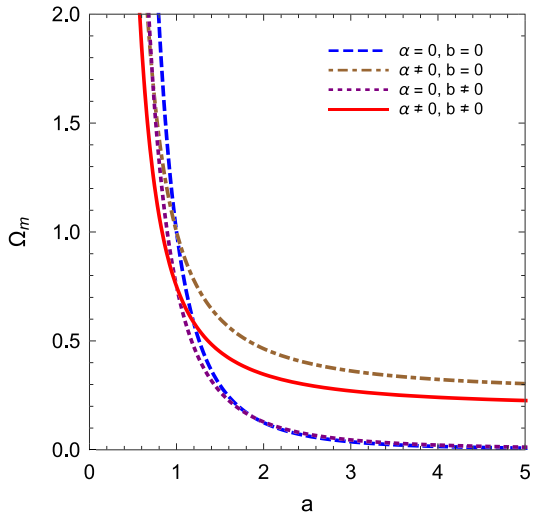


Fig. 4 Evolution of the mass density parameter with the scale factor for different combinations of best fit values of the model parameters where the combined data $SNe + OHD + BAO/CMB$ is used

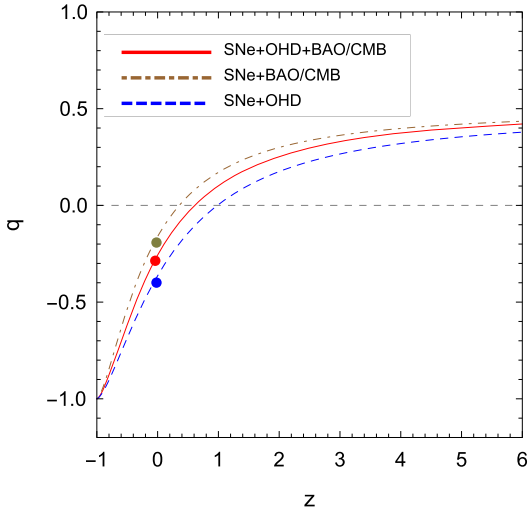


Fig. 5 Plot of the deceleration parameter q as a function of redshift z for the best-fit values of parameters. A dot on each trajectory denotes the current value, q_0

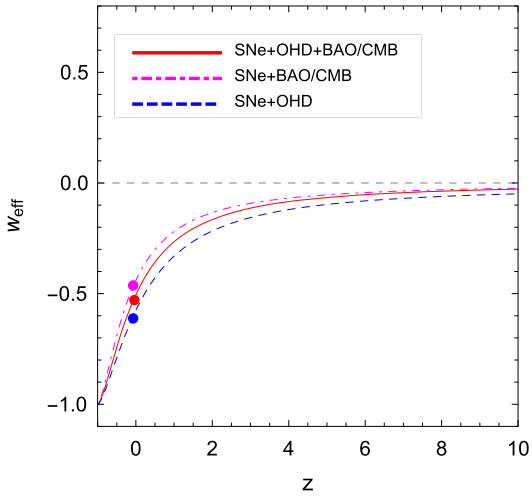


Fig. 6 Behavior of effective equation of state parameter $\omega_{\rm eff}$ with respect to the redshift z for best-fit values of parameters. A dot on each trajectory denotes the current value, $\omega_{eff}(z=0)$

In Fig. 5, we plot the evolution of the deceleration parameter q with respect to the redshift z for the best-fitted values of the parameters. From the figure, we can observe that the trajectories of q change in sign from positive to negative as the redshift decreases. This implies that the HDE model with matter creation shows transition from decelerated phase to accelerated phase. In Fig. 5, a dot on each trajectory denotes the current value, q_0 at $z=0$. The current values of q are given in Table 2 for different set of data, which lies in $-1 < q < 0$.

Using (28) we plot the trajectory for $\omega_{\rm eff}$ with respect to the redshift z as shown in Fig. 6. It is observed that as $z \to -1$, $\omega_{\rm eff} \to -1$, which means that the HDE model behaves as a de Sitter model in late-time. It is also observed

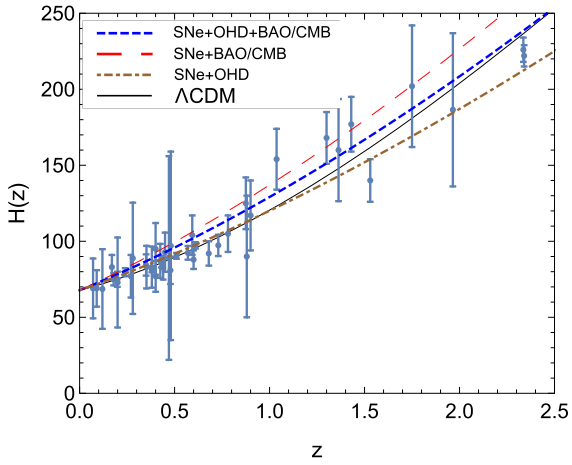


Fig. 7 The Hubble evolution and ΛCDM model with error bar plots from Hubble data

that the HDE model is free from big-rip singularity as each of the trajectories does not cross the phantom divide line. A dot on each trajectory in the figure denotes the current EoS parameter value, $\omega_{\rm eff}(z=0)$. The current values of $\omega_{\rm eff}$ for the set of combined data are listed in Table 2.

In Fig. 7, we plot the error bars of Hubble data set to compare our model with the ΛCDM model. The fitting achieved from our statistical analysis for combined data $SNe+OHD+BAO/CMB$ is more compatible with the Hubble data for the HDE model whereas the cosmic expansion differ appreciably in case of $SNe+OHD$ and $SNe+BAO/CMB$ with increasing z. The Hubble function is a monotonically increasing function of the redshift for all best-fit values of parameters.

Now, to calculate the age of the universe in terms of redshift using best-fit values of parameters given in Table 1, we use the definition given by $t(z)=T(z)/H_0$, where

$$T(z)=\int_z^\infty \frac{dz'}{(1+z')(H(z')/H_0)}. \tag{37}$$

For ΛCDM model, the age parameter is (Feng and Li 2009)

$$T(z)=\int_z^\infty \frac{dz'}{(1+z')[\Omega_{m0}(1+z')^3+(1-\Omega_{m0})]^{1/2}} \tag{38}$$

Using Eq. (19) in (37), we can calculate and plot the age of the universe with respect to the redshift z for the best-fit values as shown in Fig. 8. The current ages of the universe t_0 are given in Table 2 using different combined set of data. It is found that the age of the universe corresponding to combined $SNe+OHD+BAO/CMB$ data set is found to be $t_0=13.393\ Gyr$, which is comparatively same as predicted by ΛCDM model.

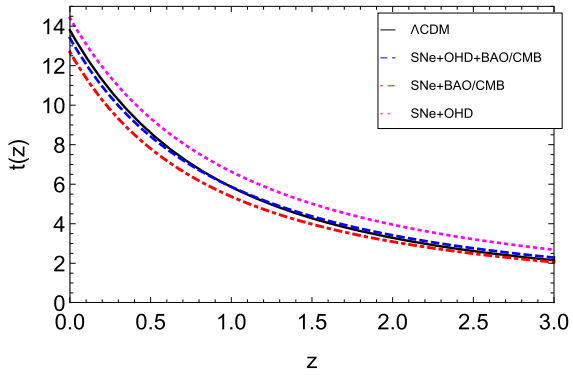


Fig. 8 The age of universe as a function of redshift for best fit values obtained from different combined observational data set. Black solid curve shows ΛCDM model, blue dashed curve for $SNe+OHD+BAO/CMB$, magenta dotted curve for $SNe+OHD$ whereas orange dashed -dotted curve shows $SNe+BAO/CMB$

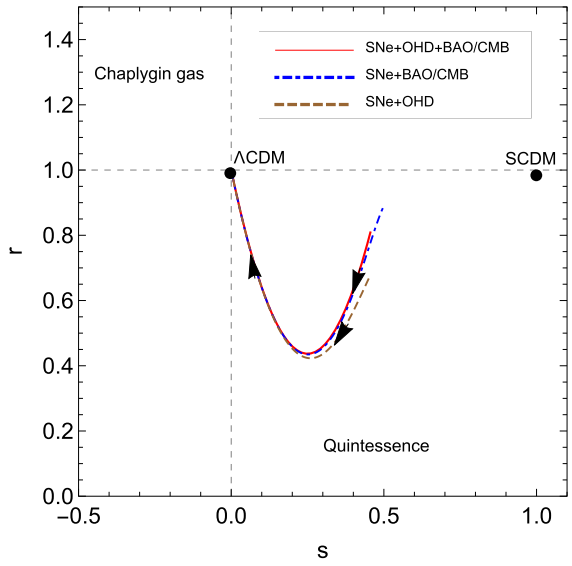


Fig. 9 The evolutionary trajectories of $\{r, s\}$ in the $s-r$ plane corresponding to the best-fit parameters of HDE model. The arrow shows the direction of the evolution of the trajectory. It can be observed that the evolution of $\{r, s\}$ are approximately coincide for the best estimates from $SNe+BAO/CMB$ $SNe+OHD$ and $SNe+OHD+BAO/CMB$ data set

5.2 Statefinder and Om(z) analysis

In this subsection, we present a comparative study of HDE model with other standard models of dark energy.

Sahni et al. (2003) and Alam et al. (2003) proposed two dimensionless and model independent parameters $\{r, s\}$, called statefinder parameters which are defined as,

$$r=\frac{1}{aH^3}\frac{d^3a}{dt^3} \text{ and } s=\frac{r-1}{3(q-1/2)} \tag{39}$$

The fixed points $\{r, s\}=\{1, 0\}$ represents the ΛCDM model of the universe whereas $\{r, s\}=\{1, 1\}$ represents the standard cold dark matter (SCDM) model. Using (19), these parameters are calculated to

$$r=1+\frac{9(\alpha(1-b^2)-(1+b^2\omega_d))\left(1-b^2\omega_d\right)}{4\,e^{\frac{3}{2}\alpha(1-b^2)H_0(t-t_0)}} + \frac{9(\alpha(1-b^2)-(1+b^2\omega_d))^2}{4\,e^{3\alpha(1-b^2)H_0(t-t_0)}} \tag{40}$$

$$s=\frac{\frac{(\alpha(1-b^2)-(1+b^2\omega_d))\left(1-b^2\omega_d\right)}{2\,e^{\frac{3}{2}\alpha(1-b^2)H_0(t-t_0)}}+\frac{(\alpha(1-b^2)-(1+b^2\omega_d))^2}{2\,e^{3\alpha(1-b^2)H_0(t-t_0)}}}{(1+b^2\omega_d-\alpha(1-b^2))e^{\frac{-3}{2}\alpha(1-b^2)H_0(t-t_0)}-1} \tag{41}$$

Equations (40) and (41) show that in the limit $a\to\infty$, the statefinder parameters $\{r, s\}\to\{1, 0\}$, a value corresponding to the ΛCDM model. Therefore, the HDE model resembles to ΛCDM model in future. The trajectory for $s-r$ plane for the best-fit parameters is shown in Fig. 9. From the figure it can be observed that in early time the trajectories lying in the region corresponding to $r<1, s>0$, which represents the quintessence region. The model is converging to the ΛCDM model in late time of its evolution.

We plot the trajectory of $\{r, q\}$ in $q-r$ plane for best-fit parameters obtained from different set of observational data as shown in Fig. 10. The SCDM model and steady state

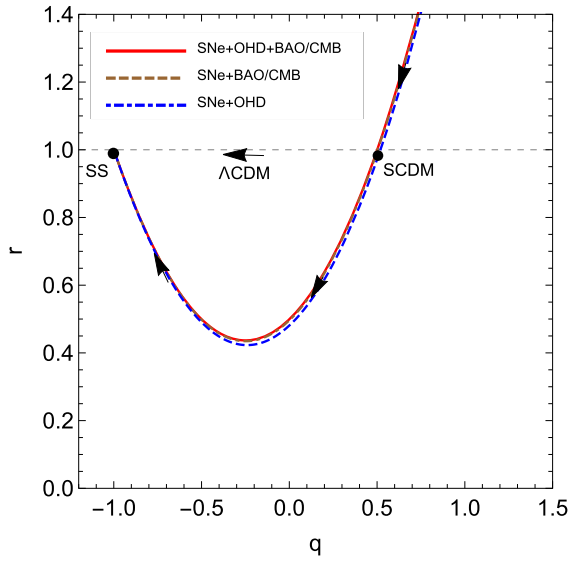


Fig. 10 The evolutionary trajectories of $\{r, q\}$ in the $q-r$ plane corresponding to the best-fitted parameters of HDE model. The arrow shows the direction of the evolution of the trajectory. The evolution of $\{r, s\}$ are approximately coincide for the best estimates from $SNe+BAO/CMB$ $SNe+OHD$ and $SNe+OHD+BAO/CMB$ data set

(SS) model correspond to fixed points $\{r, q\}=\{1, 0.5\}$ and $\{r, q\}=\{1, -1\}$, respectively. The horizontal line at $r=1$ corresponds to the ΛCDM model. We can observe that the trajectories go from positive value of q to negative value, which explains the transition from decelerated phase to accelerated one. The model approaches to the SS model in late

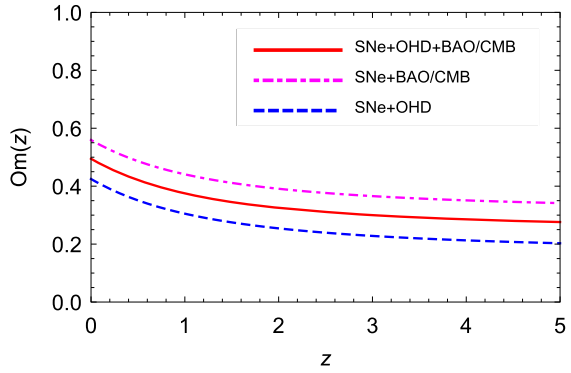


Fig. 11 The evolutionary trajectories of $Om - z$ for best-fit model parameters of HDE model

time. One more independent diagnostic parameter, called Om, was introduced by Sahni et al. (2008) to differentiate the dark energy models from ΛCDM. Om-diagnostic depends only on the first order derivative of the scale factor which makes it easier to reconstruct from observational data. We use Om to check different periods of the universe and defined it as

$$Om(z) = \frac{\frac{H^2(z)}{H_0^2} - 1}{(1+z)^3 - 1}. \tag{42}$$

Using (19) in (42), we obtain

$$Om(z) = \frac{\left[\frac{(1-b^2)\alpha}{1+b^2\omega_d} + \left(1 - \frac{(1-b^2)\alpha}{1+b^2\omega_d}\right)(1+z)^{\frac{3}{2}(1+b^2\omega_d)}\right]^2 - 1}{(1+z)^3 - 1}. \tag{43}$$

We plot the trajectories of Om-diagnostic against the redshift z in Fig. 11. The negative slope of the trajectories of $Om(z) - z$ plane indicates the behavior similar to the quintessence model.

5.3 Cosmography analysis

The geometrical investigation discussed in Sect. 5.2 on the dark energy model has been further extended by considering the Taylor series expansion of the scale factor about the present time. As a result some new model independent dimensionless geometrical parameters, called cosmography parameters (CP), have been defined as (Visser 2004, 2005)

$$j = \frac{1}{aH^3}\frac{d^3a}{dt^3}\ ,\ s = \frac{1}{aH^4}\frac{d^4a}{dt^4}\ ,\ l = \frac{1}{aH^5}\frac{d^5a}{dt^5}. \tag{44}$$

The cosmography parameters defined in (44) are called as $jerk$ (j), $snap$ (s), and $lerk$ (l) parameter, respectively. Here r and j are same; but the s parameter defined in (44) is not same with one defined in (39).

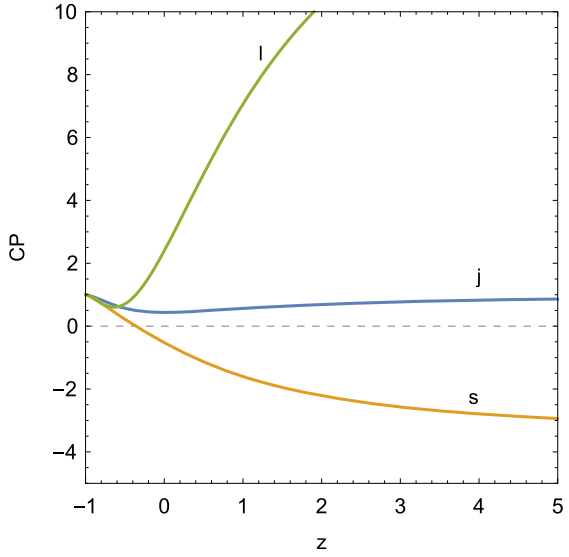


Fig. 12 The cosmographic parameters j, s and l are plotted using the combined data of $SNe + OHD + BAO/CMB$

We discuss the variations of the cosmography parameters with respect to redshift z by plotting their trajectories for combined data set of $SNe + OHD + BAO/CMB$ as shown in Fig. 12. From the figure, one can notice that during the entire evolution of the universe, the trajectories of j and l remains positive and for l it decreases sharply from high value at the early epoch. The trajectory of s shows that it is negative in early phase of the evolution and then gradually increases and becomes positive, hence it shows transition from negative values to positive values. In the late-time, j reaches to 1 as z approaches to -1 which corresponds to ΛCDM model. Similarly, other parameters s and l also approaches to 1 in the late-time of evolution. So, based on the cosmographic analysis, we can deduce that the HDE model is different from the ΛCDM model at the present time $z = 0$, but in the late-time it converges to the ΛCDM model.

5.4 Selection criteria

We assess the viability of our model against a given reference model. For this purpose we use two standard information criteria- Akaike Information Criterion (AIC) (Akaike 1974) and Bayesian or Schwarz Information Criterion (BIC) (Schwarz 1978). They are defined as follows:

$$AIC = -2\ln\mathcal{L}_{max} + 2d \tag{45}$$

and

$$BIC = -2\ln\mathcal{L}_{max} + d\ln N \tag{46}$$

where $\mathcal{L}_{max}$ denotes the maximum likelihood function, d number of model parameters and N total number of data

Table 3 Summary of the AIC and BIC values and their differences from the reference model of ΛCDM

Data	χ^2	χ^2_{red}	AIC	BIC	ΔAIC	ΔBIC
SOB	86.72	1.09	90.72	95.50	58.83	59.01
SB	34.45	0.91	38.45	41.72	16.16	16.15
SO	33.78	0.46	37.78	42.39	3.33	3.18

points used in observational analysis. For each model we can calculate AIC and BIC and check the viability by taking ΛCDM as the reference model, which is best choice for that. From Eq. (45) and (46), one can say that a model giving small AIC and a small BIC should be favored by the observations. To compare any model say k with the reference model say l, we compute $\Delta AIC_{kl} = AIC_k - AIC_l$ to come to the following conclusions: (i) For $0 \leq \Delta AIC_{kl} < 2$ then the model k has strong evidence in favour with respect to the reference model l, (ii) for $2 < \Delta AIC_{kl} < 4$, the model k has average evidence in support, (iii) for $4 < \Delta AIC_{kl} < 7$, the model k has considerably less support, and (iv) for $\Delta AIC_{kl} > 10$ implies that the model k has no support (Liddle 2007).

Similarly, we calculate $\Delta BIC_{ij} = BIC_i - BIC_j$ to find evidences against its viability where i is the candidate model and j is the reference model. The BIC analysis result to the following conclusions: (i) for $0 \leq \Delta BIC_{ij} < 2$, it is not worth more than a bare mention, (ii) $2 \leq \Delta BIC_{ij} < 6$, the positive evidence against the model i, and (iii) for $6 \leq \Delta BIC_{ij} < 10$, strong evidence against the model i (Liddle 2007).

We present the obtained results of AIC and BIC in the Table 3 with consideration of the ΛCDM as the referring model. According to the values of ΔAIC and ΔBIC, we can see that $2 < \Delta AIC < 4$ for the combined data of $SNe + OHD$, and hence the HDE model has average evidence in favour with respect to the reference model of ΛCDM cosmology corresponding to $SNe + OHD$.

Concerning ΔBIC corresponding to combined data of SNe and OHD, the HDE model has $2 < \Delta BIC < 6$ which means that it has positive evidence against it. For all other cases, the values of ΔAIC and ΔBIC are ruled out completely. The reduced $\chi^2_{red} = \chi^2_{min}/(N-d)$, where N is the total number of data and d is the degree of freedom, listed in Table 3 shows that the model provides a very good fit to these data.

6 Conclusion

In this paper, we have presented a HDE model with adiabatic matter creation process. We have considered the flat FRW universe filled with HDE and pressureless dark matter with matter creation. The exact solution with constant creation rate for scale factor has been obtained to discuss the evolution of the model. We have discussed the observational analysis by employing the MCMC package EMCEE on different combination of publicly available data sets of SNe, OHD and BAO/CMB. The best-fit values of model parameters have been obtained with different combinations of data sets. Figure 1 shows the contour plot of model parameters at 1σ and 2σ confidence level using the observational data SNe+OHD+BAO/CMB. The trajectories of evolution of the scale factor, matter density parameter, deceleration parameter, Hubble parameter have been plotted with best-fit values and discussed their evolutions. We have studied the diagnostic parameters, such as statefinder parameters and $Om(z)$ of HDE model with matter creation to distinguish it from the ΛCDM model. The cosmography analysis, like jerk, snap, etc., have been discussed to differentiate the HDE model with matter creation from other existing DE models. We summarize the results as below:

Figure 2 shows that the HDE model with matter creation starts its expansion with accelerated rate in early time followed by deceleration and then the accelerated expansion at the late time. Figure 3 gives the comparative evolution of the scale factor of different models. We have observed that the HDE model with matter creation and matter-dominated model with matter creation evolve on the same trajectory and expand with accelerated rate. However, the HDE and matter-dominated models without matter creation expand through out with decelerated rate. The effect of matter creation on matter density parameter has been shown in Fig. 4 for different models. In Fig. 5, we can observe that there is a transition from decelerated phase to accelerated phase due to sign change from positive to negative in q. Figure 6 shows the evolution of the effective EoS with redshift. It has been observed that as $z \to -1$, $\omega_{\text{eff}} \to -1$, which means that the HDE model with matter creation behaves as a de Sitter model in late-time. It has also been observed that the HDE model with matter creation is free from big-rip singularity as the trajectories of EoS do not cross the phantom divide line. The fitting achieved from our statistical analysis for combined data $SNe + OHD + BAO/CMB$ is compatible with the Hubble observational data for the HDE model with matter creation as shown in Fig. 7. The plot of the age of the universe with respect to the redshift z for the best-fit values models has been shown in Fig. 8. Using best-fit values of model parameters obtained from combined data set of $SNe + OHD + BAO/CMB$, it has observed that the age of the universe is found to be 13.393 Gyr which is very close to the observed value by concordance model. In our HDE model with matter creation, the trajectory for the model deviates very slightly from that of ΛCDM. It indicates that the HDE model with matter creation could really alleviate the age problem. This is due to the change in the energy conservation equation by addition of matter creation which makes

the matter dilution little bit slower, and then the age problem is alleviated.

The statefinder analysis shows that the trajectories of HDE model lies in the region corresponding to $r < 1, s > 0$, which represents the quintessence region. The model converges to the ΛCDM model in late time of its evolution. It can also be found from the $Om(z)$, cosmography analyses. We have also performed the information criterion of AIC and BIC to discriminate HDE model with ΛCDM model and presented in Table 3. The analysis of AIC indicates that there is average support for the HDE model with matter creation when compared to the ΛCDM model, while those based on the BIC indicates that there is positive evidence against the model with respect to the ΛCDM model when we use the combined dataset of $SNe + OHD$, while these selection criteria disfavour with other set of data discussed here. The reduced χ^2_{red} listed in Table 3 shows that the model provides a very good fit to these data. In conclusion, the present HDE model with matter creation successfully describes the expansion history of evolution of the universe from early decelerated phase to late time accelerated phase.

Acknowledgements The data fitting is based on the publicly available EMCEE package, a Markov Chain Monte Carlo code. The authors express their sincere thanks to the anonymous reviewer for constructive comments/suggestions which helped us to improve the original manuscript significantly. A. K. would like to thank University Grant Commission, New Delhi for providing Senior Research Fellowship under UGC-NET Scholarship.

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

References

Abramo, L.R.W., Lima, J.A.S.: Class. Quantum Gravity **13**, 2953 (1996)
Ade, P.A.R., et al.: Astron. Astrophys. **571**, A16 (2014)
Ade, P.A.R., et al.: Astron. Astrophys. **594**, 13 (2016)
Akaike, H.: IEEE Trans. Autom. Control **19**, 716 (1974)
Alam, U., et al.: Mon. Not. R. Astron. Soc. **344**, 1057 (2003)
Alcaniz, J.S., Lima, J.A.S.: Astron. Astrophys. **349**, 729 (1999)
Anderson, L., et al.: Mon. Not. R. Astron. Soc. **441**, 24 (2014)
Betoule, M., et al.: Astron. Astrophys. **568**, A22 (2014)
Beutler, F., et al.: Mon. Not. R. Astron. Soc. **416**, 3017 (2011)
Blake, C., et al.: Mon. Not. R. Astron. Soc. **425**, 405 (2012)
Calvao, M.O., Lima, J.A.S., Waga, I.: Phys. Lett. A **162**, 233 (1992)
Cao, S.L., et al.: Eur. Phys. J. C **78**, 170 (2018)
Cardenas, V.H.: Eur. Phys. J. C **72**, 2149 (2012)
Chen, B., et al.: Nucl. Phys. B **774**, 256 (2007)
Cohen, A., Kaplan, D., Nelson, A.: Phys. Rev. Lett. **82**, 4971 (1999)
de Haro, J., Pan, S.: Class. Quantum Gravity **33**, 165007 (2016)
dos Santos, M.V., Reis, R.R.R., Waga, I.: J. Cosmol. Astropart. Phys. **02**, 066 (2016)
Feng, C.J., Li, X.Z.: Phys. Lett. B **680**, 355 (2009)
Feng, C., et al.: J. Cosmol. Astropart. Phys. **0709**, 005 (2007)
Foreman-Mackey, D., et al.: Publ. Astron. Soc. Pac. **125**, 306 (2013)
Gunzig, E., Gehenian, J., Prigogine, I.: Nature **330**, 621 (1983)
Gunzig, E., Maartens, R., Nesteruk, A.V.: Class. Quantum Gravity **15**, 923 (1998)
Hinshaw, G., et al.: Astrophys. J. Suppl. Ser. **208**, 19 (2013)
Hsu, S.D.H.: Phys. Lett. B **594**, 13 (2004)
Huang, Q.G., Li, M.: J. Cosmol. Astropart. Phys. **08**, 013 (2004)
Komatsu, E., et al.: Astrophys. J. **192**, 18 (2011)
Li, M.: Phys. Lett. B **603**, 1 (2004)
Li, M., Li, X.D., Wang, S., Zhang, X.: J. Cosmol. Astropart. Phys. **0906**, 036 (2009)
Li, M., et al.: J. Cosmol. Astropart. Phys. **09**, 021 (2013)
Liddle, A.R.: Mon. Not. R. Astron. Soc. **377**, L74 (2007)
Lima, J.A.S., Germano, A.S.: Phys. Lett. A **170**, 373 (1992)
Lima, J.A.S., Calvao, M.O., Waga, I.: Cosmology, Thermodynamics and Matter Creation, vol. 317. World Scientific, Singapore (1991)
Lima, J.A.S., Germano, A.S., Abramo, L.R.W.: Phys. Rev. D **53**, 4287 (1996)
Lima, J.A.S., Silva, F.E., Santos, R.C.: Class. Quantum Gravity **25**, 205006 (2008)
Lima, J.A.S., et al.: J. Cosmol. Astropart. Phys. **10**, 042 (2014)
Luongo, O.: Adv. High Energy Phys. **2017**, 1424503 (2017)
Malekjani, M., et al.: Phys. Rev. D **98**, 063533 (2018)
Nunes, R.C., Pavon, D.: Phys. Rev. D **91**, 063526 (2015)
Padmanabhan, T.: Phys. Rep. **380**, 235 (2003)
Padmanabhan, N., et al.: Mon. Not. R. Astron. Soc. **427**, 2132 (2012)
Pan, S., Chakraborty, S.: Adv. High Energy Phys. **201**, 654025 (2015)
Pan, S., et al.: Mon. Not. R. Astron. Soc. **460**, 1445 (2016)
Pan, S., et al.: Int. J. Geom. Methods Mod. Phys. **15**, 1850042 (2018)
Peebles, P.J.E., Ratra, B.: Rev. Mod. Phys. **75**, 559 (2003)
Perlmutter, S., et al.: Astrophys. J. **517**, 565 (1999)
Prigogine, I., et al.: Proc. Natl. Acad. Sci. USA **85**, 7428 (1988)
Prigogine, I., et al.: Gen. Relativ. Gravit. **21**, 8 (1989)
Ramos, R.O., et al.: Phys. Rev. D **89**, 083524 (2014)
Riess, A.G., et al.: Astron. J. **116**, 1009 (1998)
Sahni, V., Starobinsky, A.A.: Int. J. Mod. Phys. **9**, 373 (2000)
Sahni, V., et al.: JETP Lett. **77**, 201 (2003)
Sahni, V., Shaleloo, A., Starobinsky, A.A.: Phys. Rev. D **78**, 103502 (2008)
Salzano, V., et al.: Mod. Phys. Lett. A **29**, 145008 (2014)
Sanchez, A.G., et al.: Mon. Not. R. Astron. Soc. **425**, 415 (2012)
Schwarz, G.: Ann. Stat. **6**, 461 (1978)
Singh, C.P., Kaur, S.: Phys. Rev. D **100**, 084057 (2019)
Steigman, G., Santos, R.C., Lima, J.A.S.: J. Cosmol. Astropart. Phys. **06**, 033 (2009)
Susskind, L.: J. Math. Phys. **36**, 6377 (1995)
't Hooft, G.: (1993). arXiv:gr-qc/9310026
Visser, M.: Class. Quantum Gravity **21**, 2603 (2004)
Visser, M.: Class. Quantum Gravity **37**, 1541 (2005)
Weinberg, S.: Rev. Mod. Phys. **61**, 1 (1989)
Zhang, X., Wu, F.Q.: Phys. Rev. D **72**, 043524 (2005)

Sādhanā (2020)45:124
https://doi.org/10.1007/s12046-020-01368-4

© Indian Academy of Sciences

Optimization of machining parameters during cryogenic turning of AISI D3 steel

ANURAG SHARMA, R C SINGH and RANGANATH M SINGARI*

Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India
e-mail: ranganath@dce.ac.in

MS received 6 February 2019; revised 27 January 2020; accepted 25 February 2020

Abstract. This research paper depicts the process of used liquid nitrogen at the interface of TiN coated carbide cutting tool insert (rake face) and AISI D3 workpiece. Design of experiments (DoE) was planned according to Taguchi L_9 (OA) orthogonal array. The experimental results during machining such as cutting force, machining time and temperature were optimized by Taguchi S/N ratio and analysed by ANOVA. The contribution of machining parameters of (i) speed, (ii) feed and (iii) depth of cut for each response were evaluated. Feed had the highest effect on the percentage of contribution of 57.21% and 52.21% for cutting force and machining time, respectively. Speed had the highest effect on the contribution as 79.57% for the temperature at the interface of insert and workpiece. The predicted values at the optimum level of machining parameters for cutting force, machining time and temperature were 44.49 N, 37.09 sec. and 24.99°C, respectively. Regression models were made. The R-Sq values were 96.59, 89.34 and 96.09% for cutting force, machining time and temperature, respectively. The ratio of an average thickness of generated chip and feed was considered as the chip compression ratio. It was observed that the generated chips during cryogenic turning were thin, discontinuous, long snarled and most of the material had side flow on either side.

Keywords. Cryogenic turning; liquid nitrogen; optimization; ANOVA; Taguchi S/N ratio; AISI D3 steel.

1. Introduction

The modern competitive society is concerned with the global environment to affect, new engineering materials and the fabrication processes involve in manufacturing of the component. There is a challenge for the scientists and researchers to fulfil intrinsic international norms up to the mark for the fabrication. Machining is a common manufacturing process to remove the material from the parent material to have required finished shape and size of the object. The conventional cutting fluid is not eco-friendly during machining operations to get the required finished product [1–4]. Dry machining (without lubricant) could be considered as eco-friendly, but the machining characteristics such as surface roughness, cutting force, temperature, coefficient of friction, tool wear, etc. were almost satisfactory at low machining parameters. This slowed down the manufacturing rate [5–8]. The need for eco- friendly cutting fluid was generated to support machining operations with satisfactory results at low, medium to high levels of machining parameters. Liquid nitrogen could be used during machining. Liquid nitrogen is colourless, tasteless, odourless, non-toxic and having a boiling point of −196°C [9–12]. Health-related issues like breathing, nausea, infection to hands, allergies, etc. for operator, were declined or almost became negligible with the utilisation of LN_2 during machining. Low adhesion and abrasion wear mechanism was found on the cutting tool. Low debris was found on the machined surface. LN_2 improved tool life and surface properties of the workpiece as compared to dry machining [13, 14].

1.1 *Cryogenic cooling*

Liquid nitrogen was used by researchers as direct regulated supply to the interface of cutting tool and workpiece during machining operations. Manimaran *et al* [15, 16] used liquid nitrogen supply at the interface of workpiece (steel) and grinding wheel. It was found that during cryogenic cooling, grinding forces were declined by 32–36% and 13–26% as compared to dry and wet grinding, respectively. Surface roughness was declined by 26–59% and 32–43% as compared to dry and wet grinding. Dhanachnezian *et al* [17] investigated cryogenic turning with a direct regular supply of LN_2 at flank and rake parts of single point cutting insert and found that cutting temperature, cutting force, surface roughness and tool wear were declined by 61–66%, 35–42%, 35% and 39%, respectively during cryogenic turning as compared with flood supply. Sartori *et al* [18]

*For correspondence
Published online: 14 May 2020

Table 1. Chemical analysis of $AISID_3$ steel.

C	Si	Mn	S	P	Cr	Ni	Mo	Co	Nb	V	W	Fe
2.02	0.259	0.430	0.029	0.020	11.00	0.074	<0.10	0.011	0.023	0.038	0.087	85.909

Table 2. Chemical analysis of cutting inserts.

C	Co	Mn	V	Nb	Ni	W	Mo	Ti
7.10	16.25	0.30	2.05	5.23	4.95	15.88	0.69	47.55

studied the tool wear and surface roughness during dry, wet, MQL, CO_2 (rake) + MQL (flank), LN_2, LN_2 (rake) + MQL(flank) and found that surface roughness was low with LN_2. Hong *et al* [19–21] found that friction, cutting force, cutting temperature were low and tool life improved during cryogenic machining as compared to dry machining of titanium alloy. Dhar *et al* [22–25] investigated the influence of cryogenic cooling on steel alloys AISI 1040, E4340C, AISI 4140, AISI 4037 and concluded that the tool wear was low and surface roughness was better as compared to wet and dry machining. The surface produced during cryogenic machining was clean and had a negligible amount of debris. The machined surface had low cracks. The surface profile graph was uniform and within a small range. The chips produced were discontinuous, thin and had clean surface [26–29]. Sun *et al* [30] used LN_2 for cooling down the air and further used chilled air at the interface of the cutting tool and workpiece and found improved surface roughness as compared to dry machining. Chip thickness declined by 9% at cutting speed lower than 50 m/minute. Mia *et al* [31] compared machining characteristics in dry, LN_2 supply at rake face named as single LN_2 jet and liquid nitrogen supply at rake and flank face named as duplex jets and found that surface roughness, cutting force and cutting temperature were low in duplex supply of LN_2.

The hardness of cutting inserts was increased by cryogenic treatment using liquid nitrogen. Cutting inserts were placed in cryoprocessor. The temperature declined slowly and gradually at approximate 2°C per hour in it. After reaching −196°C, cutting inserts were kept for 12–48 hours. The temperature was increased slowly and gradually to reach room temperature. Tempering was conducted for 2 hours to relieve residual stresses. Grain refinement of tool material took place during the treatment. This might increase hardness and wear resistant property of the material. The number of hours was decided according to the hardness required in the material. The number of hours could be varied from 4 h to 24 h according to the nature of material and extent of hardness was required. Cryogenically treated cutting inserts coated and uncoated could perform better than untreated coated and uncoated cutting inserts for titanium alloys, Nickel-based alloys and AISI D3 alloys and TiN coating with cryogenic treatment gave 20–45% extra tool life [32, 33]. Lal *et al* [34] found that the wear resistance of cryogenically treated cutting inserts was improved by 95.5%.

1.2 *Optimization*

Machining process consisted of various cutting parameters for a definite combination of cutting tool and workpiece. To save energy and decrease in economic aspects it was necessary to know and apply optimum cutting parameters for better surface integrity, lower cutting forces, lower tool wear and lower energy consumption [35]. Optimization was achieved by using multiple desirability function and RSM models to achieve low surface roughness, cutting force, power consumption and high tool life during cryogenic turning with LN_2 [36]. Taguchi technique was discovered in 1950 by Genichi Taguchi. It was based on the signal-to-noise (S/N) ratio. It minimized the practice of hit and trial for selecting the best combination of controlled factors (cutting parameters in machining) for the better conditional output of machining properties [37]. In dry turning of hardened steel alloys AISI D3, AISI4340, AISI H13, EN19, high- alloy white cast iron and magnesium alloy the surface roughness, tool wear, cutting force and cutting temperature were optimized by Taguchi approach [38–50]. The machinability characteristics during cryogenic cooling with LN_2 in turning, milling and grinding were better than dry, conventional cutting fluid, minimum quantity lubrication technique (MQL) with castor oil and frozen CO_2 [51–55].

The review of research papers showed that steel AISI D3 (HRC 60) is used in manufacturing blanking and forming dies, forming tools, press tools, punches, bushes and wear resistant moulds. It is categorised as difficult to machine materials. The combination of TiN coated carbide cutting tool and AISI D3 as workpiece was rarely used during cryogenic turning with LN_2 for the analysis of machinability characteristics. This explored a new dimension for trying experimentation with other noble gases such as frozen CO_2, liquid H_2, liquid Helium, etc. and a hybrid of gases with eco-friendly oils like castor, olive, etc. There is a need for experimental investigation of the parameters in LN_2 environment during turning of AISID3 workpiece and TiN coated carbide cutting insert. To apply the modern tools as Taguchi and ANOVA techniques for human beings and to fulfil the current requirement up to the international standards.

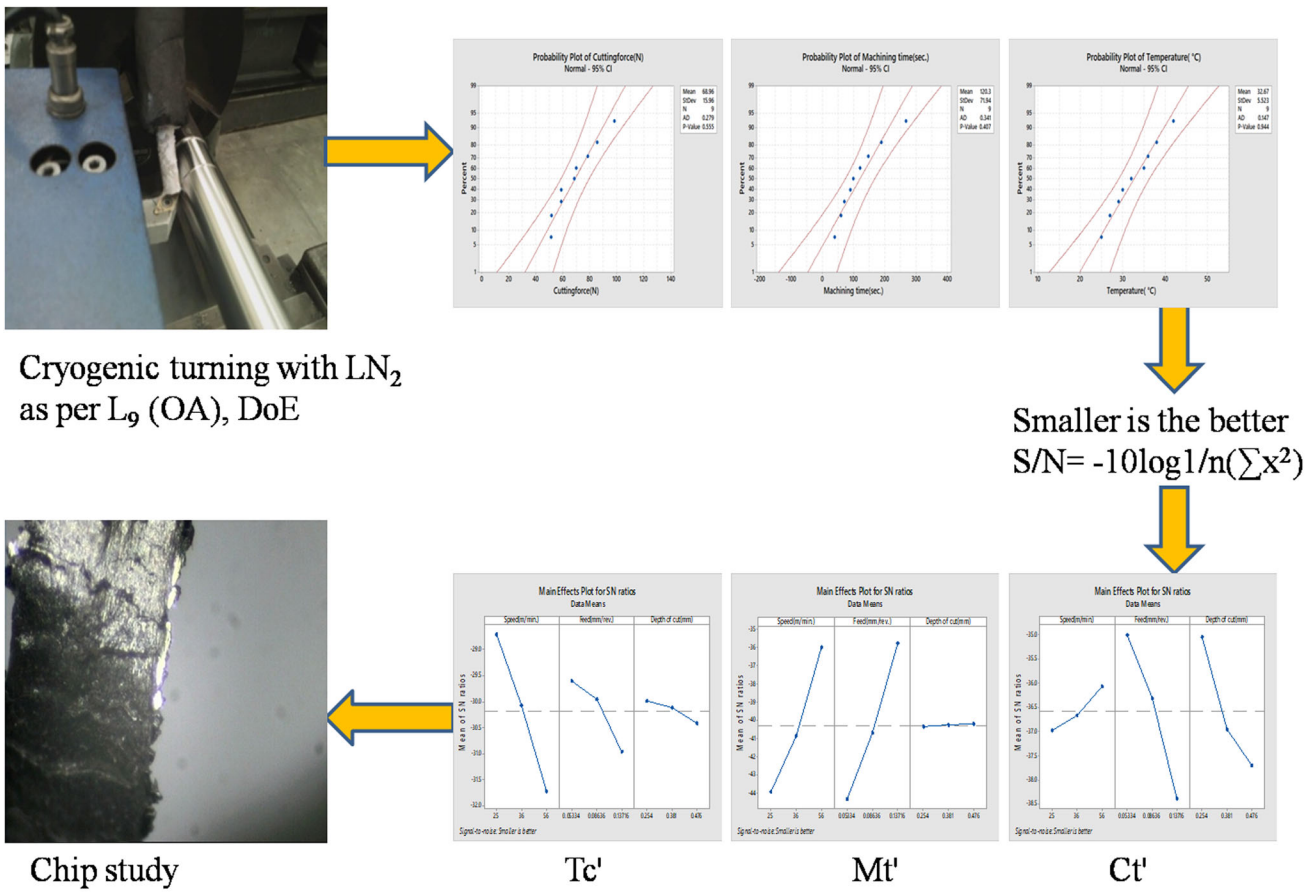


Figure 1. Workflow diagram during turning with LN_2.

Table 3. Machining parameters with different levels.

Levels	Machining parameters		
	Vd′ (m/min.)	Fd′ (mm/rev.)	Dt′ (mm)
Level 1	25	0.05334	0.254
Level 2	36	0.08636	0.381
Level 3	56	0.13716	0.476

2. Experimental procedure

2.1 *Materials and methods*

AISI D3 was used as the material for the workpiece. The shape was cylindrical. The diameter was 28 mm and length was 300 mm. The sample of the material was chemically tested for checking the composition of elements. The percentage of each element in weight was measured and depicted in table 1.

Cutting inserts were used in the shape of a diamond. The material was tungsten carbide with a coating of titanium nitride. ISO specification of cutting insert was DCMT 11T 3087 HQ with a grade of PV20 and tool holder was SDJCR 1212F 11. Tool geometry as per orthogonal rake system was 0°, 0°, 7°, 7°, 60°, 93° and 0.8 mm (λ′, α′, β′, γ′, φ′, θ′ and r′), respectively. The sample of cutting tool insert was examined chemically for checking for the composition of elements available. The percentage of each element in weight was measured and depicted in table 2.

2.2 *Machining set-up and methods*

The experiments were performed on the three-jaw lathe machine. Dewar container with the capacity of 50 kg was used as the storage of liquid nitrogen. The insulated pipe was fitted to container and another free end was fitted at the interface of cutting insert and workpiece. The regulated compressed air supply was maintained by a pressure regulator and pressure relief valve fitted between pipe connected to container and compressor. The temperature was measured by IR thermal image non-contacting thermometer with an accuracy of ±0.05°C of reading. The beam was focused at the interface of cutting insert (rake face) and workpiece. This was done with constant incident angle and distance by using a fixture near the lathe machine. The

Table 4. Taguchi L_9 (OA).

Run	Coded Values of factors			Un-coded values of factors		
	Vd′ (m/min.)	Fd′ (mm/rev.)	Dt′ (mm)	Vd′ (m/min.)	Fd′ (mm/rev.)	Dt′ (mm)
1	1	1	1	25	0.05334	0.254
2	1	2	2	25	0.08636	0.381
3	1	3	3	25	0.13716	0.476
4	2	1	2	36	0.05334	0.381
5	2	2	3	36	0.08636	0.476
6	2	3	1	36	0.13716	0.254
7	3	1	3	56	0.05334	0.476
8	3	2	1	56	0.08636	0.254
9	3	3	2	56	0.13716	0.381

Table 5. Experimental results of responses and respective S/N ratio.

Run	Responses			S/N ratio of responses		
	Ct′	Mt′	Tc′	Ct′	Mt′	Tc′
1	51.5	268.14	25	−34.2361	−48.5672	−27.9588
2	69.8	147.35	27	−36.8771	−43.3670	−28.6273
3	98.2	99.30	30	−39.8422	−39.9390	−29.5424
4	58.8	188.64	29	−35.3875	−45.5127	−29.2480
5	78.4	121.12	32	−37.8863	−41.6643	−30.1030
6	68.6	59.35	35	−36.7265	−35.4684	−30.8814
7	58.8	89.03	38	−35.3875	−38.9907	−31.5957
8	51.2	70.88	36	−34.1854	−37.0105	−31.1261
9	85.3	39.29	42	−38.6190	−31.8856	−32.4650

cutting force was measured by piezo-electric three component dynamometer. The machining time was measured by a dedicated digital stopwatch. Initially, the stopwatch was set at zero reading and started to record time with the start of an experimental run. Stopped with the end of the experimental run. The time shown on stopwatch was noted down. After the completion of each experimental run, worn out used cutting insert was replaced by a new unused cutting insert. Chips were collected for each set of experiment. The chip thickness was measured by optical CNC vision inspection microscope at ten different places. The average value was calculated for getting the final chip thickness value. Figure 1 shows the sequences of stages followed during the experimentation process. Experiments were performed during cryogenic turning with Taguchi based L_9 (OA), DoE. Probability plots were formed for the values obtained for each response (cutting force, machining time and temperature). After checking for acceptance of the null hypothesis with a normal probability distribution, the values were further used for analysis. Mean effect plots were formed. Chips were optically examined for understanding the morphology and compression ratio.

2.3 *Design of experiments (DoE)*

The experiments were performed with three machining parameters such as speed, feed and depth of cut. Each

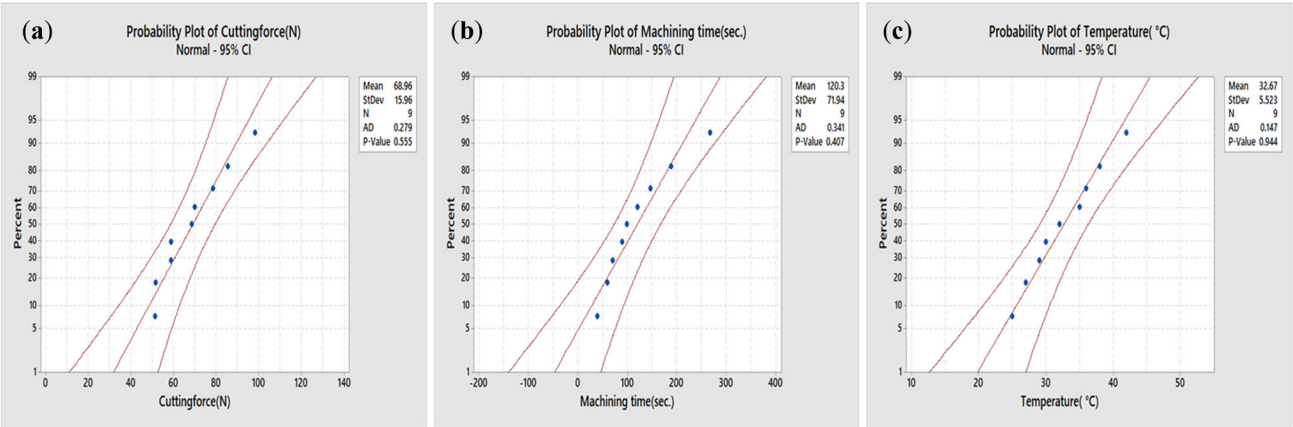


Figure 2. Probability plots for **(a)** cutting force, **(b)** machining time and **(c)** temperature.

Table 6. Response table for mean S/N ratios for (Ct′) cutting force, (Mt′) machining time and (Tc′) temperature.

Response	Level	Vd′ (m/min.)	Fd′ (mm/rev.)	Dt′ (mm)
Ct′	1	−36.99	**−35.00**	**−35.05**
	2	36.67	−36.32	−36.96
	3	**−36.06**	−38.40	−37.71
	Delta	0.92	3.39	2.66
	Rank	3	1	2
Mt′	1	−43.96	−44.36	−40.35
	2	−40.88	−40.68	−40.26
	3	**−35.96**	**−35.76**	**−40.20**
	Delta	8.00	8.59	0.15
	Rank	2	1	3
Tc′	1	**−28.71**	**−29.60**	**−29.99**
	2	−30.08	29.95	−30.11
	3	−31.73	−30.96	−30.41
	Delta	3.02	1.36	0.42
	Rank	1	2	3

Table 7. Response table for means for (Ct′) cutting force, (Mt′) machining time and (Tc′) temperature.

Response	Level	Vd′	Fd′	Dt′
Ct′	1	73.17	**56.37**	**57.10**
	2	68.60	66.47	71.30
	3	**65.10**	84.03	78.47
	Delta	8.07	27.67	21.37
	Rank	3	1	2
Mt′	1	171.60	181.94	132.79
	2	123.04	113.12	125.09
	3	**66.40**	**65.98**	**103.15**
	Delta	105.20	115.96	29.64
	Rank	2	1	3
Tc′	1	**27.33**	**30.67**	**32.00**
	2	32.00	31.67	32.67
	3	38.67	35.67	33.33
	Delta	11.33	5.00	1.33
	Rank	1	2	3

parameter was varied to three levels (Level = 1, 2 and 3). The full factorial design needed $3^3 = 27$ experiments for L_9 (OA) orthogonal array. But, according to Taguchi, the total number of experiments reduced to 9. This saved more than 50% of experiments. Table 3 shows the machining parameters with variations. Speed, feed and depth of cut were varied from (25–56 m/min.), (0.05334–0.13716 mm/rev.) and (0.254–0.476 mm), respectively. Table 4 shows coded values of factors with equivalent un-coded factors for the design of experiments. Level 1, 2 and 3 show minimum, medium and maximum value of each machining parameter such as speed, feed and depth of cut.

3. Experimental results and discussion

3.1 *Proability analysis*

Experiments were performed according to Taguchi based S/N ratio L_9 (OA) of DoE. The experimental results of responses with respective values of S/N ratio are shown in table 5. Probability plots were made at 95% confidence level. Each plot showed mean, StDev., AD and P-value. The acceptance or rejection of the null hypothesis with the normal probability distribution was dependent on P-value [44]. Figures 2(a), (b) and (c) showed that response values were approximately inclined to the middle straight line, P-value was greater than 0.01 and AD statistic was low. The experimental data could be used for further optimization, experimental investigation and interpretation.

3.2 *Taguchi based optimization (Signal to Noise ratio)*

Optimization based on Taguchi (S/N) ratio involves a decrease in variability and shifting towards target value. In any process, variability may arise due to factors having no control are termed as uncontrollable factor or noise. The expected value of the response is termed as target or signal. S/N is the ratio of expected target values to unexpected noise [40]. Eq. (1) shows smaller is the better characteristic [44, 58]

$$\frac{S}{N} = -10log\frac{1}{n}\left(\sum x^2\right) \tag{1}$$

Where, x represents the measuring responses (Ct′, Mt′, Tc′). Table 4 shows the coded and un-coded values of factors according to L_9 (OA) orthogonal array (Design of experiments). Table 5 shows the experimental results of responses with the respective S/N ratio. Table 6 shows the description for the mean S/N ratio for each response. Delta was the difference between the maximum and minimum value of each level. On the basis of delta value highest to lowest (rank) was given. This showed a priority given to each control factor. Optimum level value of the factor was evaluated by considering the highest S/N ratio value of a particular factor. Table 7 shows a description for mean values for each response. Optimum level value of the factor was evaluated by considering the lowest value of a particular factor. This was supported by the statistical analysis of variance (ANOVA). Table 8 shows ANOVA which had a degree of freedom (df), adjoint sum of squares (AdjSS), adjoint mean of square (Adj MS), F-Value, P-Value and percentage of contribution. The P-value was defined for the significance level of 5% (confidence level of 95%) for all responses [44, 57, 58].

Table 8. Analysis of variance for means of (Ct′) cutting force, (Mt′) machining time and (Tc′) temperature.

Response	Source	DF	Adj SS	Adj MS	F-Value	P-Value	% Cont.
Ct′	Vd′	2	1.3133	0.6567	2.36	0.297	4.28
	Fd′	2	17.5543	8.7771	31.58	0.031	57.21
	Dt′	2	11.2635	5.6317	20.26	0.047	36.70
	Error	20	0.5558	0.2779	–	–	1.81
	Total	26	30.6869	–	–	–	–
Mt′	Vd′	2	97.590	48.7952	21.94	0.044	45.96
	Fd′	2	111.516	55.7581	25.08	0.038	52.21
	Dt′	2	0.035	0.0174	0.01	0.992	0.02%
	Error	20	4.447	2.2236	–	–	2.08
	Total	26	213.589	–	–	–	–
Tc′	Vd′	2	13.7154	6.8577	58.31	0.0.017	79.57
	Fd′	2	3.0005	1.5003	12.76	0.073	17.41
	Dt′	2	0.2863	0.1432	1.22	0.451	1.66
	Error	20	0.2352	0.1176	–	–	1.36
	Total	26	17.2374	–	–	–	–

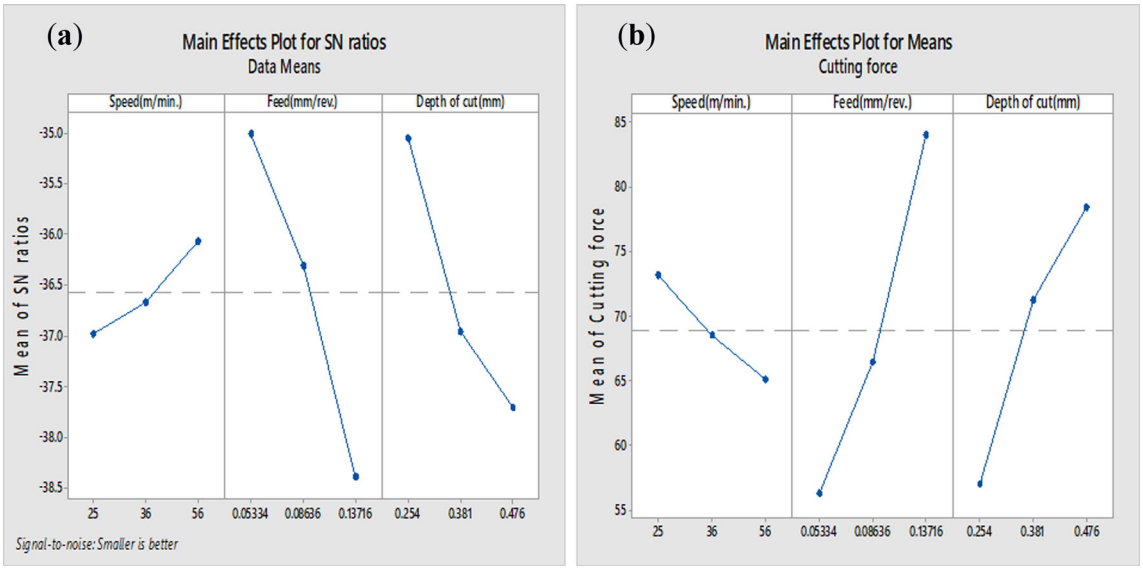


Figure 3. **(a)** Variation of mean S/N ratio Ct′. **(b)** Variation of mean Ct′ with different actor levels.

3.3 *Cutting Force*

Figure 3(a) shows mean effects plot for the mean of S/N ratios of cutting force during cryogenic turning with different levels of machining parameters. Optimum level value of cutting force Ct′ at Vd′ (level 3 = 56 m/min.), Fd′ (level 1 = 0.05334 mm/rev.) and Dt′ (level 1 = 0.254 mm).The predicted value of the response was calculated from Eqs. (2) and (3). d_p was the average S/N ratios of all variables, d_p was predicted response, $\bar{N}_o$ was the average S/N ratio when variable Vd′ (speed) was at optimum level, $\bar{F}_o$ was the average S/N ratio when variable Fd′ (feed) was at optimum level and $\bar{D}_o$ was the average S/N ratio when variable Dt′ (depth of cut) was at optimum level, Ap was predicted responses. Predicted value of cutting force was 44.49 N. (ANOVA) table 8 shows for cutting force, Fd′ had the highest percentage contribution of 57.32% followed by Dt′ and Vd′ in consecutive decreasing order of percentage contribution. F-value showed the relative importance firstly Fd′, secondly Dt′ and lastly was Vd′. P- value was significant at the value of level equal to or less than 0.05. P-value was significant for Fd′ and Dt′. Figure 3(b) shows the variation of mean Ct′ with different factor levels (individual machining parameters). On incrementing cutting speed (Vd′) cutting force declined and on incrementing feed (Fd′) and depth of cut (Dt′) cutting force increased.

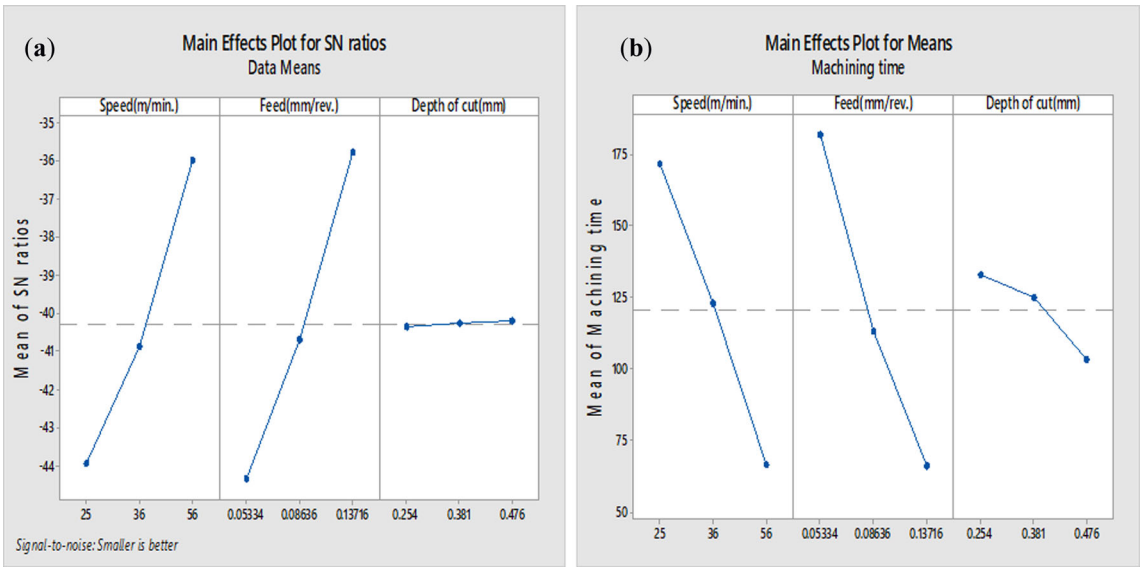


Figure 4. **(a)** Variation of mean S/N ratio Mt′. **(b)** Variation of mean Mt′ with different factor levels.

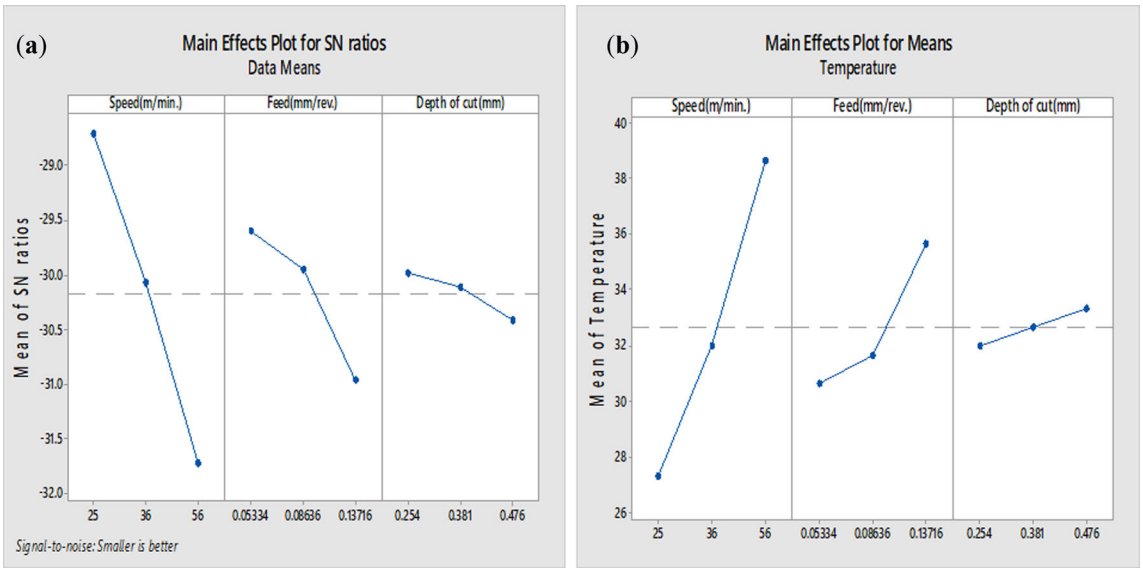


Figure 5. **(a)** Variation of mean S/N ratio Tc′. **(b)** Variation of mean Tc′ with different factor levels.

Table 9. Confirmation test results.

Response	Predicted value of response	Confidence Level(CI)	Actual value of response at experimental (optimized level)
Ct′	44.49 N	±0.85 N	45 N
Mt′	37.09 sec.	±2.39 sec.	37 sec.
Tc′	24.99°C	±0.55°C	25°C

Table 10. Regression models and R-Sq values for the experimental results.

	Regression Models	R-Sq%
Ct′	12.16 − 0.2499 Speed (m/min.) + 331.4 Feed (mm/rev.) + 97.1 Depth of cut (mm)	96.59
Mt′	422.1 − 3.125 Speed (m/min.) − 1345 Feed (mm/rev.) − 130 Depth of cut (mm)	89.34
Tc′	10.70 + 0.3617 Speed (m/min.) + 61.3 Feed (mm/rev.) − 5.96 Depth of cut (mm)	96.99

$$d_p = \overline{d_p} + \left(\overline{N_o} - \bar{d}_p\right) + \left(\overline{F_o} - \bar{d}_p\right) + \left(\overline{D_o} - \bar{d}_p\right) \quad (2)$$

$$A_p = 10^{-d_p/20} \quad (3)$$

3.4 *Machining time*

Machining time was measured in seconds as the time taken by tool travelling from the start of cut to end of cut in each experiment. Figure 4(a) shows mean effects plot for mean of S/N ratios for machining time during cryogenic turning with different levels of cutting parameters. Table 6 shows the optimum level of Mt′ was at Vd′ (level 3 = 56 m/min.), Fd′(level 3 = 0.13716 mm/rev.) and Dt′(level 3 = 0.476 mm). Predicated value of machining time was calculated from Eqs. (2) & (3) as 37.09 seconds. ANOVA table 8 shows Fd′ had the highest effect on the percentage of contribution (52.2%) followed by Vd′ and Dt′ in consecutive decreasing order of percentage contribution.

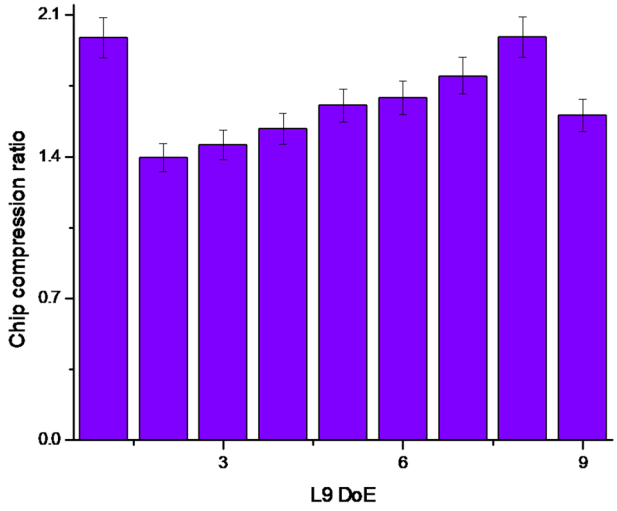


Figure 6. Chip compression according to experiments L_9 DoE.

F-value showed the relative importance firstly Fd′, secondly Vd′ and lastly was Dt′. P-value was significant for Vd′ and Fd′. Figure 4(b) shows the variation of mean Mt′ with different factor levels (cutting parameters). On incrementing cutting speed (Vd′), feed (Fd′) and depth of cut (Dt′) machining time declined.

3.5 *Temperature*

Temperature at the interface of tool and workpiece influences surface integrity of the machined component. This affects service providing the life period of the machined component. Figure 5(a) shows mean effects plot for the mean of S/N ratios for temperature during cryogenic turning with different levels of cutting parameters. Table 6 shows the optimum level of Tc′ was at Vd′ (level 1 = 25 m/min.), Fd′(level 1 = 0.05334 mm/rev.) and Dt′ (level 1 = 0.254 mm). Predicated value of temperature was calculated from Eqs. (2) & (3) as 24.99°C. ANOVA table 8 shows Vt′ had the highest effect on the percentage of contribution (79.57%), followed by Fd′ and Dt′ in consecutive decreasing order of percentage contribution. F-value showed the relative importance firstly Vd′, secondly Fd′ and lastly was Dt′. P-value was significant for Vd′. Figure 5(b) shows the variation of mean Tc′ with different factor levels (cutting parameters). On incrementing cutting speed (Vd′), feed (Fd′) and depth of cut (Dt′) temperature increased.

3.6 *Confirmation Experiment*

In Taguchi method confirmation experiments were performed to calculate the difference between actual values and predicted values of response at optimum levels. If the reliability of the condition was assumed to be 95%, then the confidence level (CI) was calculated from Eq. (4) [56]

$$CI = \sqrt{\left[F(\alpha, 1, f_e) V_e \left\{1/N_{eff} + (1/R)\right\}\right]} \quad (4)$$

$$N_{eff} = N/\left(1 + T_{dof}\right) Neff \; = \; N \quad (5)$$

Where, $F(\alpha,1,f_e)$ is the F-ratio required for $100(1-\alpha)$ percent confidence level, f_e, DOF for error = 20, Ve = AdjMs for error, N = total number of experiment($9 \times 3 = 27$), R = number of replications for confirmation of experiments (3) and T_{dof} T_f = total degree of freedom associated with mean optimum is ($2 \times 3 = 6$). From standard statistical table, the value of F ratio for $\alpha = 0.05$ is $F(0.05,1,20) = 4.351$. Substituting values form respective tables 6, 7 and 8.

CI value for cutting force was ±0.85, machining time was ±2.39 and temperature was ±0.5. The predicted optimal ranges of cutting force, machining time and temperature were $43.49 \leq Ct' \leq 45.34$, $34.7 \leq Mt' \leq 39.48$ and

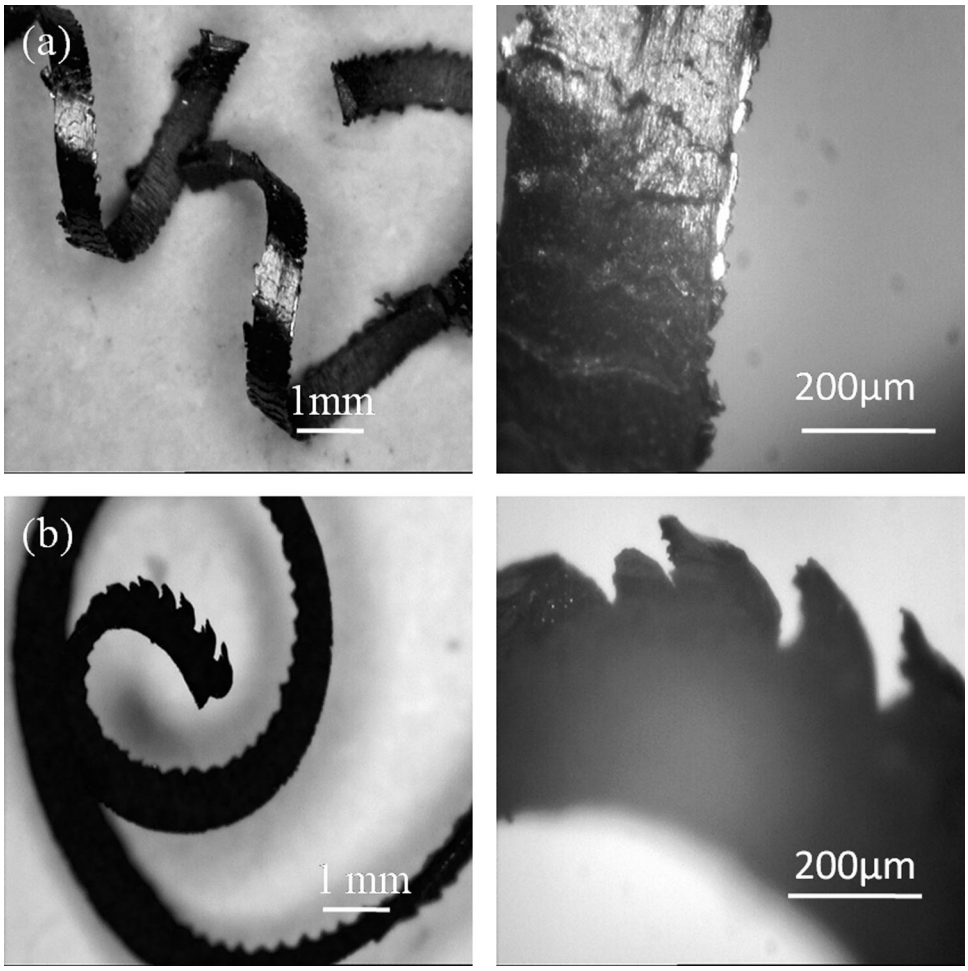


Figure 7. **(a)** Optical images of chip at Vd′ 25 m/min., Fd′ 0.05334 mm/rev. and Dt′ 0.254 mm. **(b)** Vd′ 56 m/min., Fd′ 0.13716 mm/rev. and Dt′ 0.381 mm.

$24.44 \leq Tc' 25.54$ respectively. The confirmatory tests were performed for the optimum level of parameters for cutting force, machining time and temperature. It was found that response values were 45 N, 37 seconds and 25°C, respectively. This was within the range of the confidence interval (CI) (table 9).

3.7 *Confirmation test results*

The experiments were conducted again for validating the predicted values obtained for each response. It was found that the values obtained after performing experiments at the optimum level were within the predicted range.

Regression models were developed for studying the relationship between machining parameters and the correlation between independent and dependent parameters. From table 10, it was observed that R-Sq value was greater than 89–97% and approaching 100% and a good correlation was obtained between cutting parameters and experimental outputs.

3.8 *Chip compression ratio*

The chips were collected during cryogenic turning. The ratio of deformed chip thickness to undeformed chip thickness is defined as the chip compression ratio by utilizing Eq. (6).

$$CR = \frac{t_d}{t_u} \tag{6}$$

Chip compression ratio showed the frictional condition on the tool surface. Generally, at a low cutting speed of less than 30 m/minute, the chip compression ratio was low due to discontinuous chips. Chip compression ratio decreased with increased cutting speed, feed and depth of cut for cryogenically treated cutting inserts due to increase in hardenability and wear resistance created in the cutting inserts [32]. From figure 6 it was found that chip compression ratio was less than 2 for experiments performed under cryogenic turning with LN_2 and with the increase in cutting parameters at experiment performance of 9 of L_9 DoE chip compression ratio was less than at lower cutting

parameter at experiment performance at 1 of L_9DoE. This may be due to the low temperature created at the interface of cutting insert and workpiece. The effect of thermal softening of the tool was declined. The tool became hard with low wear rate.

The chips generated during cryogenic turning were analysed by optical images as shown in figure 7(a) and (b). It was observed that they were staggered and most of the material had one sided flow. Chips were found discontinuous, brittle, easily breakable and long snarled. The low temperature created by a liquid nitrogen environment played an important role for such chip formation.

4. Conclusions

The exhaustive literature survey was conducted on machining for humankind and found that there was the need for experimental investigation during cryogenic turning in LN_2 environment at the mating zone of TiN coated cutting insert and AISI D3 workpiece. The experiment was carried out on lathe in created LN_2 environment. The following conclusions were drawn.

1. Chips generated during cryogenic turning were staggered and most of the material had one sided flow. The chip thickness increased with the increase of the machining parameters (speed, feed and depth of cut).
2. Chips were discontinuous, brittle, and easily breakable. The chips were long snarled. This may be due to the cooling effect created by the liquid nitrogen environment.
3. Taguchi based design of experiments [orthogonal array] was applied to reduce the number of the experimental run. This saved energy, raw materials for workpiece and cutting tool.
4. During analysis, the optimized value of cutting force was found at speed (56 m/min.), feed (0.05334 mm/rev.), and depth of cut (0.254 mm). According to mean effect plot, it was found that cutting force was inversely proportional to the speed and directly proportional to feed and depth of cut.
5. According to mean effect plot, it was found that the machining time was directly proportional to speed, feed and depth of cut. This may be due to the faster tool movement and reduction in friction between the cutting tool and workpiece. LN_2 absorbed heat generated between the cutting tool and workpiece and built a fluid layer which reduced friction. The optimized value of machining time was found at speed (56 m/min.), feed (0.13716 mm/rev.), and depth of cut (0.476 mm).
6. The optimized value of temperature was found at speed (25 m/min.), feed (0.05334 mm /rev.) and depth of cut (0.254 mm). According to mean effect plot, it was found that the temperature was directly proportional to speed, feed and depth of cut. This may be due to increasing friction at the interface of cutting tool and workpiece.

Acknowledgements

Authors are thankful to the workshop and laboratories facilities shared by Delhi Technological University and Indian Institute of Technology, Delhi (India).

Nomenclature

Ct'	Cutting Force during cryogenic turning with LN_2
Mt'	Machining time during cryogenic turning with LN_2
Tc'	Temperature during cryogenic turning with LN_2
Vd'	Speed (m/min.)
Fd'	feed (mm/rev.)
Dt'	Depth of cut (mm)
λ'	Inclination angle
α'	Orthogonal rake angle
β'	Orthogonal clearance angle of principal flank
γ'	Auxiliary orthogonal clearance angle
φ'	Auxiliary cutting edge angle
θ'	Principal cutting edge angle
r'	Nose radius (mm)
OA	Orthogonal array
DoE	Design of Experiment
LN_2	Liquid Nitrogen
CO_2	Frozen Carbon dioxide
AD	Anderson-Darling value

References

[1] Byrne G and Scholta E 1993 Environmentaly clean machining processes - a strategic approach. *CRIP Ann. Manuf. Technol.* 42(1): 471–474

[2] Shokrani A, Dhokia V and Newman S T 2012 Environmentally conscious machining of difficult-to- machine materials with regard to cutting fluids. *Int. J. Mach. Tools Manuf.* 57: 83–101

[3] Pereira O, Rodriguez A, Fernandez-A A I, Barreiro and Lopezde L N 2016 Cryogenic and minimum quantity lubrication for eco-friendly turning of AISI 304. *J. Clean Prod.* 139: 440–449

[4] Hong S Y and Broomer M 2000 Economical and ecological cryogenic machining of AISI 304 austenitic stainless steel. *Clean Technol. Environ. Policy* 2(3): 157–166

[5] Boubekri N and Foster P R 2015 A technology enabler for green machining: minimum quantity lubrication (MQL). *J. Manuf. Manag.* 212(5): 556–566

[6] Pusavec F, Kramar D, Krajnik P and Kopac J 2010 Transitioning to sustainable production-Part-II: Evaluation of sustainable machining technologies. *J. Clean Prod.* 118(12): 1211–1221

[7] Debnath S, Reddy M M and Yi Q S 2014 Environmental friendly cutting fluids and cooling techniques in machining: A review. *J. Clean Prod.* 83: 33–47

[8] Liew P J, Shaaroni A, Sidik N A C and Yan J 2017 An Overview of current status of cutting fluids and cooling techniques of turning hard steel. *Int. J. Heat Mass Transf.* 114: 380–394

[9] Hong S Y and Zaho Z 1999 Thermal aspects, material considerations and cooling strategies in cryogenic machining. *Clean Prod. Process.* 1: 107–116

[10] Yang W H and Tarang Y S 1998 Design optimization of cutting parameters for turning based on Taguchi method. *J. Mater. Process. Technol.* 84: 122–129

[11] Mia M 2017 Multi-response optimization of end milling parameters under through-tool cryogenic cooling condition. *Measurement* 111:134–145

[12] Chinchanikar S and Choudhary S K 2015 Machining of hardened steel—Experimental investigations, performance modelling and cooling techniques: A review. *Int. J. Mach. Tools Manuf.* 89: 95–109

[13] Sharma V S, Dogra M and Suri N M 2009 Cooling techniques for improved productivity in turning. *Int. J. Mach. Tool Manuf.* 49: 435–453

[14] Yildiz Y and Nalbant M 2008 A review of cryogenic cooling in machining processes. *Int. J. Mach. Tools Manuf.* 48: 947–964

[15] Manimaran G, Kumar M P and Venkatasamy R 2014 Influence of cryogenic cooling on surface grinding. *Cryogenics* 59: 76–83

[16] Manimaran G and Kumar M. Pradeep 2013 Effect of cryogenic cooling and sol-gel alumina wheel on grinding performance of AISI 316 stainless steel. *Arch. Civ. Mech. Eng.* 13: 304–312

[17] Dhananchnezian M and Kumar M P 2011 Cryogenic turning of the Ti-6Al-4V alloy with modified cutting tool inserts. *Cryogenics* 51: 34–40

[18] Sartori S, Ghiotti A and Bruschi S 2017 Hybrid lubricating/cooling strategies to reduce the tool wear in finishing turning of difficult-to-cut alloys. *Wear* 376–377: 107–114

[19] Hong S Y and Ding Y 2001 Cooling approaches and cutting temperatures in cryogenic machining of Ti-6Al-4V. *Int. J. Mach. Tools Manuf.* 41(10): 1417–1437

[20] Hong S Y, Dong Y and Jeong W 2001 Friction and cutting forces in cryogenic machining of Ti-6Al-4V. *Int. J. Mach. Tools Manuf.* 41(15): 2271–2285

[21] Hong S Y, Markus I and Jeong W 2001 New cooling approach and tool life improvement in cryogenic machining of titanium alloy. *Int. J. Mach. Tools Manuf.* 41(15): 2245–2260

[22] Dhar N R, Paul S and Chattopadhyay A B 2002 The influence of cryogenic cooling on tool wear, dimensional accuracy and surface finish in turning AISI 1040 and E4340C steels. *Wear* 249: 932–942

[23] Dhar N R, Paul S and Chattopadhyay A B 2002 Machining of AISI 4140 steel under cryogenic cooling—toolwear, surfaceroughness and dimensional deviation. *J. Mater. Process. Technol.* 123: 483–489

[24] Dhar N R and Kamruzzaman M 2007 Cutting temperature, tool wear, surface roughness and dimensional deviation in turning AISI-4037 steel under cryogenic condition. *Int. J. Mach. Tools Manuf.* 47: 754–759

[25] Dhar N R, Nand Kishore S V, Paul S and Chattopadhay A B 2002 Effects of cryogenic cooling on chips and cutting forces in turning AISI 1040 and 4320 steel. *Proc. IMeche Part B J. Eng. Manuf.* 216: 713–724

[26] Bermingham M J, Kirsch J, Sun S, Palanisamy S, Dargusch M S 2011 New observations on tool life, cutting forces and chip morphology in cryogenic machining Ti6AL4V. *Int. J. Mach. Tools Manuf.* 51: 500–511

[27] Diniz A E, Machado A R and Correa J G 2016 Tool wear mechanisms in the machining of steels and stainless steels. *Int. J. Adv. Manuf. Technol.* 87: 3157–3168

[28] Strano M, Chiappini E, Tirelli S, Albertelli P and Monno Michele 2013 Comparison of Ti6Al4 V machining forces and tool life for cryogenic versus conventional cooling. *Proc. IMechE Part B J. Eng. Manuf.* 227(9): 1403–1408

[29] Venugopal K A, Tawade R, Prashanth P G, Paul S and Chattopadhyay A B 2003 Turning of titanium alloy with TiB2-coated carbides under cryogenic cooling. *Proc. IMechE Part B J. Eng. Manuf.* 217: 1697–1707

[30] Sun S, Brandt M and Dargush M S 2010 Machining Ti-6Al-4V alloy with cryogenic compressed air cooling. *Int. J. Mach. Tools Manuf.* 50: 933–942

[31] Mia M and Dhar N R 2017 Influence of single and dual cryogenic jets on machinability characteristics in turning of Ti-6Al-4V. *Proc. IMechE Part B J. Eng. Manuf.* 1–16

[32] Khan A and Maity K 2017 Comparative study of some machinability aspects in turning of pure titanium with untreated and cryogenically treated carbide inserts. *J. Manuf. Process.* 28: 272–284

[33] Chetan, Ghosh S and Rao P V 2017 Performance evaluation of deep cryogenic processed carbide inserts during dry turning of Nimonic 90 aerospace grade alloy. *Tribol. Int.* 115: 397–408

[34] Lal D M, Renganarayanan S and Kalanidhi A 2001 Cryogenic treatment to augment wear resistance of tool and die steels. *Cryogenic*s 41: 149–155

[35] Shokarani A, Dhokia V and Newman S T 2016 Energy conscious cryogenic machining of Ti-6Al-4v titanium alloy. *Proc. IMechE Part B J. Eng. Manuf.* 232(10): 1690–1706

[36] Aggarwal A, Singh H, Kumar P and Singh M 2008 Optimization of multiple quality characteristics for CNC turning under cryogenic cutting environment using desirability function. *J. Mater. Process. Technol.* 205: 42–50

[37] Yilmaz B, Karabulut S and Gullu A 2018 Performance of analysis of new chip breaker for efficient machining of Inconel 718 and optimization of the cutting parameters. *J. Manuf. Process.* 32: 553–563

[38] Kalyan K K V B S and Choudhury S K 2008 Investigation of tool wear and cutting force in cryogenic machining using design of experiments. *J. Mater. Process. Technol.* 203: 95–101

[39] Viswanathan R, Ramesh S and Subburam V 2018 Measurement and optimization of performance characteristics in turning of Mg alloy under dry and MQL conditions. *Measurement* 120: 107–113

[40] Gupta M K and Sood P K 2016 Optimizing multi-characteristics in machining of AISI 4340 steel using Taguchi's approach and utility concept. *J. Inst. Eng. (India) Ser. C.* 97(1): 63–69

[41] Dureja J S, Singh R and Bhatt M S 2014 Optimization flank wear and surface roughness during hard turning of AISI D3 steel by Taguchi and RSM methods. *Prod. Manuf. Res.* 2: 767–783

[42] Mandal N, Doloi B and Mondal B 2016 Surface roughness predication model using zirconica toughned aluminia (ZTA): Taguchi method and regression analysis. *J. Inst. Eng. (India) Ser.* C. 97(1): 77–84

[43] Kumar Dr V, Kiran K B J and Rudresha N 2018 Optimization of machining parameters in CNC turning of stainless steel (EN19) by Taguchi's orthogonal array experiments. *Mater. Today Proc.* 5: 11395–11407

[44] Mozammel M M, Prithbey, Dey R, Hossain M S, Arafat S M T, Asaduzzaman Md, Ullah Md. S and Zobaer S M T 2018 Taguchi S/N based optimization of machining parameters for surface roughness, too wear and material removal rate in hard turning under MQL cutting condition. *Measurement* 122: 380–391

[45] Gunay M and Yucel E 2013 Application of Taguchi method for determining optimum surface roughness in turning of high-alloy white cast iron. *Measurement* 46: 913–919

[46] Durakbasa M N, Akdogan A, Vanil A S and Bulutsuz A G 2015 Optimization of end milling parameters and determination of the effects of edge profile for high surface quality of AISI H13 steel by using precise and fast measurements. *Measurement* 68: 92–99

[47] Rath D, Panda S and Pal K 2018 Prediction of surface quality using chip morphology with nodal temperature signatures in hard turning of AISI D3. *Mater. Today* 5: 12368–12375

[48] Zerti O, Yallease M A, Khettabi R, Chaoui K and Marbrouki T 2017 Design optimization for minimum technological parameters when dry turning of AISI D3 steel using Taguchi method. *Int. J. Adv. Manuf. Technol.* 89: 1915–1934

[49] Hasclik A and Caydas U 2008 Optimization of turning parameters for surface roughness and tool life based on Taguchi method. *Int. J. Adv. Manuf. Technol.* 38: 896–903

[50] Mandal N, Doloi B, Mondal B and Das R 2011 Optimization of flank wear using Zirconia toughned aluminia (ZTA) cutting tool: Taguchi method and Regression analysis. *Measurement* 44: 2129–2155

[51] Shokrani A, Dhokia V and Newman S T 2016 Investigation of the effects of cryogenic machining on surface integrity in CNC end milling of Ti-6Al-4 V titanium alloy. *J. Manuf. Process.* 21: 172–179

[52] Thirumalai R, Senthilkumar J S, Selvarani P and Ramesh S 2012 Machining characteristics of Inconel 718 under several cutting conditions based on Taguchi method. *Proc. IMeche Part C J. Mech. Eng. Sci.* 227(9): 1889–1897

[53] Fredj N, Habibsidhon and Braham C 2006 Ground surface improvements of the austentic stainless steel AISI 304 using cryogenic cooling. *Surf. Coat. Technol*. 200: 4846–4860

[54] Sivaiah P and Charkardha D 2018 A comparison with MQL, wet, dry machining. *CIRP J. Manuf. Sci. Technol.* 21: 86–96

[55] Zebia W and Kowlczyk R 2015 Estimating the effect of cutting data on surface roughness and cutting force during WC-CO turning with PCD tool using Taguchi design and ANOVA analysis. *Int. J. Adv. Manuf. Technol.* 77: 2241–2256

[56] Ross P J 1996 *Taguchi Techniques for quality engineering*. New York: McGraw-Hill: pp. 181–196

[57] Mohanty SD, Mahapatra SS and Mohanty R C 2019 PCA based hybrid Taguchi philosophy for optimization of multiple responses in EDM. *Sadhna* 44(2):1–9 https://doi.org/10.1007/s12046-018-0982-z

[58] Ghodsiyeh D, Akbarzadeh S, Izman S and Morad M 2019 Experimental investigation of surface integrity after wire electro-discharge machining of Ti-6Al-4V. *Sadhna* 44(196): 1–15 https://doi.org/10.1007/s12046-019-1184-z123456789

2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)
GLA University, Mathura, UP, India. Feb 28-29, 2020

2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) 978-1-7281-6575-2/20/$31.00 ©2020 IEEE 10.1109/PARC49193.2020.236577

Real Time Experimental control and Design of FOPI^{λ} and L-PID Controllers tuned by Invasive-weed optimization for Trajectory Control of TRAS

Hem Prabha
Dept. of Electrical Engineering
Delhi Technological University
Delhi-110042, India
prabhasingh1686@gmail.com,

Rajul Kumar
Dept. of Electrical Engineering
Delhi Technological University
Delhi-110042, India
rajul_bt2k17@dtu.ac.in

***Abstract*— This article presents formational designing and real time experimental validation of Fractional order-Proportional integral (FOPI^{λ}) and Linear-Proportional Integral Derivative (L-PID) controllers for trajectory stabilization of Twin Rotor Aerial System (TRAS) which is conceded as helicopter model having pitch and yaw actuators for hovering of TRMS in vertical and horizontal planes respectively. Both the controller parameters are tuned using robust metamorphic algorithm called Invasive Weed Optimization. The control performances of both the designed controllers are examined for trajectory tracking and time taken to stabilize after the given manual disturbances. It is experimentally validated that FOPI^{λ} shows superior and robust control behavior then L-PID for benchmark TRAS plant.**

Keywords—IWO, FOPI$^{\lambda}$, L-PID, TRAS, Stabilization Control

I. INTRODUCTION

In contemporary year's automatic control of aerial vehicles are in curious interest because of their non-linear, complex system dynamics and wide range of applications in sky vigilance, detecting forest fire, natural calamity relief assistance[1]-[4]. Twin Rotor Ariel System (TRAS) plant mimics 2-DOF helicopter model having a two 44 watt DC motor propellers for pitch and yaw angle control and for achieving desired set-point hovering in vertical and horizontal planes without significant cross coupling between the two planes. The pitch and yaw DC motors receives voltage control signals from PC via PCI-1711 card and vertical beam which is mounted over electrical base plate. These control signals decisive the needed pneumatic torque by altering the rpm of pitch and yaw rotors resulting change in pitch and yaw hovering angles[5], [6].

TRAS is recognized as most non linear and complex behavior because of three moments i.e. gravity, friction and gyroscopic acting on pitch and yaw propellers [7], [8]. The stabilization control of such complex plants cannot be fulfilled by adoption traditional controllers like L-PID [9], [10].

To realize these complex system uncertainties two Fractional order-PI$^{\lambda}$ controllers are designed and experimentally validated in real time for pitch and yaw angle stabilization control. The designed FOPI$^{\lambda}$ controller have capability to stabilize close loop dynamics from dynamic environment through optimized fractional power of integral scheme [11], [12].

For optimal utilization and tuning of designed controllers a ecologically inspired catechistic algorithm [13], [14] called invasive weed optimization is used. Invasive weed optimization (IWO)[15]-[17] is influenced from robust pioneering of invasive weeds in plant cultivation. Invasive weed Colonize very vigorously and shows very adaptive nature to environmental diversification. Therefore the robust acclimating behavior of invasive weeds provides an edge for dealing with optimization of complex plants in control engineering practices. After tuning with IWO trajectory control performance both FOPI$^{\lambda}$ and L-PID controller are validated and compared mathematically and experimentally by giving real time manual disturbances for a regular time interval, these manual disturbances provided through counterbalance resembles aerial, wind and pneumatic disturbances on helicopter.

The paper is fragmented into five sections, Introduction then section II describing TRAS plant dynamical Modelling followed by Section III comprises designing of FOPI$^{\lambda}$ and L-PID controllers both for pitch and yaw propellers. Section IV intimates the experimental comparisons and results followed by conclusion and references in Section IV and V.

II. DYNAMICAL MODELLING OF TWIN ROTOR ARIEL SYSTEM (TRAS)

A. Dynamical Momentum Equations

The experimental laboratory setup for TRAS is given in fig.1 with pitch and yaw two angular rotors.

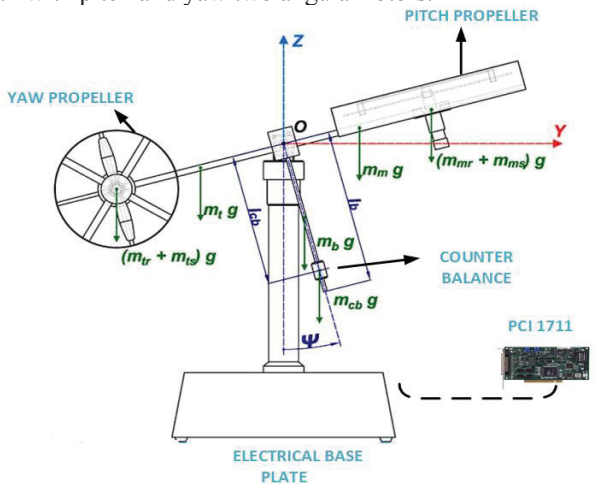


Fig. 1. TRAS Experimental system

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:15:17 UTC from IEEE Xplore. Restrictions apply.

The mathematical statements of pitch angle (Ψ) propeller are provided in equation (1) - (5).

$$I_1\ddot{\Psi} = M_1 - M_{FG} - M_{B\Psi} - M_G \quad (1)$$

$$M_1 = a_1\tau_1^2 + b_1\tau_1 \quad (2)$$

$$M_{FG} = M_g sin\Psi \quad (3)$$

$$M_{B\Psi} = B_{1\Psi}\dot{\Psi} + B_{2\Psi}sign\dot{\Psi} \quad (4)$$

$$M_G = K_{gy}M_1\dot{\varphi}cos\Psi \quad (5)$$

The mathematical statements of yaw angle (φ) propeller are provided in equation (6) - (8).

$$I_2\ddot{\varphi} = M_2 - M_{B\varphi} - M_R \quad (6)$$

$$M_2 = a_2\tau_2^2 + b_2\tau_2 \quad (7)$$

$$M_{B\varphi} = B_{1\varphi}\dot{\Psi} + B_{2\varphi}sign\dot{\varphi} \quad (8)$$

Dynamical systemization of pitch and yaw rotors consist of gyroscopic, frictional and non-linear, gravity characteristics momentums that influence TRAS hovering non-linearities and panorama. The numerical values of system parameters including static, frictional and cross-coupling constants are given in Table.1.

TABLE I. TRAS PARAMETRIC VALUES

TRMS modeling Constants	Values
$\boldsymbol{I_1}$- M.I of Pitch Propeller	$\mathbf{7.9 \times 10^{-2} Kgm^2}$
$\boldsymbol{I_2}$- M.I of Yaw Propeller	$\mathbf{4 \times 10^{-3} Kgm^2}$
$\boldsymbol{a_1}$- Pitch Propeller Static Constant	$\mathbf{0.4134}$
$\boldsymbol{b_1}$- Pitch Propeller Static Constant	$\mathbf{0.655}$
$\boldsymbol{a_2}$- Yaw Propeller Static Constant	$\mathbf{0.67}$
$\boldsymbol{b_2}$- Yaw Propeller Static Constant	$\mathbf{0.74}$
$\boldsymbol{M_g}$- Gravity Moment	$\mathbf{0.399}$
$\boldsymbol{B_{1\psi}}$- Frictional Moment of Pitch Propeller	$\mathbf{5 \times 10^{-0.6}}$
$\boldsymbol{B_{2\psi}}$- Frictional Moment of Pitch Propeller	$\mathbf{0.7 \times 10^{-0.6}}$
$\boldsymbol{B_{1\varphi}}$- Friction Moment of Yaw Propeller	$\mathbf{2 \times 10^{-0.5}}$
$\boldsymbol{B_{2\varphi}}$- Friction moment of Yaw Propeller	$\mathbf{3.5 \times 10^{-0.5}}$
$\boldsymbol{K_{gy}}$- Gyroscopic Moment	$\mathbf{0.053}$
$\mathbf{k_1}$- DC Gain of Pitch Propeller	$\mathbf{1.45}$
$\boldsymbol{k_2}$- DC Gain of yaw Propeller	$\mathbf{0.79}$
$\boldsymbol{T_{11}}$- Moment Parameter of Pitch Propeller	$\mathbf{1.01}$
$\boldsymbol{T_{10}}$- Moment Parameter of Pitch Propeller	$\mathbf{1}$
$\boldsymbol{T_{21}}$- Moment Parameter of yaw Propeller	$\mathbf{1}$
$\boldsymbol{T_{20}}$- Moment Parameter of yaw Propeller	$\mathbf{1}$
$\boldsymbol{T_p}$- Cross-coupling value	$\mathbf{2.1}$
$\boldsymbol{T_0}$- Cross-coupling value	$\mathbf{3.85}$
$\boldsymbol{k_c}$- Gain of Cross-coupling	$\mathbf{-0.34}$

After Solving and substituting TRAS system constants in non-linear pitch and yaw rotor equations gives transfer function in Laplace state-space form in equation (9) and (10).

$$T_P = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \quad (9)$$

$$T_{p(s)} = \begin{bmatrix} \frac{2.346}{(s+0.934)(s^2+0.0765s+3.689)} & 0 \\ \frac{2.567(s+0.365)}{s(s+6)(s+0.944)(s+0.95)} & \frac{4.8}{s(s+9)(s+3)} \end{bmatrix} \quad (10)$$

III. FOPI^λ AND L-PID CONTROLLER DESIGNING

Two FOPI^λ and L-PID controllers are designed for pitch and yaw propellers stabilization. The general control scheme is given in equation (11) consisting proportional, integral and derivative control arrangements where as equation (12) shows error function as difference of desired propeller angle and experimental propeller angle.

$$c(t) = k_p e(t) + k_i \int e(t)dt + k_d \frac{de(t)}{dt} \quad (11)$$

$$e(t) = z_{desired\ angles} - z_{actual\ angles} \quad (12)$$

A. L-PID CONTROL SCHEME

The control structural of L-PID control arrangement with control equation (13) in mathematical form is provided in Fig.2.

$$C(s) = k_p + \frac{k_i}{s} + k_d s \quad (13)$$

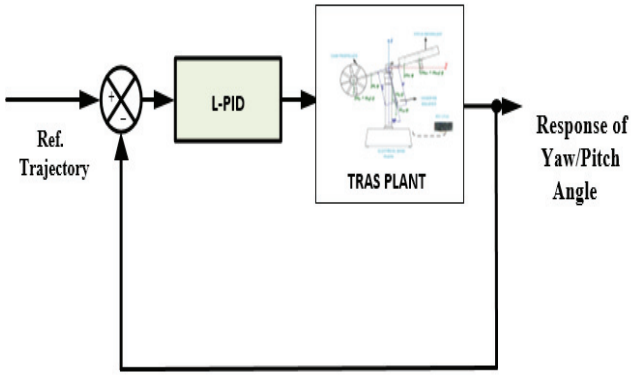


Fig. 2. L-PID control arrangement

B. FOPI^λ CONTROL SCHEME

According to equation (8), the modified transfer function of IO-PID controller structure is written in (10)

$$C(s) = k_p + \frac{k_i}{s^\lambda} \quad (10)$$

The design of IO-PID controller with TRMS plant is shown in Fig.3.

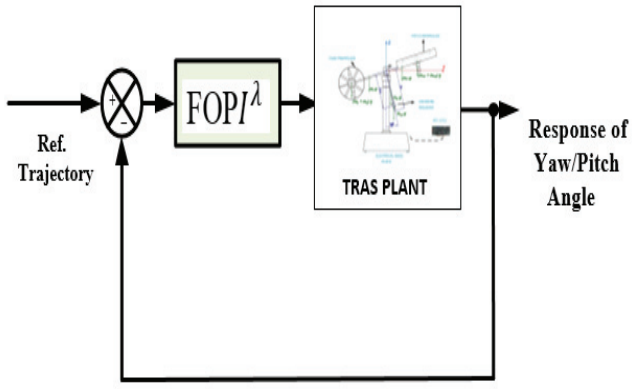


Fig. 3. FOPI$^\lambda$ control arrangement

IV. INVASIVE WEED OPTIMIZATION (IWO) OF DESIGNED FOPI^λ AND L-PID CONTROLLERS

In recent years scientists and industries are fascinated towards solving non-definitive or discontinuous control engineering problems by nature influenced non-gradient

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:15:17 UTC from IEEE Xplore. Restrictions apply.

stochastic algorithms. Invasive weed optimization is one of the ecological influenced metamorphic optimization technique. In scientific language Weeds can be elaborated as any undesirable shrubs, trees, saplings in our cultivation or optimal solution search space. Invasive traits of weeds poses critical hazard to cultivation, so many attempts like tillage, defoliation and machine weeding are rendered to get rid of these nasty and stubborn plants.

They can reproduce by seeds or sex cells, when ancestor weed forms a seed by fertilization of egg through pollen. They flourish and reproduce at the last cycle of their life, although the way of living their entire life cycle depends upon two selections called r and k force criterion. The r criteria includes more abundance, little troop and low value to permanent colonization for winning in competition this cooperates them to succeed in insecure and unpredictable ecological conditions, whereas k criteria consist of bigger troops, slow fecundity and give more emphasis on permanent colonization. After reproduction process seeds are haphazardly dispersed in solution search space according to equations (11)-(13), through random values such that they domicile near their parent weeds shown in Fig.4, Fig.5, Fig.6.

Fig. 4. Distributing of initial spore

Fig. 5. Dispersing of spores in neighborhood orbits

Fig. 6. Rolling down of spores into most distant orbits

Weeds are very adaptive and flexible to their surroundings and environmental changes, they overrun and seize the search space by permanent colonization and also getting improved by having best fitness among plant community through interacting with peer weeds based upon their capability traits. Weeds are allowed to clone based upon the linear fitness enhancement of whole community as shown in Fig 7.

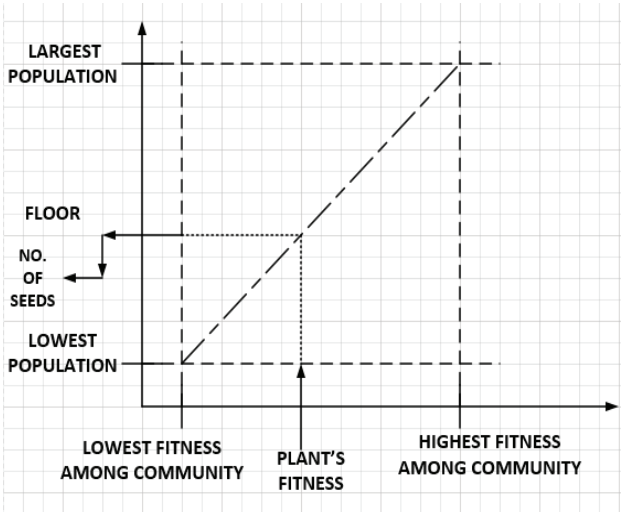


Fig. 7. Linear increase in reproduction of seeds in weeds community

$$Mean = 0 \tag{11}$$

$$Variance = varied\ over\ search\ space \tag{12}$$

Standard deviation is given by σ_{loops} ,

$$\sigma_{loops} = \frac{(loop_{max} - loop)^a}{(loop_{max})^a}(\sigma_{start} - \sigma_{end}) + \sigma_{end} \tag{13}$$

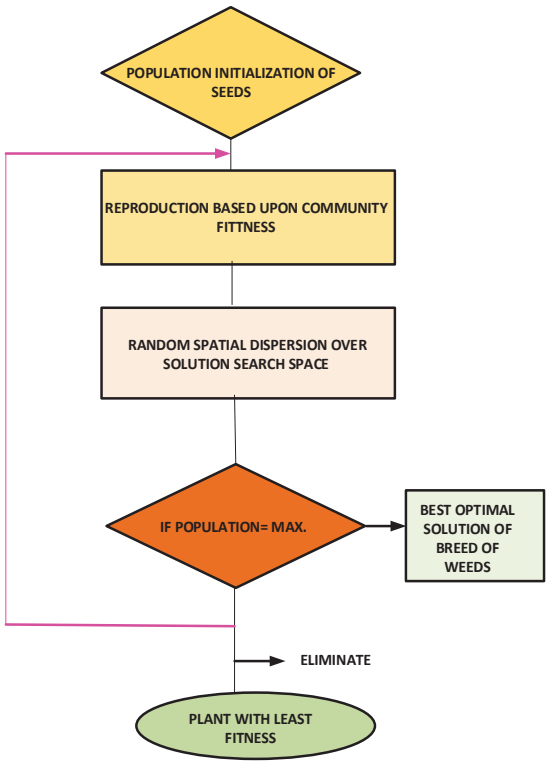


Fig. 8. IWO flowchart

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:15:17 UTC from IEEE Xplore. Restrictions apply.

The loops of dispersion of seeds in search space works according to equitation (13) ensuring less chances to disperse in larger radial orbit in each successive iterations. The process includes elimination of less fit weeds, and revolution of their way of living from r to k force criterion. When maximum population is reached in community, weeds with highest fitness are segregated together and one with lower fitness are disqualified and not allowed to breed further. The outflow of IWO is elaborated in form of flow chart provided in Fig.8.

The tuned values of both the controller parameters are given in table.2.

TABLE II. PITCH AND YAW PROPELLERS OPTIMIZED VALUES FOR L-PID CONTROLLER

Rotor	Optimization parameter	Tuned Values
Pitch Controller	k_{p1}	6.95
	k_{i1}	8.923
	k_{d1}	7.784
Yaw Controller	k_{p2}	2.798
	k_{i2}	4.554
	k_{d2}	3.843

TABLE III. PITCH AND YAW PROPELLERS OPTIMIZED VALUES FOR FOPI^{λ} CONTROLLER

Rotor	Optimization Parameter	Tuned Values
Pitch Controller	k_{p1}	7.213
	k_{i1}	18.101
	λ_1	0.634
Yaw Controller	k_{p2}	4.051
	k_{i2}	19.544
	λ_2	0.863

These IWO tuned values of controller parameters are tested and validated on real time and stabilization graphs are obtained mentioned in section 5.

V. STABILIZATION RESULTS AND EXPERIMENTAL COMPARISION OF CONTROLLERS

The L-PID and FOPI^{λ} controller response on pitch and yaw propellers along with control signal are given in Fig.9, Fig.10, Fig.11, Fig.12

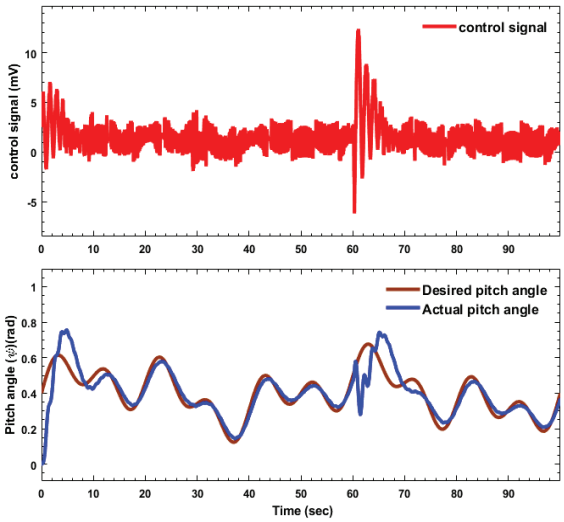


Fig. 9. L-PID controller Pitch angle response

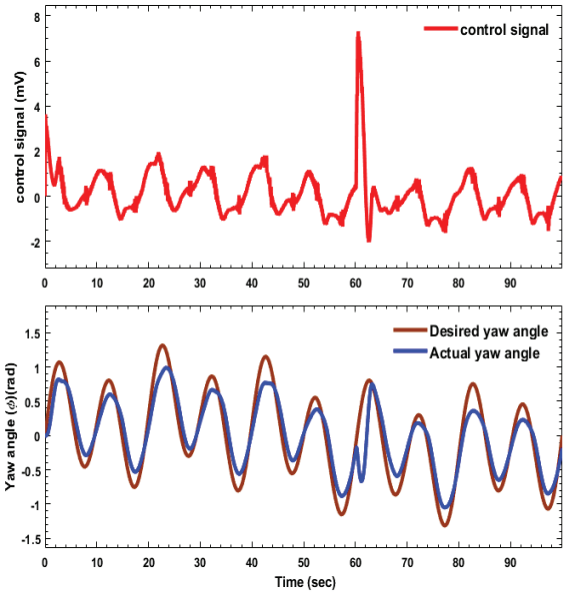


Fig. 10. L-PID Yaw angle Response

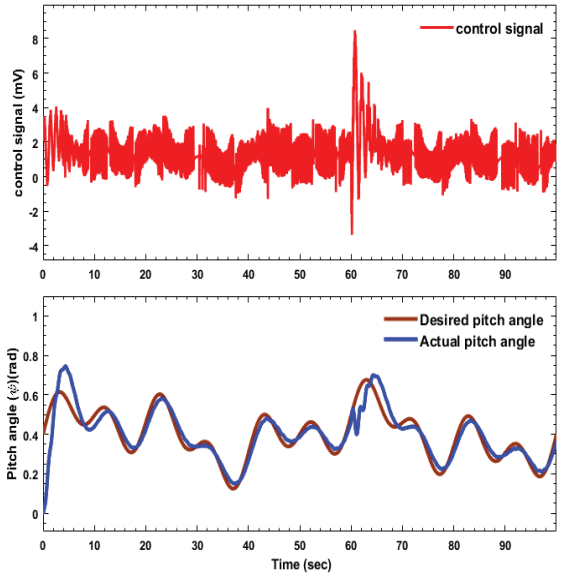


Fig. 11. FOPI^{λ} controller pitch response

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:15:17 UTC from IEEE Xplore. Restrictions apply.

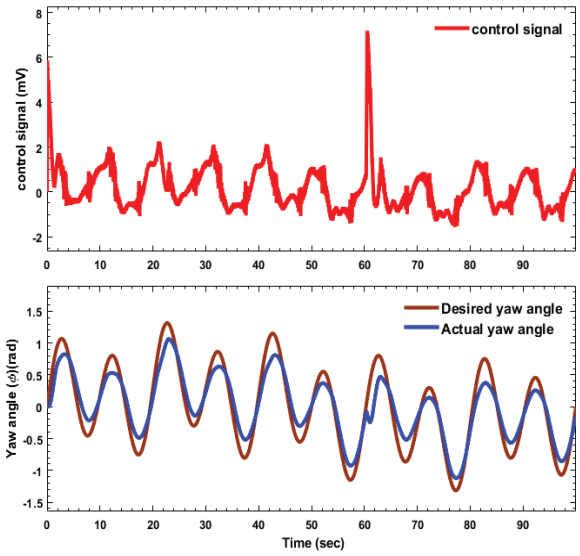


Fig. 12. FOPI^{λ} Controller yaw response

A comparison table describing time taken to stabilize after the given manual disturbance are given in table.4 and table.5 for L-PID and FOPI^{λ} controllers respectively.

TABLE IV. L-PID CONTROLLER RESPONSES

Rotor	**L-PID**		
Pitch	Distrubence Initiated (Sec)	Stabilization Instant (Sec)	Stabilization time (Sec)
	60	74	14
Yaw	60	65	5

TABLE V. FOPI^{λ} CONTROLLER RESPONSES

Rotor	***FOPI*$^{\lambda}$**		
Pitch	Distrubence Initiated (Sec)	Stabilization Instant (Sec)	Stabilization time (Sec)
	60	66	6
Yaw	60	62	2

It is observed from Table IV and Table V, that FOPI^{λ} controller takes less time to stabilize pitch and yaw angles of TRAS after the given manual disturbances and shows robust control signal then L-PID controller

VI. CONCLUSION

In this paper, optimal control of Twin rotor aerial system (TRAS) has been achieved by designing and real time experimental validation of L-PID and FOPI^{λ} controllers. These controller parameters are optimized between constraints through analytical metamorphic algorithm inspired from robust and stubborn colonization traits of invasive weeds. After experimental verification it is found that FOPI^{λ} shows better control behavior than L-PID controller for stabilization of pitch and yaw rotors of TRAS aerial plant, this signifies that it is also capable of receding wind and atmospheric disturbances on real helicopter and UAVs.

ACKNOWLEDGMENT

We would like express our warm gratitude to Indian government and DTU to provide the real time experimental plant of aerial system.

REFERENCES

[1] H. Shakhatreh *et al*., "Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges," in *IEEE Access*, vol. 7, pp. 48572-48634, 2019.

[2] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," arXiv preprint arXiv:1803.00680, 2018..

[3] Y. Sun, D. Wing Kwan Ng, D. Xu, L. Dai, and R. Schober, "Resource allocation for solar powered uav communication systems," in 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), June 2018, pp. 1–5.

[4] G. Chmaj and H. Selvaraj, "Distributed processing applications for UAV/Drones: a survey," in Progress in Systems Engineering. Springer, 2015, pp. 449–454

[5] Ahmad, S. M., Chipperfield, A. J., & Tokhi, M. O. (2000). Dynamic modelling and optimal control of a twin rotor MIMO system. In Proceedings of the 2000 IEEE international conference on national aerospace and electronics (pp. 391–398)..

[6] Ahmad, S. M., Shaheed, M. H., Chipperfield, A. J., & Tokhi, M. O. (2000). Nonlinear modelling of a twin rotor MIMO system using radial basis function networks. In Proceedings of the 2000 IEEE international conference on national aerospace and electronics (pp. 313–320).

[7] Aldebrez, F. M., Mat-Darus, I. Z., & Tokhi, M. O. (2004). Dynamic modelling of a twin rotor system in hovering position. In Proceedings of the 2004 international symposium on control, communications and signal processing (pp. 823–826).

[8] Avila-Vilchis, J. C., Brogliato, B., Dzul, R., & Lozano, R. (2003). Nonlinear modelling and control of helicopters. Automatica, 39(9), 1583–1596.

[9] S. K. Pandey and V. Laxmi, "Control of twin rotor MIMO system using PID controller with derivative filter coefficient," *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*, Bhopal, 2014, pp. 1-6.

[10] V. S. Rao, V. I. George, S. Kamath and C. Shreesha, "Reliable robust PID controller design for TRMS," *2017 11th Asian Control Conference (ASCC)*, Gold Coast, QLD, 2017, pp. 565-569.

[11] S. K. Mishra and S. Purwar, "To design optimally tuned FOPID controller for twin rotor MIMO system," *2014 Students Conference on Engineering and Systems*, Allahabad, 2014, pp. 1-6.

[12] S. Chaudhary and A. Kumar, "Control of Twin Rotor MIMO System Using 1-Degree-of-Freedom PID, 2-Degree-of-Freedom PID and Fractional order PID Controller," *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2019, pp. 746-751.

[13] H. N. Al-Duwaish and M. Bettayeb, "Control of nonlinear dynamical systems using genetic algorithms," *Proceedings of 12th IEEE International Symposium on Intelligent Control*, Istanbul, Turkey, 1997, pp. 421-424.

[14] P. Biswas, R. Maiti, A. Kolay, K. Das Sharma and G. Sarkar, "PSO based PID controller design for twin rotor MIMO system," *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*, Calcutta, 2014, pp. 56-60.

[15] Prithwish Chakraborty, Gourab Ghosh Roy, Swagatam Das and B. K. Panigrahi, "On population variance and explorative power of invasive weed optimization algorithm," *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, Coimbatore, 2009, pp. 227-232.

[16] D. Vyas and G. Parmar, "Reduced order modelling of ltisystems using invasive weed optimization," *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*, Indore, 2017, pp. 1-4.

[17] N. C. Patel, M. K. Debnath, B. K. Sahu, S. S. Dash and R. Bayindir, "Multi-Staged PID Controller Tuned by Invasive Weed optimization Algorithm for LFC Issues," *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*, Paris, 2018, pp. 1358-1362

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:15:17 UTC from IEEE Xplore. Restrictions apply.

Purakala

ISSN:0971-2143

Vol-31-Issue NO.-37-May-2020

(UGC Care Journal)

Relegated, Stigmatized and Peripheral Voices: A Study of Unsung Mythological Women Characters in Chitra Banerjee Divakaruni's *The Forest of Enchantments*

Poorva Gulati
English Faculty at Delhi Technological University and a Research Scholar
poorvagulati94@gmail.com

Abstract

Gender is the self-representation of a person as male or female based on societal norms. Judith Butler elucidates that we cannot define a stable subjectivity about 'performing' various gender roles, it is the very act of performing gender that constitutes our identity. In our religious and mythological texts, we see that punishments for misdeeds meted out to women are biased. When the great epics were written, these norms were established which not only expect people to behave accordingly but impose stringent penalties on those who did not abide by them. Women are forced to get into the allotted roles prescribed for them in epics and historical accounts. Chitra Banerjee Divakaruni brings forth the ignored voices by giving different perspectives on various women characters like Urmila, Kaikeyi, Shabri, Mandodiri, Sarama, Shanta, Ahilya, and Surpanakha, who were relegated at the margins even in our sacred and religious narratives. Mythology is paradoxical in a way, on the one hand, it deifies them on a pedestal, and on the other hand, they are time and again asked to give parikshas (tests) to prove their chastity.The paper aims to analyze the critical corpus of problematic sisterhood characterized by male preconceptions where a woman is no less than the embodiment of betrayal, transgressions, and indiscretion. Through a study of Chitra Banerjee Divakaruni'sThe Forest of Enchantments, this paper attempts astudy ofSurpanakha, Mandodari, Sarama (Vibheeshan's wife) and Kaikeyi. It telegraphs the readers into Sita's search for selfhood, gives an ability of articulation to the margins, and thus margins transcend into the center of the narrative. If a character like Sita is an embodiment of unflinching dedication, forbearance, stoicism, she also carries the quintessence of knowledge, rationality, and finer sensibility. The Study gives us an insight as to how objectified goddess is worshipped from a distance as a magnificent work of art, withouthuman emotions but only reifying her to heights, oblivious of her troubles that she was enduring in her troubled existence. Divakaruni's rendition of women characters is a comment on the engulfing silences of duty and betrayal where the story of gods and goddesses or Dharma is not always colored with polar hues, they are more complexed, more layered and more nuanced.

Keywords: Sitayan, deification,subjugation, performing gender roles, identity

Copyright © 2020Author

Purakala

ISSN:0971-2143

Vol-31-Issue NO.-37-May-2020

(UGC Care Journal)

Introduction:

Indian scriptures, Vedas, Epics, and Upanishads are replete with mythological stories that are reinterpreted, retold, and revised numerous times by various contemporary writers because these stories are deeply ingrainedin the collective unconscious of Indian society and form the ideological basis of thinking. Myths have always captivated theimaginative thoughts of writers and summoned theirinterpretive capabilities to rework on the traditional and culture-specific ideological structure of a society. Myths have been attributed with dual understanding, first as a rationalistic approach where we tend to find evidence, empirical proofs to substantiate it and it is here, we tend to define myth as falsehood or unreliable story or just merely a belief. Secondly, writers and theologists try to understand myth in its intuitive mode of cosmic understanding which is deployed more often in its literary contexts.

Indian Myths and mythical stories pay a pivotal role in defining the power structure of society and present stigmatized images of women. The contemporary modern literature attempts toquestion and review the stereotyped images of women in mythology and either try to subvert or invert the ideological basis ofthe imbalanced power structure. Thus the sacred epics of Hindu mythology are reinterpreted, stories are reanalyzed, female characters are given voice to question the patriarchal norms and revisionist strategies. The ancient epics present the conventionalimages of women. The images of Sita, Savitri, Draupadi, Mandodri, Shurpanakhaare deeply ingrained in the social structure of Indian society that resonatesthe voices of submission and subordination irrespective of any era.

Method:

Chitra Banerjee Divakaruniresurrects the long-forgotten Indian myths, beliefs, and unrecorded historical accounts of women, which are essential for sustaining culture. This paper studies how diasporic writings give another vantage point to see myth and mythology from an altogether different direction. They are perceived from the dimensions of two worlds; from a domain where there is a sense of nostalgia for the homeland, but also acceptability to see everyday life with that shared sense of nostalgic compositeness in an alien land. In the paper, there is an attempt to create new myths through ways of re-mythification and dissolve older

Copyright © 2020Author

myths through the process of de-mythification. In the creation of new myths binaries are coalesced by blending: reality with imagination; conscious with the subconscious; truth with falsehood; physical with metaphysical; one woman rescue another woman through the bonds of sisterhood without taking the support of men.

Findings and Results:

Chitra Banerjee Divakaruni's most recent publication, *The Forest of Enchantments*, is the retelling of *The Ramayana* from Sita's perspective. Valmiki convinced her to write for her own self and her narrative illuminates the darker recesses and the fatal silences in her retelling *Sitayan.*The Forest full of enchantments becomes a gender-neutral space where it converges the polar gender dichotomies laid down by society. The forest becomes a realm where the prescriptive gender roles are diluted and more freedom is given to characters in contrast to the life of a kingdom. It is not just Sita, but a human narrative of other women in the epic who are often misunderstood and relegated to the margins like Surpanakha, Kaikeyi, Mandodari, Urmila, and Sunaina. *Sitayan* unearths the saga of a wronged woman. It is a powerful portrayal of human frailties, human emotions, and human weaknesses that a being faces in the *mrityulok*. Ram and Sita are gods and goddesses but they too are affected by the notions of duty, betrayal, honor, and infidelity. Divakaruni gives a realistic description of women's struggle to retain autonomy in a world that privileges men. Divakaruni's efforts to converge ancient story into a contemporary battle of choices are unique and aim to transcend gender polarities.

Indian immigrants in America and Canada face problems of identity and are always in conflict whether to identify themselves between a home or the outer world; between a world governed by fantasy or reality. The research will try to answer as to how the affinity with myth and mythology resolve this conflict in the characters of the novel. The vantage point from where diasporic writers perceive myths and mythical narratives is different than the understanding of Indian writers. The difference in understanding is reflected in the works of revisionist writers who write from the distance, as they will be above the prejudices that bind them and will be

Copyright © 2020Author

critical to those myths or mythological characters that are deified in the 'home' and worshipped on a pedestal.

Discussion:

The Forest of Enchantments is a meditation on the nature of love, on different shades of love and hence different nuances are reflected through different female characters. There is an aspect of love, where Ram is forced to choose between the public role of a king and his private role of a husband and a lover. There is another aspect of love where Kaikeyi out of her blind devotion for her son wrested out the throne of Ayodhya only to be hated by her son for the very same throne. Shades of love can be traced out in the possessiveness of *Rishi*Gautam that out of his fury of temper, Ahalya turned to stone. There is further layering in the nuances of love, when Surpanakha, wild enchantress of the forest whose gravest crime was to desire a wrong man. Mandodari, wife to the legendary demon king, Ravan, her love is forced to watch her kingdom into ruin and her beloved son perish because her husband was obsessed with another woman. The conflicting love is best expressed in Urmila's devotion; her love was enthralled with dual causes of pain where Lakshman had refused her heart's offer in serving his duty of righteousness by following Ram, who was his first duty and abandoned her. The other cause of Urmila's despair was Sita, her motherly sister whom she looked up to for everything before; she left for the forest.

Every female character in the text is the reflection of different shades of love and by expressing their problematic aspects of love, they are expressing their femininity in public and private space; in their apprehensions; dilemmas and turmoils that shape their characters. Sita's story of love can be traced back to the bonds of empathy, compassion, and sisterhood that she shares with each of the female characters. Sita, during her stay at the Valmiki's *ashram,* saw divine visions where she was surrounded by the women, she met in her entire journey from girlhood to her banishment. She saw Sunaina, her mother; Ahalya; Mandodari; Sarama, Vibheeshan's wife; Kaikeyi; Tara; Surpanakha and Shabari. They use to say in dreams, "endure, endure as we do. endure your challenges" (Divakaruni 322). Sita all through her days in

Copyright © 2020Author

wilderness use to ponder the word 'endure' and confronted that does it mean to give up; it did not mean to be weak and accept injustice rather it meant to take the challenges thrown at them and dealing with them as intelligently as they knew until they grew stronger than them because each one of them stood up against all the unjust that was inflicted on them. All that had happened was without their choice, but they had survived through it: "Wasn't that all we could do as imperfect human beings. I could not control what was done to me but my response to those situations was in my own control" (Divakaruni 322).

Chitra Banerjee Divakaruni creates a matriarchal home that is built on the premise of the strength and influence of female bonding. Women in association with other women can rise above the isolation and oppression without falling into the symmetry of any constructs. Divakaruni deploys mythology from the vantage point where gender is not seen as a domain comprising of compartmentalized constructs. Her characters reflect consolidated bonds of sisterhood, without falling prey to any feminist agenda; they encompass various constructs as they are not confined in the perceptions of these constructs. Divakaruni in her works cites attributes from *Bhagavad Gita* and *Manu Smriti* to put forth her argument that the universal human traits though feminine in forms such as *kirti*(fame), *smriti*(memory), *sri*(prosperity), *dhairya*(patience), *medha*(intelligence) and *dhriti*(firmness) are gender-neutral attributes.

The tale of Surpanakha is always thought of as a cautionary tale in our oral culture. She is also called by the name *Kaamarupini*, the desirable one. She possessed an altogether different shade of love that is a desire which is considered as an antithesis to love. Ravan's sister, with whom Sita felt the bond of empathy; innocent and natural; she possessed her beauty like the wild enchantress of the forest. Surpanakha was that arrow that set in motion the chain of events leading to the destruction of Ravan. She is one such character who was deeply misunderstood and often maligned in *Ramayana*. Divakaruni's portrayal of her makes us realize that the story of gods and goddesses or Dharma is not always colored with polar hues, they are more complexed, more layered, and more nuanced. *Ramayana* portrays Sita as a goddess and Surpanakha as a woman consumed by lust with the mutilated face. Divakaruni renders different shades between

Copyright © 2020Author

the extremities of good and bad; evil and great; physical and eternal. She proposed Ram and assured him that with the spell of her wild magic she could provide him with all the pleasures that a man could ever ask for. Ram refused and in return sent a lightning glance to Lakshman, asked her to demand Lakshman's hand as he was alone in the forest and would surely appreciate the female company. In perfect synchrony, they exchanged glances and mocked her, laughing at her desperation. The girl had gone beyond the norms of maidenly behavior offering herself differently to a man she barely knew but was that her fault to express herself without paying much heed to her dignity.

Kaamarupini accused Sita as the obstacle hindering way between her and Ram and then swelling to twice of her size, turned black and monstrous with her red pulsed eyes, hair writhed as snakes and teeth grew into fangs. She rushed at Sita with her fiercer form but Ram indicated his brother and Lakshman released an arrow that chopped off her nose and ears. She looked at Ram with eyes full of disbelief that someone like him could do such a thing to her when she had offered him, love. Divakaruni'sSita empathized with her that living with a mutilated face was no easier than death, especially for a woman who had so badly wanted a mate. That night, Surpanakha's screams were haunting Sita's heart and she confronted the brothers that both of them did have their share of fun at her expense, tossing her proposition from one to the other, and that became the reason of her anger: "Did you need to be so harsh? To mutilate her so horribly? She was just an infatuated girl" (Divakaruni 150).

Valmiki's Ramayana described Surpanakha as an ugly woman with a potbelly, cross eyes, and grating voice that is harsh on the ears. She had oversized breasts with a heart full of wickedness. That is not the just description with a darker shade rendered to her by Valmiki and therefore various poets and scholars differ drastically in their descriptions of Surpanakha from Valmiki describing her as a beautiful woman with beautiful fish-shaped eyes and thus naming her as Minakshi. Divakaruni's*Sitayan* asserts that she possesses magical powers and can assume any shape and form as she puts these powers to good use when she first proposes Ram. Sita in her dream that night viewed mutilated Kaamarupini knelt in front of an enormous, jeweled

Copyright © 2020Author

throne in a courtroom. It was this very act of chopping out her senses that provoked Ravan to kidnap Sita as an act of revenge but there are other versions of *Ramayana* that assert that Surpanakha had no real interest in brothers. She had orchestrated the war between Ram and Ravan to seek revenge against her husband's murder, as she knew about Ram and his divine powers and felt that Ram could be a perfect match for the might of Ravan and thus pitching the two in battle through her designs. The result of which was she orchestrated her encounter with Ram; the abduction of Sita by Ravan and the resulting war between them and the sole intention was of having her brother killed.

Surpanakha wasn't submissive in her character, she had powers but she didn't like being made a fool of it. Thus there was a woman who was caught up between the swirls of beauty and coyness but here the writer questions as to why our *shastras* have no major description of Surpanakha? Why was she insignificant or for that matter rendered an ugly lusty experience? Many critics oscillate in their perceptions of her as either a liberated feminist of her times or merely a victim. After Ram's coronation upon returning to Ayodhya, his sister Shanta arrived to greet Ram and Sita. Shanta retreated into her mother's chambers to sleep and Surpanakha in the guise of Shanta have come to Ayodhya, following Sita from Lanka. Instead of ears, there were gaping holes. She first asked Sita about Ravan and about the days of captivity and how he fell in love with Sita. Then with a piece of chalk and under a magnificent spell, Surpanakha in the guise of Shanta persuaded Sita to draw a portrait of Ravan on the floor. The chalk glided over the stone and with the spell of magic, moving rapidly by itself, the portrait of Ravan was completed in its entirety. The demon princess, full of hate and rage gloats that Sita would be rejected in a similar way as she herself was rejected, and that's the only way when she would get a justice:

> You suffer as I do, bereft of the status as I was bereft of beauty. Tell me, Sita, did you think you'd get away so easily? Did you think I'd give up and forget about you who had destroyed my home and everyone I loved? Did you think there were no consequences? That you could cause the devastation of Lanka behind you; my people killed; their spirits are broken and you can live happily ever after in Ayodhya, loved by your husband, your family, your subjects, and your sons? I

Copyright © 2020Author

> will not let you escape so easily. Live on Sita. Live without love, as I have been forced to. See how it feels to lose everything you cherish. I will give you a little hint: you will long for death. (Divakaruni 295)

Mandodari, the wife of Ravan, whom he truly loved suffered from the problematics of love mingled with pain and horror. Ravan had countless wives and concubines in Lanka for his intoxicating pleasures; he had been obsessed with Sita to the point of ruination but Mandodariwas the wife of his youth. She continued to love Ravan with all her heart and soul no matter how many women he brought to Lanka. But this river of love followed a complicated course, where one tide of love was in conflict with the other. Pangs of Mandodari's conflicting love were because of the myth which she used to believe so firmly that Sita is her daughter. She gave birth to a girl who would be the cause of her father's death and the ruin of his entire kingdom and hence Ravan ordered the baby to be killed. The night before the girl was to be taken away, Mandodari managed to bribe the rakshasa who was given the duty of killing the baby. The rakshasa promised her that he would leave the baby somewhere far away and safe. Mandodari's conflict lied in either saving her husband, the mighty ruler of the three worlds, and his kingdom or protecting her daughter from the wrath of her father. With the raging pangs of love, she wrapped the baby in beautiful yellow silk which she had woven as a good luck symbol and handed her to the asura. Days and nights, she was torn from the quarrel within herself all those years until Ravan brought Sita to Lanka.

The rakshasa took the infant to a distant land of Mithila, where Janak found the girl from the furrow of the earth and brought her up but he never disclosed the place to the queen for her safety. Ravan was certain that the ill-fated girl was dead but Divakaruni portrays Mandodari's dilemma, her apprehensions, her conflicts in a subtle and nuanced way. During Sita's days of captivity, Mandodari began to watch her in the Ashoka grove at nights and the more she watched her, the more convinced she became that Sita was her long-lost daughter. The bravery with which Sita refused to obey Ravan and enter his palace; he could have killed her for this refusal; the stubborn way of tilting her chin; the way she slept, all her ways were similar to that of the

Copyright © 2020Author

Queen of Lanka. Mandodari love for Sita delighted her but at the same time terrified her for if Ravan would find that Sita was the cursed daughter, he would kill both of them:

> I am too frightened of what he will do in retaliation for such a great crime by going not just against him but endangering the whole country. You haven't seen Ravan in one of his rages. When it happens, even the earth trembles and the gods hide their faces. If he punished me with death, I could endure it. But he would consider that too easy. He won't care that I am his queen. He will put me in his dungeons. Please forgive me, my daughter, and let me make your life easier for you. (Divakaruni 194)

There was a reluctant sympathy in Sita's heart when she heard of Mandodiri's story which annoyed her all the more. The shade of Mandodiri's love for Sita was filled with anguishes, agony, and most of all with the guilt. The greatest guilt of watching and wondering fate's predestined occurring; unable to speak her suspicions to her husband, who was infatuated with a woman who might be his daughter.

The*Sitayan* gives us an insight into the saga's of unsung women characters who were enmeshed in the paradoxical love in such a way, that their sacrifices were never highlighted; their glories were never sung; their tales were never foretold. One such character was Sarama, Vibheeshan's wife who was Sita's only strength in the Lanka of sorrow and fear. Sarama visited Sita in Ashoka grave where they walked around the gardens together, told Sita stories of Lanka as one of the magical places that Ravan brought it from Kuber. Sarama's magic on Sita was stronger than rakshasis magic. Sarama's love for Sita was royal in innumerable ways. She had every reason to hate Sita because she was the cause of all her troubles but strangely Sarama loved her the most amidst all the fiercer rakshasis. Sarama had her worries as Ravan had accused Vibheeshan in the public court for he was not a proper asura because he kept entreating Ravan to send Sita away. Vibheeshan was too sympathetic to humans; he uses to worship the enemy gods in secret and most of all he was a devout *bhakt* of Lord Ram, Ravan's enemy. Sita was the only person in Lanka, with whom she could share all her worries.

Copyright © 2020Author

Sarama was enmeshed between the bonds of love of a husband and her son, Taranisen. One was on the side of Ram and the other was a Ravan's follower. Taranisen was a boy of tender age, on his way of becoming a proficient warrior. Son of Vibheeshan, who was attached to Ravan, was loyal to him and the one who fought against his own father in the war between Ram and Ravan. Divakaruni presents the contradictions of love in such a finest way in her characters that how it filled them with joys but also with worries for the welfare of the loved ones and pain for their sufferings. Sarama's turmoil is best expressed in the lines:

"Taranisen has been asked to lead the army tomorrow. . . I could imagine the turmoil in her. One one side, her husband and on the other, her only child. And in front of her, the woman responsible for all her troubles" (Divakaruni 222).

The war became a tragedy for Sarama's motherhood as the finest warrior like Taranisen who had been trained by Vibheeshan himself; fought marvelously but Ram called on the *Brahmastra*and killed him. Vibheeshan, his father was the only one who knew the secret of his death and it was the paradoxical love that in the *dharmayudh*, a father was responsible for the death of his own son. Sarama never forgave her husband for this betrayal after Ram won the war and Vibheeshan was made the king of Lanka.

Love has many faces and it can lead beings along the twisted paths and crossed trajectories. It made us ready to wreak havoc on the people whom we cared for; it can lead to ruination in the lives of those whom we cherished more. Wasn't it the same force that impelled kaikeyi to turn on Ram to guard her own son's interest? Wasn't it the same force that led Manthara, the hunchback, the orchestrator of the entire disaster who whispered directives to Kaikeyi whom she still saw as the innocent child to whom she had devoted her entire mishappen life? If love for Ram had strengthened Kaushalya, making her shine brighter and stronger amidst all odds then it was the love for Bharat that had weakened Kaikeyi so much so that to give her son zenith of happiness, she had turned his and her life into a state of ultimate despair.

Kaikeyi, daughter of Ashwapati was the favorite queen of Dasharath. The king was captivated by her beauty, intellect, and warrior-like courage. Kaikeyi had been trained in the art

Copyright © 2020Author

of war that whenever there was a battle, she drove Dasharath's chariot. Dasharath won the battle which further strengthened her hold over him. She drove his chariot in other wars and twice saved his life because of which the king was indebted to her. Kaikeyi was the only woman in Ayodhya who transcended the norms of femininity by practicing sword fighting; by assisting Dasharath in the matters of state. There was a symbol of the leaping tiger in her elegant and opulent chambers. The symbol epitomized a woman who knew her power and allure and was willing to make the most of them. Kaikeyi challenged Sita to formal duel in order to judge her martial art skills. Sunaina had pushed against tradition so that her daughter learned to defend herself. The result of it was that even Kaikeyi, an experienced warrior had given up in front of Sita's skills. Kaikeyi oozed trouble from every pore of her beautiful and carefully maintained body.

She had an anger chamber in her quarters which was seen as unwomanly and barbaric to the other two queens. Whenever she wants Dasharath to fulfill her demands, she threw off all her ornaments and silks and refused to eat or drink anything until he went to the chamber and persuaded her to come back. Kaikeyi's two boons, firstly that her son Bharat be made the king of Ayodhya instead of Ram and secondly that Ram should be banished to the forest for fourteen years had wreaked havoc in the life of every character. Love of Kaikeyi was full of contradictions, she loved Ram with genuine fondness but overnight was ready to wrest his kingdom away from him and pushed him into a dungeon of darkness and danger. Divakaruni through the portrayal of Kaikeyi questions her readers once again about the tides of conflicting love; that could love, which we have taken to be powerful and everlasting, be so frail as well? Could the pangs of eternal love be fulfilled with the succession of a throne or by accomplishing material gains? Could we pluck it out of our heart as easily as we would pull a weed of flowers? The love that defines our being can certainly be not that shallow. Divakaruni through the vivid description of her characters left a series of perturbing questions and the answers of which lie in further problematizing the nuances of love.

Chitra Banerjee Divakaruni through her modern-day retelling of the *Ramayana* is compassionate on the thoughts and actions of its myriad women characters. They are painted as

Copyright © 2020Author

strong, relatable, and inspirational characters who are no longer meek, servile but a rebel with warrior instincts. Sita, as she begins to write verses: "Everything was about to change" for Divakaruni's rendition of women characters is a comment on the engulfing silences of duty, betrayal, and how love becomes a medium to balance the two. The retelling of Sita's story is incomplete without Sunaina, Urmila, Mandodari, Sarama, and Surpanakha. Few of them were portrayed as a villain with a cautionary tale and the others were brushed off to the margins. They got their due interpretation in various works by the revisionist writers; in many retellings of the folk epic, the long lost women characters come to life, claiming their own lives, reinventing, and rephrasing them. The writer delves intensely into their unexplored selves and lays their beauty out. It is not that Divakaruni rendition of women characters is portrayed with brighter hues but her men are also treated with equal thoughtfulness. Her Ram is also a complex being as a noble, earnest, devout husband but beset by challenges, he is forced to choose between *Raj Dharma* and his role as a husband. She has filled the void between the adventures undertaken by the male characters in the epic, their triumphs, and their defeats. After reading this retelling with an unprejudiced perception, one can find a gradual fondness for Ravan for his wisdom, intellect, sovereignty, and most of all for his respect for Sita. This study aims to change our polar, biased perceptions for characters like Kaikeyi, an accomplished charioteer; Surpanakha, wronged by two men and Mandodari, brushed off as a demon. The author worked effortlessly to remove such prejudices and biases from the reader's psyche. Be it, Ram or Sita, all are portrayed with human colors; human emotions; human failings; with quirks and fallacies.

We as readers will absorb the retelling in the way of our own experiences and worldviews. There are traces of all perspectives, an attempt to examine the story from multiple new angles; interpretations, and explanations that are both compelling and enlightening. Divakaruni's style of oscillating between past and present; reality and fantasy by picturing the concept of divine dreams is similar to magic realism, a powerful weapon which she deploys in almost all her works. Her language is lyrical, pleasing to the senses but also carries its own concreteness and strength. Apart from her stylistic flair, her major strength is the unraveling of infinite dimensions of a character. Myriad of shades are illustrated in Sita's character as a dutiful

Copyright © 2020Author

but bold daughter; a protective and consoling sister; a conservationist; loving yet willful wife; a perfect helpmate; a skilled healer; a medical woman; a lover; nurturer and adventurer; a kind yet firm daughter-in-law and a courageous fighter. She enjoys in her being, mindful of happiness and grief; of empathy and discernment, and is full of dignity not only for herself but for others. When Ram says to his people, "In my kingdom, every man will have a voice" (Divakaruni 232). Sita wanted to ask, what about women . . .? because not all women are meek and helpless. In these finer expressions, Divakaruni highlights that the destructive masculinity not only endangers Sita and other women but somewhere affects Ram.

Sita's vindication to other women is explicit in her redeeming assertions at the unjust treatment of a woman. In contemporary times, it directly parallels with survivors of abuse and discrimination. Divakaruni's portrayal aptly raises pertinent questions on racism, casteism, sexism, and inequality. Myth, divine dreams, symbols, and forebodings are placed in the violent cartography of present-day scenarios so that the readers aspire the emancipation, royalty, willingness of the characters. Divakaruni maintains an air of mystery and suspense in the text adding with shocking encounters. The sustained and controlled fervor of her storytelling is filled with vibrant imagery and keeps the readers captivated.

The Forest of Enchantments is a work of pluralities and possibilities with a rich taste of thoughtfulness. The text has an aura that surrounds us and makes us grateful to be living in the times of post-truth, pluralities, and diverse interpretations. Thus after reading *Sitayan,* one should imbibe Sita's courage and will to express herself; to baffle amidst the chaos and to need one's own integrity when the situation demands you to say "enough of it. Now no more. . ." (Divakaruni 357).

Works Cited and Consulted

Agarwal, Beena. *Women Writers and Indian Diaspora*. Authors Press, 2011.

Divakaruni, Chitra Banerjee. *The Forest of Enchantments*. India: HarperCollins, 2019. Print.

Copyright © 2020Author

Srivastava, Noopur, and Payel Dutta Chowdhury. "Myths and Mythical Narratives Based on Indian Epics *The Ramayana and The Mahabharata*: A Study through the Revisionist Texts." *Vedas's Journal of English Language and Literature JOELL*, Vol.5, no. 1, 2018, pp. 171-178.

Taskeen, Shaista. "Punctuated Margins and Womanist Valences in Chitra Banerjee Divakaruni's *The Forest of Enchantments*." *The Criterion: An International Journal in English* Vol.10, no.5, 2019, pp. 39-49.

Copyright © 2020Author

Sentiment Analysis using Machine Learning and Deep Learning

Yogesh Chandra
ISSA, DRDO, Delhi, INDIA
yogesh@issa.drdo.in

Antoreep Jana
Delhi Technological University, Delhi, INDIA
antoreepjana@gmail.com

Abstract – **With the increasing rate at which data is created by internet users on various platforms, it becomes necessary to analyze and make use of the data by the Defense and other Government Organizations and know the sentiment of the people. This shall help the organizations take control of their actions and decide the steps to be taken shortly. Added to it, when something crucial is happening in the nation, it is of paramount importance to decide every step without hurting/violating the sentiments of the people. In the era of Microblogging, which has become quite a popular tool of communication, millions of users share their views and opinions on various day-to-day life issues concerning them directly or indirectly through social media platforms like Twitter, Reddit, Tumblr, Facebook. Data from these sites can be efficiently used for marketing or social studies. In this paper, we have taken into account various methods to perform sentiment analysis. Sentiment Analysis has been performed by using Machine Learning Classifiers. Polarity-based sentiment analysis, and Deep Learning Models are used to classify user's tweets as having 'positive' or 'negative' sentiment. The idea behind taking in various model architectures was to account for the variance in the opinions and thoughts existing on such social media platforms. These classification models can further be implemented to classify live tweets on twitter on any topic.**

Keywords- -twitter, sentiment analysis, polarity, machine learning, deep learning, LSTM, CNN

I. INTRODUCTION

Sentiment analysis is the field dealing and analyzing people's opinions, sentiments, and attitudes, behavioral responses to certain events or incidents based on the written language available. It has become one of the most active areas in research due to surge in the fields like machine learning and deep learning, and the blending of such areas with the earlier used statistical methods of Natural Language Processing.

Sudden bloom in sentiment analysis comes with the growing interest and active participation in social media such as reviews of various topics and arts, forums discussions about prevalent issues, blogs and micro-blogs about opinions and information sharing, Twitter and social networks talk about something which is trending in the local area or globally. With the growth in users using digital media to communicate and express themselves, organizations and individuals can extract out the information needed from the opinions and sentiments of the users[1].

As per dsayce website, every second, on average, around 6,000 tweets as reported in November 2018[2]. Twitter user in Japan set a new world record of 143,199 tweets per second (TPS) by tweeting "balus" during a television broadcast of Hayao Miyazaki's anime classic Castle in the Sky on Aug 2.

Also, in the year 2016, Every second, on average, around 6,000 tweets are tweeted on Twitter, which corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year.

This paper covers techniques and methods that can be implemented to directly enable opinion-oriented information-seeking systems. For this paper, python library named "tweepy", "scikit-learn", and framework "keras" was also used to test as per user's will on live sentiments and analyze the present state about any issue or topic.

In this paper, we propose effective ways to perform sentiment analysis or opinion-mining using user tweets. The rest of the paper is organized as follows: Section II of the paper presents the related works that have taken place. Section III briefs about the research methodology. Section IV talks about the dataset used for this paper. Section V talks about the experiments conducted and the results obtained from them. The paper is concluded in Section VI by what and how work can be further extended.

II. RELATED WORK

In this section, we discuss the related works that have taken in the field of sentiment analysis anopinion mining and the various ways and techniques used to exploit and make use of themicroblogging data. With the large availability of blogs and social network websites, many researchers have delved into opinion mining and sentiment analysis. Similar brief insights were presented by [18]. The authors described existing techniques and approaches for opinion-oriented information retrieval.

In a paper on Large-Scale Sentiment Analysis for News and Blogs, by [3], They talk about the opinions expressed in newspapers and blogs while reporting on recent events. They assigned a score to each distinct entity in the text corpus indicating positive or negative opinion.

The idea behind it is that news can be either good or bad. For this sentiment identification phase was used and associated expressed opinion with each relevant entity was given, and for

978-93-80544-38-0 /20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:13:06 UTC from IEEE Xplore. Restrictions apply.

relative evaluation with others in the same class, a sentiment aggregation and score phase are used.

The process was in three stages – 1) Algorithmic Construction of Sentiment Dictionaries 2) Sentiment Index Formation 3) Evaluation of Significance. They figured out how sentiment can vary by demographic group, news source or geographic location. Also, the degree to which the sentiment indices predict future changes in popularity or market behavior.

In another paper by [4], to detect sentiment and topic simultaneously from the text, a probabilistic modeling framework has been based on LDA (Latent Dirichlet Allocation), called joint sentiment/topic model (JST). This model is fully unsupervised, unlike other machine learning models that require supervision. The model has been evaluated on the movie review dataset and the preliminary results have shown promising results achieved by JST.

A paper by [5] uses the concept of recognizing contextual polarity in Phrase-level Sentiment Analysis. It first determines whether an expression is neutral or polar and then removes the ambiguity of the polar expressions. Doing this, the contextual polarity is identified by the system from a large subset of sentiment expressions, giving way better results. Sentence level or even phrase-level are required for multi-perspective question answering and summarization, opinion-oriented information extraction, and mining product reviews. The main motive is to be able to successfully pinpoint positive and negative sentiments.

Since the aim of the paper was to improve the percentage accuracy of the classifier models some similar work was performed by [6], method where publicized stream of tweets from Twitter microblogging site are preprocessed and classified based on their emotional content and then, precision and recall are used to perform the analysis of performance in such cases.

Besides the typical machine learning techniques, there exist a plethora of other techniques in sentiment analysis. In [7], lexicon models are used for the description of verbs, nouns, and adjectives. The model achieves to find the different subjectivity relations expressing separate attitudes among the actors in a sentence.

In another work, by [8], they have applied statistical approaches in which they try to differentiate between features having sentiment and features not having sentiment. A set of new feature selection schemes was proposed that would use the Content and Syntax Model to automatically learn a set of features in review document by separating entities that are being reviewed from the subjective expressions that describe those entities in terms of polarities.

Similar to [7] work, [9] an enhanced lexical resource was developed. It is the result of automatic annotation of all synsets of WORDNET according to notions of "positivity", "negativity", "neutral". Sysnet talks about the 'positivity', 'negativity' and objectivity associated with it.

An interesting paper by [10], tried to present a new method for sentiment classification based on extracting and analyzing appraisal groups such as "very good" . A set of attribute values in several task-independent semantic taxonomies, based on Appraisal theory is an appraisal group. Lexicon of appraising adjectives and their modifiers was built using Semi-automated methods to classify the movie reviews using features based upon these taxonomies combined with standard "bag of words" features.

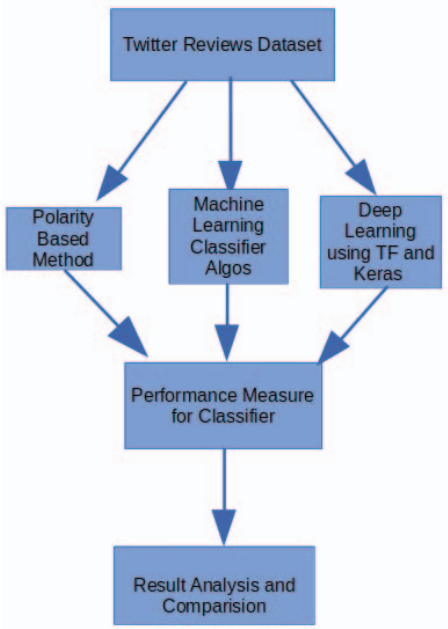


Fig. 1. Workflow for Twitter Reviews Dataset

Deep learning methods can also be utilized to perform a sentiment analysis of a text. In [11], authors proposed a deep convolutional neural network that exploits from character-to-sentence level information to perform sentiment analysis of short texts. Two different domains were used for application: The Stanford Sentiment Tree bank (SSTb), which contains sentences from movie reviews; and the Stanford twitter sentiment corpus (STS), which contains twitter messages.

In [12], exploited Twitter for the purposes of sentiment analysis, and it was done for another time by they approached the sentiment analysis in form of two tasks. The first type takes into consideration a five-point scale, which confers an ordinal character to the classification task. The second type is about the correct estimation of the prevalence of each class of interest, which is called quantification in supervised learning.

Combining various methods to perform a task can yield better results. A paper by [13], shows a new combined method by the combination of rule-based classification, supervised learning and machine learning. The results then showed that a hybrid classification can improve the classification effectiveness in terms of macro- and micro- averaged F1.

A Paper utilizing convolution neural network by[17], utilizes deep convolutional networks to perform sentiment

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:13:06 UTC from IEEE Xplore. Restrictions apply.

analysis in multiple languages without using machine translation and is achieving good accuracy scores. This is done by convoluting on character level embeddings for finding out the polarity of tweets.

III. The Approach

Earlier experiments to perform sentiment analysis on Twitter Dataset were based on polarity-based sentiment analysis and some machine learning classifiers. In this research paper, we have adopted a voting-based classification system and the polarity-based sentiment analysis is done on live tweets fetched using Twitter API.

Then Machine Learning Classifier algorithms were also used to check the performance. Usually, one machine or different machine learning algorithms are used and their performances are compared. We have used a confidence measure by combining several machine learning algorithms.

The models (Machine Learning Models, Deep Learning Models, or Polarity-based methods) are trained using features. The core idea of machine learning is feature engineering.

Machine Learning needs feature extraction and feature engineering. What would happen if that could be automated and that's how we decided to try deep learning models. We used several combinations of deep learning models and compared the relative performance.

IV. Problem Assessment and Tools Description

To perform the research, we have chosen dataset(s) available on Kaggle/Twitter API. For experimental purposes, Kaggle kernels were used.

The datasets used to experiment were 'First GOP Debate' [14], 'Bitcoin Tweets' [15] and 'IMDb movie reviews' and live tweets using Twitter API.

In the training of machine learning models, the features were made by tokenizing the sentences/tweets after performing data cleaning of the tweets. A separate script was made to perform data cleaning of the tweets being received from the twitter API and store them in a dataframe. This can be used by simply importing the python script file and getting the dataframe of the twitter tweets with a keyword as required by the user and the number of tweets/dataset size as mentioned by the user/analyst. Now, after having the dataset, tokenization and lemmatization of the tweets were performed to have tokens of tweets and stopwords removal was done to eliminate the common words. The machine learning classifier models have then trained on these word features of the tweets and the corresponding sentiment associated with a tweet as obtained from a labeled dataset. The predictions were made on the test dataset. The predictions can also be performed from the dataframe obtained from the python script using tweepy to analyze the sentiment on the 'keyword' of user concern.

When working with deep learning models for sentiment analysis, the feature selection and mapping (feature engineering) are performed automatically. Padding is done to maintain uniform input length of all the tweets being passed, the tweets are passed through the deep learning models and training is performed for the tweets. Later on, testing of the models is performed on the test dataset and the performances of various deep learning models were calculated and evaluated. The trained model can be saved and loaded back to save the training time. The model can then be used for predicting the sentiments of new tweets as obtained from the dataframe using the tweepy script file. The input length of the tweets must be adjusted appropriately by using the padding of the tweets.

Classifiers taken for comparison were Naive Bayes, MNB Classifier, Bernoulli Classifier, Logistic Regression, Stochastic Gradient Descent, Linear SVC, NuSVC Classifier. Also, polarity-based methods were used to identify the positive and negative sentiment of tweet. The classification results showed that Deep Learning models using LSTM-CNN gave the best classification results.

V. Experiments and Results Analysis

Using the Twitter API available on apps.twitter.com and various datasets of Twitter on various topics, results were generated.

For comparative study, these results were plotted in a bar graph and results were analyzed.

A custom python script was made which would utilize the tweepy library and fetch tweets from twitter by specifying the keywords, like 'rafale' or any other term associated with the military. The number of tweets/sizes of the dataset required by the analyst can be passed as a function argument. This prepares us with the live test dataset on which we/the analyst wants to have predictions to be made on.

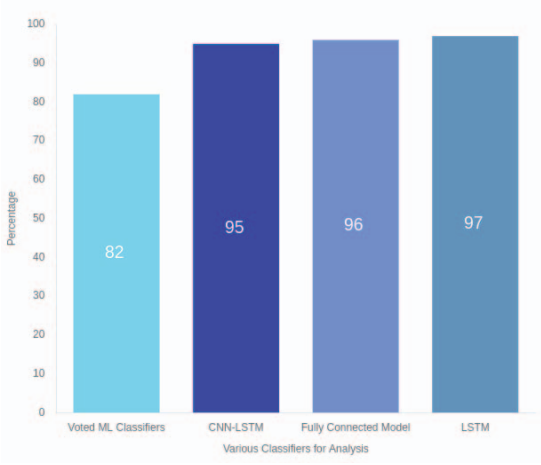


Fig. 2. Cumulative Analysis of various classifiers for Twitter Data Analysis

Various machine learning classifiers can be used to perform training. In our paper, we have used seven different machine learning classifiers and then used a voting system to perform the classification and the associated confidence of the classification is calculated by calculating the ratio of the number of votes (classifications in favour) of a class to the total number of votes (classifications or classifiers). The models are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:13:06 UTC from IEEE Xplore. Restrictions apply.

saved and can be utilized to perform prediction on a dataset that is being loaded from tweepy using the custom python script.

The accuracy percentage for machine learning classifiers lingers around 81 percent to 90 percent for training and testing data.

The deep learning models are fed with padding of the tweets and trained on the training dataset. The trained model is tested on a test set, and the results for training and testing were much better, in the range of 85 percent to 97 percent for training and testing data. The reason being deep learning models like RNN or LSTM remember to utilize the context of the token or the feature which wasn't happening in the case of machine learning models, in which features were simply adjacent words. Due to taking into context the words in the whole sentence in the scenario, the percentage accuracy of the deep learning models was better than the machine learning models. The deep learning models take much longer to be trained (even on GPU) and it is recommended to save the model after training is performed and load the model when required to perform the predictions.

Therefore, we were able to conclude that using deep learning models definitely gives better classification results compared to machine learning classifiers or polarity-based methods on Twitter Dataset(s) and can be reliably used to serve the purpose of sentiment assessment of the crowd which is not possible to do manually when the number of tweets you are dealing with is in millions.

VI. Conclusion and Future Scope

Analyzing the amount of data that is being generated through internet media and microblogging by users of facilities available like Twitter, Reddit, Facebook, etc., and understanding the behavior of people using the internet is useful; it can generate both informative insights and revenue for the business.

In this paper, the tweets data is collected and then passed through machine learning classifiers. After being classified by the individual classifiers, a voted classification mechanism has been used to finally obtain the class of the 'tweet' and the percentage confidence on it. Polarity method for classification has also been used to find the percentage of positive and negative tweets. Lastly, Deep Learning Models have been implied to classify the tweets. Models like RNN, LSTM, CNN-RNN have been utilized to classify the tweets.

Deep Learning models like CNN-RNN, LSTM, etc., and their various combinations have shown better performance compared to the machine learning algorithms. Added to this, Deep Learning Models were trained on various datasets of different domains and the models were achieving very high percentage accuracy for test data in respective domains. Doing this ensures that the final model will account for all the possible variance in the social media. The final prediction to be made by the model will be a voted system of all the models.

As a future outline, work can be done to find out which architecture can be designed and implemented so that the Deep Learning Model achieves an accuracy percentage as good as human accuracy. This will lead to further enhancement for this domain of work.

Acknowledgement

Authors convey their gratitude to Director, ISSA for guiding, supporting and allowing to publish and present the paper at the conference.

References

[1] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," LREc. Vol. 10. No. 2010, 2010.

[2] D. Sayce, "David Sayce," [Online]. Available: https://www.dsayce.com/social-media/tweets-day/.

[3] N. Godbole, M. Srinivasaiah and S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs," UVM, 2007.

[4] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," ACM, 2009.

[5] T. Wilson, J. Wiebe and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," ACLWEB, 2005.

[6] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath and A. Perera, "Opinion mining and sentiment analysis on a Twitter data stream," IEEE, 2012.

[7] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," ScienceDirect, 2012.

[8] A. Duric and F. Song, "Feature selection for sentiment analysis based on content and syntax models," ScienceDirect, 2012.

[9] S. Baccianella, A. Esuli and F. Sebastiani, "SENTIWORDNET 3.0: An Enhanced Lexical Resource," Researchgate.

[10] C. Whitelaw, N. Garg and S. Argamon, "Using appraisal groups for sentiment analysis," ACM, 2005.

[11] C. N. d. Santos and M. Gatti, "Deep Convolutional Neural Networks for," ACLWEB, 2009.

[12] P. Nakov, A. Ritter, S. Rosenthal, F.(http://bvicam.ac.in/indiacom/Downloads.asp) Sebastiani and V. Stoyanov, "SemEval-2016 Task 4: Sentiment Analysis in Twitter," ACLWEB, 2016.

[13] M. Thelwall and R. Prabowo, "Sentiment analysis: A combined approach," Science Direct, 2009.

[14] C. D. f. E. library, "Kaggle," 2016. [Online]. https://www.kaggle.com/crowdflower/first-gop-debate-twitter-sentiment. [Accessed 2019].

[15] Suran, "Kaggle," 2018. [Online]. https://www.kaggle.com/skularat/bitcoin-tweets. [Accessed 2019].

[16] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," Springer Link, 2012.

[17] Wehrmann, Joonatas, Willian Becker, Henry EL Cagnini, and Rodrigo C. Barros. "A character-based convolutional neural network for language-agnostic Twitter Sentiment Analysis", IEEE, 2017..

[18] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in Information Retrieval 2, no. 1–2 (2008): 1-135.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:13:06 UTC from IEEE Xplore. Restrictions apply.

Standoff Quantum Cascade Laser Coupled Acoustic Sensor for the Detection of Hazardous Chemicals

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, Delhi - 110042, India[2]

Abstract:

A standoff photo acoustic detection of explosives using tunable quantum cascade laser (QCL) were studied for a distance of 0.5 meter. Two samples, that is Penta-erythritol Tetranitrate (PETN), and Trinitro-toluene (TNT) were used and studied. The laser beam is focused to the target molecules which are adsorbed on a cloth/plastic bag surface. Photo acoustic Sensor is demonstrated with ultrasonic frequency of ~25 kHz microphone, resonator cavity in form of a cylinder, and a parabolic acoustic reflector. Photoacoustic results are compared in three detection experimental conditions for enhancing the signals with resonator cavity, and a parabolic reflector. Standoff photo acoustic theoretical calculation has been studied. We obtained photoacoustic signals is decreasing inversely proportional with the standoff range. Photoacoustic sensitivity of the sensor is 5µg/cm^2 at standoff distance of ~4.0 meter.

Keywords: *Tunable Quantum Cascade Laser, Ultrasonic Detector, Resonator Cavity, Photoacoustic Spectroscopy, Parabolic Reflector.*

** Corresponding Author: Dr Ramesh C Sharma

Laser Science and Technology Centre, DRDO, Delhi - 110054, India[1]

E-mail: Rameshsharma@lastec.drdo.in

1. Introduction

Over the past four decades, photoacoustic (PA) spectroscopy has been extensively used in numerous applications for example in field of detection of trace gases, environmental sensing and medical diagnostics etc. [1, 2]. Firstly, it was A.G. Bell in the year 1880, who detected the sound waves from the interaction of electromagnetic wave with matter, the PA effect or earlier also known as opto-acoustic effect [3]. The resonator amplification of the PA signal in solid, liquid and gas form was again first described by Bell in 1881 [4]. Terrorist activities possess threat to human life around the globe and these activities have increased in past few year [5]. So there is a demand of early detection of hazardous chemicals, nerve gases and explosives to prevent the loss of human beings. Standoff PA technique has emerged as a successful method in detection of chemical explosives. Several scientists have described the sensing of high energy molecules, hazardous chemicals and explosive materials using Mid – IR vibrational spectroscopy [6-13]. In recent research works QCL has attracted much more attention of researcher due to its better-quality output power, mode quality and room temperature operations as compared to other diode lasers [14-17]. QCL based has been made new development using PA spectroscopic technique for the study of post blast materials [18]. Detection of Standoff Ultrasonic wave using QCL- PA spectroscopy has been made by [19-23]. In the present work we have studied for the detection of hazardous molecules on adsorbed surfaces plastic bags and cloths bags. The main advantages of LPAS include high selectivity, simple setup, immune to other interferences, easily portable, and real time data analysis.

The aim of the present study is to detect the suspicious objects containing hazardous chemicals and explosives adsorbed on the surface by screening them at strategic places for Defence and national security/safety. Quantum Cascade Laser is used as a source, emitting electromagnetic radiation in mid infrared wavelength region. An ultrasonic microphone is employed as a standoff sensor coupled with resonator tube of cylindrical shape. The sensor has peak sensitivity at 25 kHz. We have also enhanced the acoustic signal collection efficacy by using resonator cavity coupled with the acoustic reflector. The system is capable to detect < 5.0 µg/cm^2 at a distance up to 4.0 meters.

2. Experimental

A schematic of experimental setup for Standoff laser photo acoustic sensor, adsorbed on cloth/plastic surfaces (Suspicious objects) is shown in Fig 1. The microphone is used as a PA detector. A Quantum Cascade Laser source operating in the tuning wave-number range 1130 cm^{-1} to 1420 cm^{-1} (i.e. 7 µm to 9 µm in wavelength) commencing M/s. Daylight Solutions;

model no. Uber Tuner (UT-8). It can operate at a small step size of 1.0 cm^{-1} is used. A function generator is used to modulate the laser beam of QCL at 25 kHz with pulse duration of 500 ns. The average peak power of laser beam was nearly 600mW running entirely through the tuning range. Lock-in-Amplifier is provided with a reference signal from the function generator. The laser beam is made to fall on adsorbed cloth/plastic surface which results in absorption of electromagnetic radiation followed by non-radiative decay as a result pressure waves are generated. These waves further results into the formation of acoustic waves which are sensed by microphone. The microphone used has a narrow frequency response which is centered at ~25 kHz. The acoustic signals captured by microphone are then transferred to pre-amplifier. After amplification the signals are feed to lock-in-amplifier card. USB 4716 Advantech data acquisition card is used to digitize the analog signal from lock-in amplifier card and the data is traced in a Lab View program based GUI. Samples of TNT and PETN are prepared using acetone and the dried uniform solution is adsorbed on cloth/plastic bag surface. The concentration of sample 50µg/cm^2 is deposited on surfaces (Plastic and Cloths). Experimental spectra are recorded in the three categories using Standoff laser photo acoustic Sensor adsorbed on cloth/plastic surfaces; (1) only used microphone (2) Microphone with resonator Tube and (3) Microphone with resonator tube coupled with parabolic acoustic Reflector. A schematic of experimental setup for Standoff laser photo acoustic sensor, adsorbed on cloth/plastic surfaces containing microphone at one end of the resonator tube placed at the focus point of parabolic reflector is depicted in Fig 1. The aperture diameter of parabolic reflector is 25 cm.

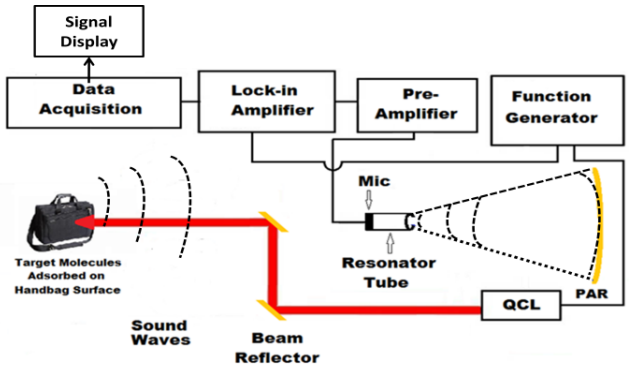


Fig. 1: Schematic of experimental setup for Standoff laser photo acoustic Sensor for adsorbed on cloth/plastic surfaces with microphone, resonator tube and parabolic acoustical reflector (PAR).

3. Photoacoustic Theoretical Calculations

Development and Design of Cylindrical Tube made up of stainless steel material acting as an Acoustic Resonator for Standoff Detection at 25 kHz (Ultrasonic Frequency): Simulations 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 are carried out. A stainless steel cylinder tube is developed with a length of 2.40 cm and inner diameter of 1.70 cm [21]. 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 *(f)* of acoustic resonator open cylinder can be obtained theoretically by equation (1) as written,

$$f = n.s/4L \qquad (1)$$

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 resonatorusing Finite-element based method. The Eigen modes and Eigen frequency are calculated by solving the following partial differential equation,

$$\nabla^2[p(r)/\rho] + [(\omega/s)^2.(p(r)/\rho)] = 0 \qquad (2)$$

where $p(r)$ is the acoustic pressure and is given by,

$$p = p_o\, exp(i\omega t) N/m^2 \qquad (3)$$

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 [21 -23]

$$p_0 = \Gamma \alpha_{abs} F, \qquad (4)$$

where, p_0, Γ are the local pressure wave and Gruneisen parameter of target sample, α_{abs} is the absorption coefficient of sample, and F is the local light fluence. The Gruneisen parameter Γ 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 can be defined by the relation $\sim v_s^2 = (1/ k.\rho)$,

$$p_0=(\beta/k.\rho.c_p).A_e=\Gamma.A_e \quad (5)$$

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 the equation by Ref [22,23]

$$P\,(r) = (p_0 / r).\, exp\,(-8.68.\alpha_{atm}.\, r) \quad (7)$$

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 photoacoustic generated pressure amplitude, reaching at a distance *'r'* from the source is having a *'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) photoacoustic signals with standoff range are plotted in Figure 5 at resonant ultrasonic frequency 25 kHz. PAS signal is achieved up to 4.0 meter distance at modulation at 25 kHz frequency. Parameters are used for PETN and TNT explosive molecules for theoretical calculation is given in Table 1.

4. Results and Discussion

The absorption cross section of explosive molecules like RDX, TNT, PETN, TATP, are $\sim 10^{-18}$ cm^2/molecule in the wavelength region 7.0 to 9.0 micrometer. For the case PETN molecule absorption cross section at wavelength 7.7 micrometer(1286 cm^{-1}) $\sim$ 5.4 x10^{-18} cm^2/molecule. For PETN very strong absorption band is due to (NO_2 stretch + CH bend + C_5 skeletd). 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 spot sizes on a 1.0 cm^2 area of an uniform film of PETN deposited on cloth / plastic surface bags. Pulsed width of laser was 500 ns. The resultant fluence value 'F' is 3.80 mJ/m^2. For PETN, the κ-isothermal Compressibility = (1/11.83) GPa^{-1}, β-Thermal coefficient of volume expansion = 17.4 x 10^{-5} K^{-1}, density ρ = 1620 kg/m^3 and specific heat capacity c= 1250 J/Kg. K, resulting into a Grueneisen parameter Γ~1, using equation 5, [13]. The absorption depth is approximately 0.80 microns at 1286 cm^{-1}. Therefore, for film thickness greater than 1.0 microns, the whole of the laser fluence is absorbed by PETN film. From equation 6, for PETN molecules, we inferred the local pressure rise P_0 = ~ 6,500 Pascal above the atmospheric ambient. The pressure wave propagating from the source can be described to have amplitude at a distance 'r' from it, is as given in equation (8).

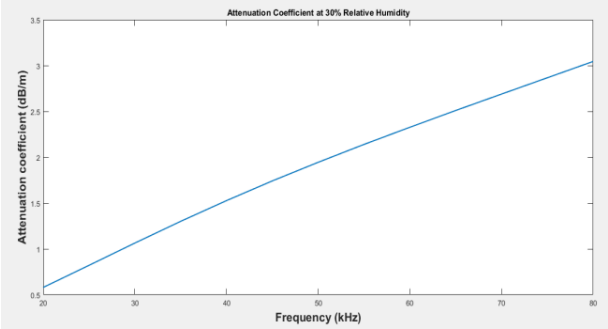


Figure 2: Ultrasonic Frequency dependent atmosphere attenuation of acoustic wave at Relative Humidity (30%). ISO:9613-1:1993

Figure(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 0.821dB/m.

Table 1: Parameters are used for PETN and TNT Explosive molecules for theoretical calculation. Coefficient of Atmospheric attenuation, Alpha (α_{atm}) ~ (25 kHz) =0.821.

Target Samples	Iso-thermal compressibility, k (GPa^{-1})	Thermal coefficient of volume expansion, beta, B (K^{-1})	Density (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

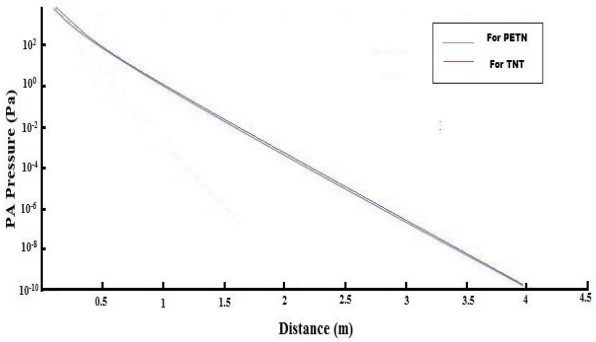


Figure 3: Calculated Standoff Photoacoustic Pressure (Pa) with Distance (m) at ultrasonic frequency of 25 kHz.

The photoacoustical pressure of 10^{-10} Pa can be measured up to distance of 4.0 meters for both the analytes at 25 kHz frequency.

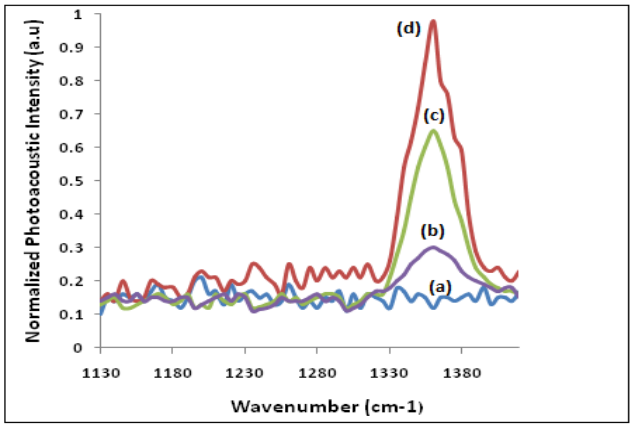


Fig 4: Standoff Photo acoustic Spectra of TNT (a) Background without Sample (b) Only Microphone (c) At 25 kHz with microphone and resonator tube (d) At 25kHz with microphone and resonator tube coupled acoustic reflector (Quantity of each case 50μg/cm² and Standoff Distance of 0.5m for each case)

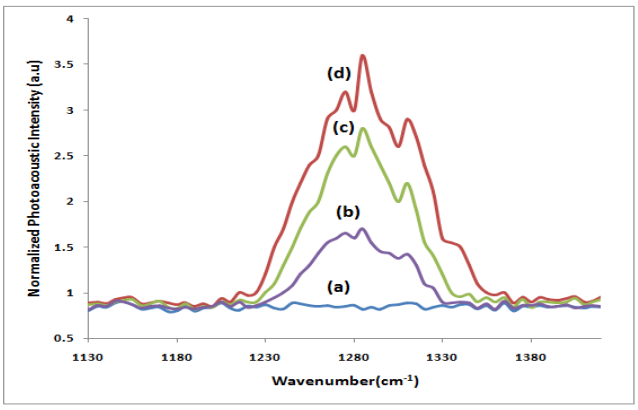


Fig 5: Standoff Photo acoustic Spectra of PETN (a) Background without Sample (b) Only Microphone (c) At 25kHz with microphone and resonator tube (d) At 25kHz with microphone and resonator tube coupled acoustic reflector (Quantity of each case 50µg/cm² and Standoff Distance of 0.5m for each case)

Fig 4 and Fig 5 depicts the standoff photo acoustic spectra of TNT and PETN respectively, the background of the signal is represented by (a) when there is no sample placed on the cloth/plastic bag surface. When the experiment is performed with microphone as one of the component a peak is obtained at 1360cm^{-1} in case of TNT and 1276cm^{-1}, 1286cm^{-1}, 1310cm^{-1} in case of PETN as shown by (b) line. When microphone is used with resonator tube as shown as (c), the photo- acoustical intensity gets enhanced by approximately about twice the signal received in previous setup, since more acoustical waves are collected by the resonator tube. The much better acoustical signals are received depicted as (d) line, when parabolic reflector is used along with microphone and resonator tube. In this setup the sound collection efficacy of the system improved to a greater extent.

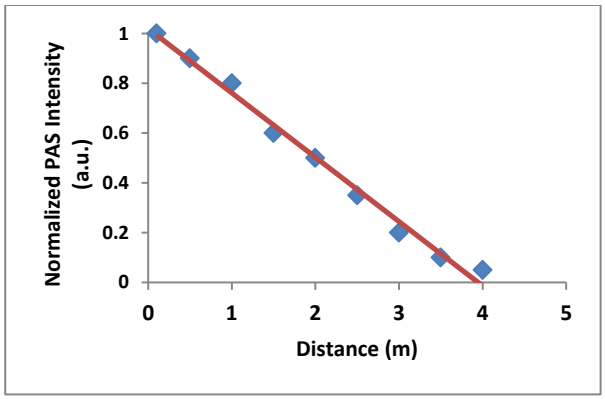


Fig 6: Photoacoustic Signal of TNT in quantity 5µg/cm² adsorbed on surface at different distances.

Fig 6 represents the detection limit of the setup for TNT sample at very low concentration of 5µg/cm² adsorbed on cloth/plastic surface. The dots in the graph represents normalized photo-acoustic signal and the line represents linear fitting of the experimental data. The detection limit obtained experimentally from the graph is up to 4.0 meters at SNR of 3.0.

Conclusions

We demonstrated standoff PA detection of explosives using a QCL as the source, an ultrasensitive microphone as the acoustical detector, a photo acoustic reflecting parabolic reflector and a resonator tube. We presented a simplified model describing pulsed laser induced PA signals on different matters. The study indicated that the PA signal strength is proportional to laser pulse energy and inversely proportional to standoff detection distance. With the help of a parabolic sound reflector, narrow band-pass filters, and amplifiers, the standoff detection distance was extended up to few meters, with a strong SNR. PA theoretical calculation is shown as similar pattern as experimental results. The compact technology is very capable for screening of suspicious objects. PA sensor will be made in handy product for Defence and Security. Future studies include using higher-power QCLs to achieve a longer detection range and trace amount detection capability.

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.

Acknowledgments: The authors are highly thankful to the Dr. Subodh Kumar for his valuable suggestions during all the research work. Mr. Sudhir Kumar, Mr. Umesh Sah and Mrs. Usha Kumari for their help during the experimental work.

References

1. Rosencwaig, A. Photoacoustics and Photoacoustic Spectroscopy; Wiley: Hoboken, NJ, USA, 1980.
2. Pao, Y.-H. Optoacoustic Spectroscopy and Detection; Academic Press: New York, NY, USA, 1977.
3. Bell, A.G, American Journal of Science; XXXIV, 305-324, (1880).
4. Bell, A.G, Philosophical Magazine and Journal of Science; XI, 510-528, (1881).

5. 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).
6. Y. A. gruzdkov and Y. M. Gupta, J. Phys. Chem. A, 105, 6197, (2001).
7. K. McCann and J. Laane, Journal of Molecular Structure, 890, 36, (2008).
8. J. J. P. Stewart, S. R. Bosco, W. R. Carper, Spectrochimica Acta Part A: Molecular Spectroscopy, 42, 13 (1986).
9. C. M. Ramos, L. F. Alzate, N. M. Hernaádez, S. P. Hernaádez, N. Min, Journal of Molecular Structure: Theochem, 769, 69 (2006).
10. F. Pristera, M. Halik, A. Castelli, W. Fredericks, Anal. Chem., 32, 495 (1960).
11. 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).
12. D. L. Perkins, C. R. Lovell, B. V. Bronk, B. Setlow, P. Setlowm and M. L. Myrick, Applied Spectroscopy, 59, 893 (2005).
13. 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).
14. 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 (>10μm), IEEE Journal of Selected Topics in Quantum Electronics, 19, 1200407 (2013).
15. F. Xie, C. Caneau, H. P. Leblanc, D. P. Caffey, L. C. Hughes, T. Day and C. -en Zah, "Watt-level room temperature continuous-wave operation of Quantum Cascade Laser with λ > 10 μm", *IEEE Journal of Selected Topics in Quantum Electronics* 19, 1200407 (2013).
16. A. Talukdar, M. F. Khan, D. Lee, S. Kim, T. Thundat & G. Koley "Piezotransistivetransduction of femtoscale displacement for photoacoustic spectroscopy" *Nature Communications* 6 (2015)
17. 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);
18. 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, (2017) 59-63.
19. Ramesh C. Sharma ; Subodh Kumar , Surya Kumar Gautam ; Saurabh Gupta ; Deepak Kumar ; Hari B. Srivastava; "Detection of Ultrasonic Waves Using Resonant Cylindrical Cavity for Defense Application" IEEE Sensor Journal , 17 , 1681, (2017)
20. 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).
21. Ramesh C. Sharma, Subodh Kumar, Sudhir Kumar, Mohit Mann, Mayank and Mukul Sharma, Spectrochimica Acta Part A: Molecular and Biomolecular Specctoscopy, 224, 117445 (2020)
22. 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 ultrasensitive microphone, Applied Optics 52 (12), 2626-32 (2013).
23. Lihong V. Wang, Tutorial on Photoacoustic Microscopy and Computed Tomography IEEE Journal of Selected topics in Quantum Electronics 14, 1, 171 (2008).

Disclosure: The Authors declare no conflicts of interest.

Thermal Reliability of GaN-BTG-MOSFET for High-Performance Applications in Integrated Circuits

Harshit Soni
Electrical Engineering Department
Delhi Technological University
Delhi, India
harshit_mt2k18@dtu.ac.in

Pranav M Tripathi
Applied Physics Department
Delhi Technological University
Delhi, India
pranavmtripathi@gmail.com

M.M. Tripathi
Electrical Engineering Department
Delhi Technological University
Delhi, India
mmmtripathi@gmail.com

Ajay Kumar
ECE Department,
Jaypee Institute of Information Technology,
Noida, India
ajaykumar@dtu.ac.in

Rishu Chaujar
Applied Physics Department
Delhi Technological University
Delhi, India
chaujar.rishu@dtu.ac.in

***Abstract*—This paper presents, thermal reliability of GaN-BTG-MOSFET (Buffer Trenched Gate MOSFET) for application in Integrated Circuits (ICs) at high temperatures (300K to 600 K).A thorough comparative analysis of electrical characteristics such as transconductance, transfer characteristics, leakage current and the electric field of the designed devices have been performed using the TCAD Atlas tool. A detailed discussion is presented on the thermal stability of the device at high temperatures (300-600K). Paper also presents the performance factors such as On-Resistance (R_{on}), leakage current and threshold voltage (V_{th}). Results suggest that the introduction of GaN in place of silicon in a trenched gate structure not only improves the device's performance at room temperature (300K) but also enhances the thermal stability of the device at high temperatures. The performance of GaN-BTG-MOSFET at high temperature when compared to CTG (Conventional trenched Gate MOSFET) MOSFET suggests that it can be used in ICs and shows better thermal stability than silicon-based devices.**

***Keywords*—CTG-MOSFET, GaN-BTG-MOSFET, Reliability, EOT.**

I. Introduction

This is an era of fast-developing technology. This advancement in the semiconductor industry is supported by innovation where the focus is on high-performance transistors with improved packaging density. Improved packaging density can be associated with a reduction in the size of the transistor which is following Moore's Law [1]. Trenched gate MOSFET is one recent design for effective reduction of device size without compromising with its performance [2]. Nowadays GaN is emerging as an important material for MOSFET because of its high thermal stability than silicon [3]. Further introduction of GaN as the base material in place of silicon improves the performance of trenched gate MOSFET when compared with the conventional MOSFET at room temperature (300K). Continuous usage of the device leads to the formation of hotspots in the device which results in degraded transfer characteristics and may lead to breakdown of the transistor [4]. The performance of a device with increasing temperature can be interpreted from its various parameters like transfer characteristics, transconductance, the -threshold voltage (V_{th}), switching ratio, leakage current, saturation resistance, and Electric field.

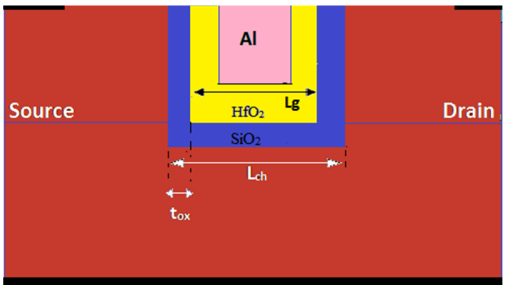


Fig.1. Device structure of GaN-BTG-MOSFET

It is desirable that these parameters be independent of temperature variations or the effect of temperature should be as minimum as possible. With the increase of temperature in thin gate oxide devices, leakage current also increases [5]. Leakage current for small EOT (equivalent oxide thickness) is small [6]. Gate stacking is used to reduce Equivalent Oxide Thickness, a high-K dielectric material Hafnium Oxide (HfO_2) is used to complement Silicon Di Oxide (SiO_2) as a gate stacking material [7]. Gate Stacking results in improved gate control from increased gate oxide capacitance (C_{ox}). In this paper, the performance of GaN-BTG-MOSFET is analyzed and compared with CTG-MOSFET at 300K, 400K, 500K, and 600K. The simulation results obtained from Silvaco TCAD ATLAS [8] show that the performance of the GaN-BTG-MOSFET does not vary much with the change in temperature between 300K to 600K and it can be used in high-temperature devices.

II. Methodology and Calibration

All simulations are done on Silvaco TCAD ATLAS [8] using optimum models which include Lombardi CVT, Fermi Dirac, Shockley-Read-Hall (SRH) Recombination model, etc. to consider real-world possibilities. The simulation models have been utilized by the following equations [8]:

978-1-7281-9713-5/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:14:55 UTC from IEEE Xplore. Restrictions apply.

1. Poisson's equation:

$$\nabla \bullet (\nabla \varphi) = -\frac{\rho}{\varepsilon} \tag{1}$$

where ε, ρ and φ represent permittivity, space charge density and potential respectively.

2. Current continuity equation:

For Holes,

$$\frac{\partial p}{\partial t} = \frac{1}{q} \operatorname{div} \vec{J}_p + G_p - R_p \tag{2}$$

For Electrons,

$$\frac{\partial n}{\partial t} = \frac{1}{q} \operatorname{div} \vec{J}_n + G_n - R_n \tag{3}$$

where, $\vec{J}_p$ and $\vec{J}_n$ are holes and electrons current density; the generation rate for holes and electrons represented by G_p and G_n respectively; R_p is the holes recombination rate and R_n is electrons recombination rate.

3. Recombination

(a) SRH (Shockley Read Hall)

$$R_{SRH} = \frac{pn - n_{ie}^2}{\tau_p [n + n_{ie} e^{(-(E_{TRAP}/kT_L))}] + \tau_n [p + n_{ie} e^{(-(E_{TRAP}/kT_L))}]} \tag{4}$$

(b) Auger recombination

$$R_{Auger} = AUGN\left(pn^2 - nn_{ie}^2\right) + AUGP\left(np^2 - pn_{ie}^2\right) \tag{5}$$

AUGN and AUGP are user-defined.

4. Parallel electric-field-dependent mobility

$$\mu(E) = \mu_0 \left[\frac{1}{1 + \left(\frac{\mu_0 E}{v_{sat}} \right)^{\beta}} \right]^{\frac{1}{\beta}} \tag{6}$$

5. Fermi Dirac and Boltzman Statistics:

Electrons in thermal equilibrium at temperature TL with a semiconductor lattice obey Fermi-Dirac statistics. That is the probability *f(ε)* that an available electron state with energy ε is occupied by an electron is:

$$\frac{1}{1 + e^{\left(\frac{\varepsilon - E}{kT}\right)}}, \tag{7}$$

where E is spatially independent reference energy known as the Fermi level and *k* is Boltzmann's constant.

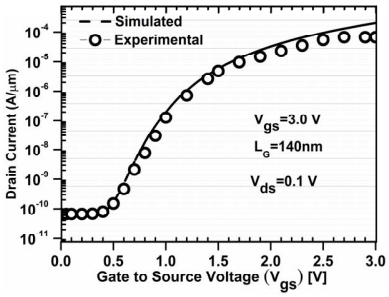


Fig. 2: Calibration of I_{ds}-V_{gs} characteristics of grooved gate MOSFET L_G=36nm by experimental and simulation data.

Further, the simulated device has been validated with the calibration of models with experimental data. A fabricated 36 nm gate length grooved MOSFET with its experimental data and a simulated device with the same dimensions have been taken for the plotting calibrated results (as shown in Fig. 2). It is observed from Fig. 2 that the modeled and simulated results are well-calibrated, thus showing the validity of simulation models.

III. Simulation and Results

Electrical Characteristics of both CTG-MOSFET and GaN-BTG MOSFET are studied with increasing temperature using the Silvaco-TCAD ATLAS tool [8]. Fig.3 shows the Transconductance to the gate voltage of GaN-BTG-MOSFET and CTG-MOSFET at different temperatures. The transconductance of GaN-BTG-MOSFET is higher than that of CTG-MOSFET at 300K, which is because of the higher electron mobility of GaN than silicon [9]. With the increase in temperature transconductance of both the devices decrease, because of the increased scattering of the electrons resulting in reduced transconductances [10]. From the plots in Fig. 3 it can also be seen that transconductance shows higher dependency on the temperature in the case of CTG MOSFET, as compared to GaN-BTG-MOSFET which indicates the thermal stability of GaN as compared to Silicon.

Fig. 4 depicts the drain current with gate voltage at different temperatures for CTG-MOSFET and GaN-BTG-MOSFET. Because of the higher electron mobility of GaN to silicon [2], a higher drain current is observed in GaN-BTG-MOSFET in spite of similar biasing in both devices. With the increase in temperature, scattering in the channel increases [1], and this results in a reduction of drain current at higher temperatures. With the increase in temperature, the change in threshold voltage of GaN-BTG-MOSFET is constant as compared to CTG-MOSFET. This small change in the threshold voltage with increasing temperatures also shows the thermal stability of GaN-BTG-MOSFET with respect to conventional MOSFET. All the electronic devices offer resistance while conducting [11] while referring to Fig. 5, it can be inferred that saturation resistance is smaller than triode region resistance for both CTG and GaN BTG MOSFET. Fig 5 also shows that resistance (R) of GaN-BTG MOSFET is very high in the case of triode, cutoff region and lower saturation mode resistance as compared to CTG MOSFET.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:14:55 UTC from IEEE Xplore. Restrictions apply.

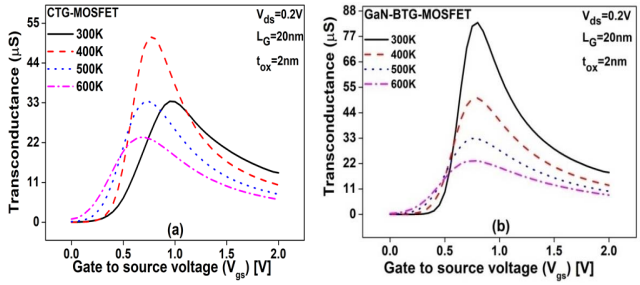


Fig. 3. (a) Transconductance of CTG-MOSFET to the gate voltage and (b) Transconductance of GaN-BTG-MOSFET to the gate voltage.

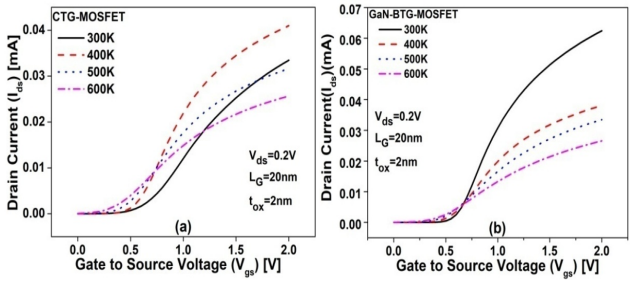


Fig. 4. (a) Transfer characteristics of CTG MOSFET at different temperatures and (b) Transfer characteristics of GaN-BTG-MOSFET at different temperatures.

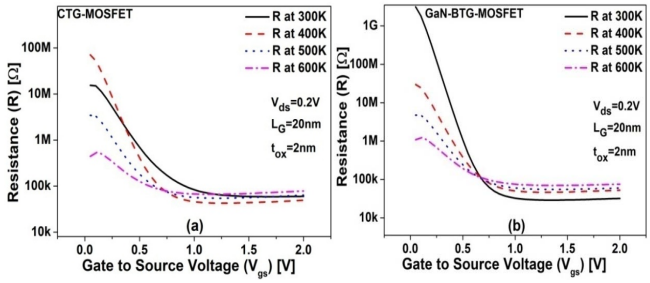


Fig. 5. (a) Resistance of CTG MOSFET at different temperatures, and (b) Resistance of GaN-BTG-MOSFET at different temperatures.

With increasing temperature, ON-Resistance for GaN-BTG-MOSFET increases while that of CTG-MOSFET decreases and this decrement in R_{on} (saturation resistance) with increasing temperature advocates the increase of leakage current in CTG-MOSFET. The continuous increasing resistance for GaN-BTG-MOSFET with increasing temperature results in lower leakage current at High temperature.

Fig. 6 compares the Switching Ratio (I_{on}/I_{off}) of GaN-BTG-MOSFET at different temperatures (300-600K) to CTG MOSFET. The switching ratio is the ratio of I_{on} obtained at V_{gs} = 0.7V and I_{off} obtained at V_{gs} = 0V [12]. At 300K it is observed that the switching ratio of GaN-BTG-MOSFET is 10^3 times that of CTG MOSFET which indicates that GaN-BTG-MOSFET has fast-switching speed with respect to CTG-MOSFET at 300K and lower leakage current which reduces power dissipation and results in a power-efficient device [2]. With the increase in temperature, switching ratio declines for GaN-BTG-MOSFET, following the reduction in drain current while leakage current decreases as shown in Fig.4, Reason behind the fall of switching ratio is the decrease in saturation mode current of MOSFET at High temperatures [2].

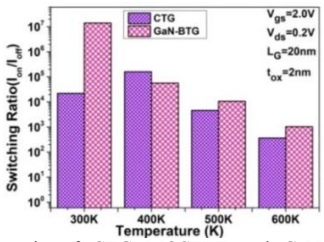


Fig. 6. Switching ratio of CTG MOSFET and GaN-BTG-MOSFET at different temperatures.

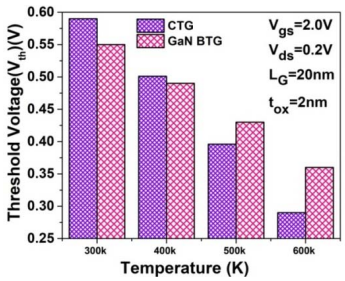
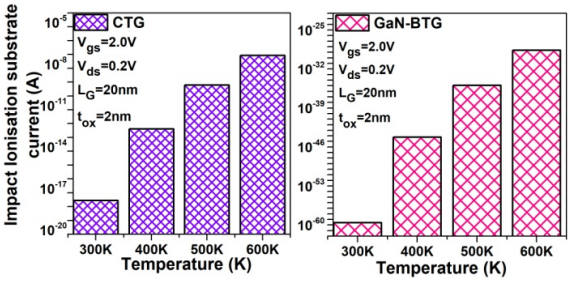


Fig. 7. Threshold voltage (V_{th}) of CTG MOSFET and GaN-BTG-MOSFET at different temperatures.

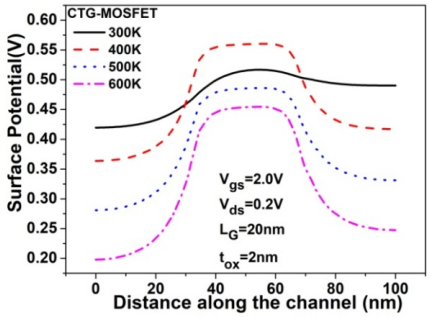
Fig. 7 showcases the threshold voltage (V_{th}) at different temperatures for CTG-MOSFET and GaN-BTG MOSFET. GaN-BTG-MOSFET shows a lower V_{th} as compared to CTG-MOSFET at 300K. With increasing temperature decrement in V_{th} is small in GaN-BTG-MOSFET as compared to CTG-MOSFET, hence GaN BTG MOSFET is thermally more stable than CTG MOSFET. With a smaller deviation in threshold voltage with temperature makes GaN-BTG-MOSFET more efficient in high-temperature applications. Fig. 8 depicts the maximum impact ionization substrate current ($I_{sub,max}$) for GaN-BTG-MOSFET at different temperatures and compared with that of CTG MOSFET. The impact ionization substrate current in the MOSFET occurs when the electron in the channel region injects into the substrate region [10]. At 300K, $I_{sub,max}$ for GaN-BTG-MOSFET is of the order of 10^{-60} and that of CTG MOSFET is in the order of 10^{-20}. This shows that GaN-BTG-MOSFET is more power-efficient. With the increase in temperature, $I_{sub,max}$ increases for both CTG and GaN-BTG-MOSFET. Increment in case of CTG is of the order 10^{-3}, while increment in case of GaN-BTG-MOSFET is of the order 10^{-7}, this relation depicts that GaN-BTG-MOSFET is thermally more stable and more power-efficient than CTG-MOSFET at both room temperature and high temperatures.



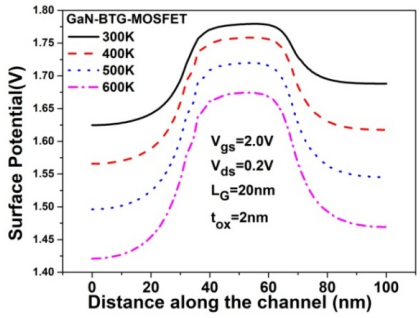
(a) (b)

Fig. 8. (a) Impact ionization substrate current ($I_{sub,max}$) of CTG MOSFET at different temperatures and (b) Impact ionization substrate current ($I_{sub,max}$) GaN-BTG-MOSFET at different temperatures.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:14:55 UTC from IEEE Xplore. Restrictions apply.

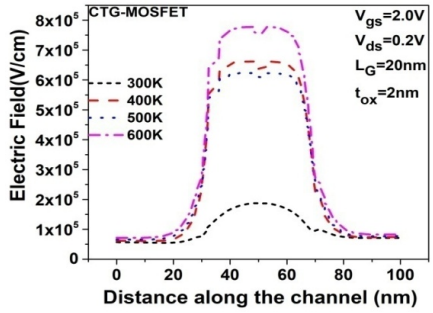


(a)

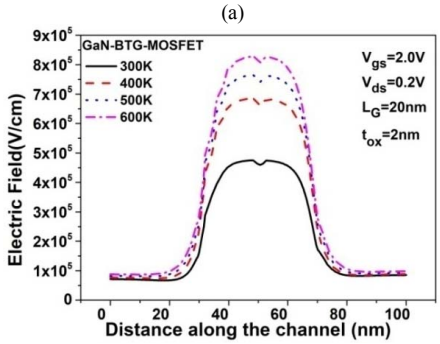


(b)

Fig. 9. (a) Surface potential of CTG MOSFET along the channel at different temperatures and (b) Surface potential of GaN BTG MOSFET along the channel at different temperatures.



(a)



(b)

Fig. 10. (a) Electric field of CTG MOSFET along the channel at different temperatures and (b) Electric field of GaN-BTG-MOSFET along the channel at different temperatures.

Fig. 9 (a) and (b) show the surface potential of GaN-BTG MOSFET along the channel length at different temperatures and compared this to the surface potential of CTG-MOSFET. The figure shows a higher surface potential for GaN-BTG-MOSFET as compared to CTG MOSFET because GaN has a large bandgap of 3.1eV with respect to silicon 1.1eV [9], which results in reduced barrier potential which owes to the high conductivity of GaN-BTG MOSFET. Fig. 10 (a-b) showcases the electric field for both CTG and GaN BTG MOSFET along the channel. Fig. 10 shows that the Electric field for GaN-BTG-MOSFET is higher than CTG MOSFET across the conducting channel region at room temperature. Electric field increases with an increase in temperature for both CTG and GaN BTG MOSFET because of high electron mobility and high saturation velocity [10] at high temperatures.

IV. Conclusion

This paper presents a comparative study of the performances of GaN-BTG-MOSFET with conventional trenched gate MOSFET for high temperature (300K-600K). Comparative analysis of electrical characteristics and performance factors of the devices such as transfer characteristics, transconductance, switching ratio leakage current, threshold voltage, electric field, and surface potential with increasing temperatures from 300K to 600K has been discussed. This paper discusses the thermal stability of the GaN-BTG-MOSFET at different temperatures as compared to CTG-MOSFET. GaN-BTG-MOSFET has very small leakage current and very small saturation resistance which infers that GaN BTG MOSFET is more power-efficient and showcases the possibility of being implied for high temperature and continuously used integrated circuits.

References

[1] N. D. Arora, *MOSFET Models for VLSI Circuit Simulation: Theory and Practice*: Springer Vienna, 2012.

[2] A. Kumar, N. Gupta, and R. Chaujar, "Analysis of novel transparent gate recessed channel (TGRC) MOSFET for improved analog behaviour," *Microsystem Technologies,* vol. 22, no. 11, pp. 2665-2671, 2016.

[3] C. Tsai, T. Wu, and A. Chin, "High-Performance GaN MOSFET With High-$\kappa $$\hbox {LaAlO} _ {3}/\hbox {SiO} _ {2} $ Gate Dielectric," *IEEE electron device letters,* vol. 33, no. 1, pp. 35-37, 2011.

[4] S. Z. Zhao, and R. R. Khan, "Thermal improvement for hotspots on dies in integrated circuit packages," Google Patents, 2015.

[5] J. Jose, A. Ravindran, and K. K. Nair, "Study of temperature dependency on MOSFET parameter using MATLAB," *International Research Journal of Engineering and Technology,* vol. 3, no. 7, pp. 1530-1533, 2016.

[6] C.-H. Choi, K.-Y. Nam, Z. Yu, and R. W. J. I. T. o. E. D. Dutton, "Impact of gate direct tunneling current on circuit performance: A simulation study," vol. 48, no. 12, pp. 2823-2829, 2001.

[7] Y. Ma, Y. Ono, L. Stecker, D. R. Evans, and S. Hsu, "Zirconium oxide based gate dielectrics with equivalent oxide thickness of less than 1.0 nm and performance of submicron MOSFET using a nitride gate replacement process." pp. 149-152, 1999.

[8] I. SILVACO, "ATLAS User's Manual," *Santa Clara, CA, Ver,* vol. 5, 2011.

[9] R. S. Pengelly, S. M. Wood, J. W. Milligan, S. T. Sheppard, and W. L. Pribble, "A review of GaN on SiC high electron-mobility power transistors and MMICs," *IEEE Transactions on Microwave Theory and Techniques,* vol. 60, no. 6, pp. 1764-1783, 2012.

[10] S. Mennillo, A. Spessot, L. Vendrame, and L. Bortesi, "An analysis of temperature impact on MOSFET mismatch." pp. 56-61.

[11] T. Kasuya, "Electrical Resistance of Ferromagnetic Metals," *Progress of Theoretical Physics,* vol. 16, no. 1, pp. 58-63, 1956.

[12] A. Kumar, M. Tripathi, and R. Chaujar, "Reliability issues of In 2 O 5 Sn gate electrode recessed channel MOSFET: Impact of interface trap charges and temperature," *IEEE Transactions on Electron Devices,* vol. 65, no. 3, pp. 860-866, 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 09:14:55 UTC from IEEE Xplore. Restrictions apply.

Circuits, Systems, and Signal Processing
https://doi.org/10.1007/s00034-020-01454-2

SHORT PAPER

Two Quadrant Analog Voltage Divider and Square-Root Circuits Using OTA and MOSFETs

Ajishek Raj[1] · **D. R. Bhaskar**[1] · **Pragati Kumar**[1]

Received: 17 January 2020 / Revised: 9 May 2020 / Accepted: 11 May 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In this communication, two novel architectures of voltage mode analog divider circuit and square-root circuit using an operational transconductance amplifier (OTA) have been presented. The proposed divider circuit employs an OTA and two MOSFETs, while the square-root circuit requires one OTA along with one MOSFET. The proposed divider circuit can also be configured as inverse voltage function generator. The performance of the proposed circuits has been validated through Cadence Virtuoso simulations using 0.18 μm CMOS technology parameters. The total power consumption for analog divider circuit is 821 μW, while for square-root circuit, it is 442 μW with ± 0.9 V power supply. Experimental results have also been provided to validate the theory.

Keywords Operational transconductance amplifier · Voltage mode · Analog divider · Analog computational circuits · Square-root · Inverse voltage

1 Introduction

Analog dividers have a wide range of applications in analog signal processing, telecommunications, and electronic systems [9], A/D–D/A converters, peak detectors, phase detectors, synthesizers as well as in analog computational systems based on biological 'neural' paradigms [27, 38]. Recently, divider circuit designs using numerous active building blocks (ABBs) have received much attention. Various researchers have proposed current mode (CM) [2, 3, 15–17, 19, 20, 24, 25, 28, 29, 34–36, 39, 41] and voltage mode (VM) analog divider circuits [6, 7, 12, 14, 21, 22,

✉ D. R. Bhaskar
drbhaskar@dtu.ac.in

Ajishek Raj
ajishekaerotek@gmail.com

Pragati Kumar
pragatikumar@dtu.ac.in

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

Published online: 26 May 2020

26, 37] in the past using a variety of ABBs such as OTAs [6, 16, 19, 34, 37], current-controlled current conveyors (CCCII) [28], current-controlled current differencing buffered amplifiers (CCCDBAs) [35], current differencing transconductance amplifiers (CDTAs) [15, 39], current-controlled current differencing transconductance amplifiers (CCCDTAs) [36], current amplifier (CA) [17], second-generation current conveyors (CCII) [2, 14, 22], multiple output current-controlled current through transconductance amplifier (MO-CCCTTA) [29], current follower transconductance amplifier (CFTA) [24], current feedback amplifiers (CFA) [7, 21] and operational transresistance amplifier (OTRA) [26]. Square-root circuits, on the other hand, are useful in measurement and instrumentation for linearization of a signal from a differential pressure flow meter, or to calculate the root mean square value of an arbitrary waveform [9]. Geometric mean of two signals can also be found employing square-root circuits [4, 18]. As the proposed work deals with the realization of analog voltage dividers [6, 7, 12, 14, 21, 22, 26, 37] and square-root circuits [5, 8, 10, 11, 23, 30, 32, 33, 40] in VM, in the following, we present a brief description of VM analog dividers and square-root circuits realized with various ABBs.

In [21], CFA-based analog divider was proposed employing two CFAs and four MOSFETs (operated in the triode region). The circuit did not employ any passive components and thus could be integrated easily. An analog divider using voltage variable resistance property of MOSFET was proposed in [12]. The authors presented three analog divider circuits in which the first two circuits employed two Op-amps, six resistors and one field effect transistor (FET), while the third circuit required three Op-amps along with six resistors and two FETs. In [37], nonlinear circuits were proposed using OTAs in which one of them is an analog divider circuit that required one multiplier (using two OTAs) circuit cascaded with an amplifier circuit. OTA-based multiplier was used in [6] to design a divider circuit that required six OTAs and two resistors. An integrable analog divider was presented in [22] using a voltage-to-current converter realized with a CCII and four MOSFETs (biased in the non-saturation region) and a multiplier circuit constructed with two CCII along with two MOSFETs (biased in non-saturation region). CFA-based differential divider circuit was presented in [7] that employed five CFAs, six matched MOS transistors (biased in the triode region) and four identical capacitors. Multiplication mode current conveyor along with a CCII and two resistors were used in [14] to design a divider circuit. In [26], an OTRA-based analog divider was presented that employed a single OTRA and four matched MOSFETs (operated in the triode region) having an active die-out area of 0.0039 mm^2.

In [11], CMOS analog square-root and squaring circuits were presented using an op-amp and two/four MOSFETs. The differential difference current conveyors (DDCCs)-based squarer and square-rooter circuits were reported in [8] employing two MOSFETs, one voltage-to-current converter and a voltage follower. The square-root circuit proposed in [32] uses an Op-amp, six resistors and a diode. This circuit is based on the use of Op-amp supply-current sensing technique. Using OTAs, two square-root circuits (one in VM and another in CM) were proposed in [33]. In the first circuit, four OTAs and one resistor were used to produce VM square-root function, whereas another CM square-root circuit comprised of four OTAs only. Another OTA-based square-root circuit was presented in [40] where three OTAs and two resistors were used. CCCII-based

VM and CM square-root circuits were reported in [10] where two CCCIIs along with three resistors/one resistor were used in VM and CM, respectively. In [23], CCII-based square-rooter circuit was presented employing one buffer, one resistor and two MOSFETs (biased in the triode region). The reported bandwidth of the presented circuit was 400 kHz. Current differencing buffered amplifiers (CDBAs) based two square-root circuits were proposed in [30]. One of them required one CDBA, one resistor; one buffer and two MOSFETs and another circuit used one CDBA and four MOSFETs. The used MOSFETs were matched and biased in the triode region. Low-voltage square-root circuit employing OTRA was presented in [5] that required single OTRA and three MOSFETs (biased in the triode region).

From the above description, it may be noted that the reported divider and square-root circuits use either more number of active and passive components [6–8, 10, 12, 14, 21–23, 26, 32, 33, 37, 40] or more number of MOSFETs [5, 7, 8, 11, 21, 23, 26, 30]. Thus, in this communication, we present a very simple architecture of analog divider that requires only one single output OTA with two NMOS transistors (operating in the triode region), while in analog square-root circuit, a single output OTA with only one NMOS transistor (operating in the saturation region) is used. To demonstrate the workability of the proposed circuits, simulations have been carried out through Cadence Virtuoso (Cadence EDA Tools) using 0.18 μm CMOS technology parameters. Sample experimental results have also been provided to validate the theory.

2 Proposed Configurations

2.1 The Proposed Analog Divider Circuit

Figure 1 shows the proposed analog divider circuit.

The proposed circuit uses one single output OTA and two matched MOSFETs (operating in the triode (linear) region) to obtain analog divider function. The expression for drain current of MOSFET in triode region (neglecting channel length modulation) is given by [1]:

$$I_{\mathrm{D}} = K_n\left\{\left(V_{\mathrm{GS}} - V_{\mathrm{TH}}\right)V_{\mathrm{DS}} - \frac{1}{2}V_{\mathrm{DS}}^2\right\} \tag{1}$$

where $V_{\mathrm{DS}} < \left(V_{\mathrm{GS}} - V_{\mathrm{TH}}\right)$; $V_{\mathrm{GS}} > V_{\mathrm{TH}}$ and $K_n = \mu_n C_{\mathrm{ox}}\left(\frac{W}{L}\right)$, μ_n is the surface mobility, C_{ox} is the gate oxide capacitance per unit area, W/L is the aspect ratio, V_{TH} is the threshold voltage, and V_{GS} and V_{DS} are the gate-to-source and drain-to-source voltages, respectively, of the NMOS transistor.

A straightforward analysis of the circuit in Fig. 1 reveals that the drain currents I_{D1} and I_{D2} are given by:

$$I_{\mathrm{D1}} = K_n\left[\left\{\left(V_1 + V_P\right) - V_{\mathrm{TH}}\right\}V_{\mathrm{out}} - \frac{1}{2}V_{\mathrm{out}}^2\right] \tag{2}$$

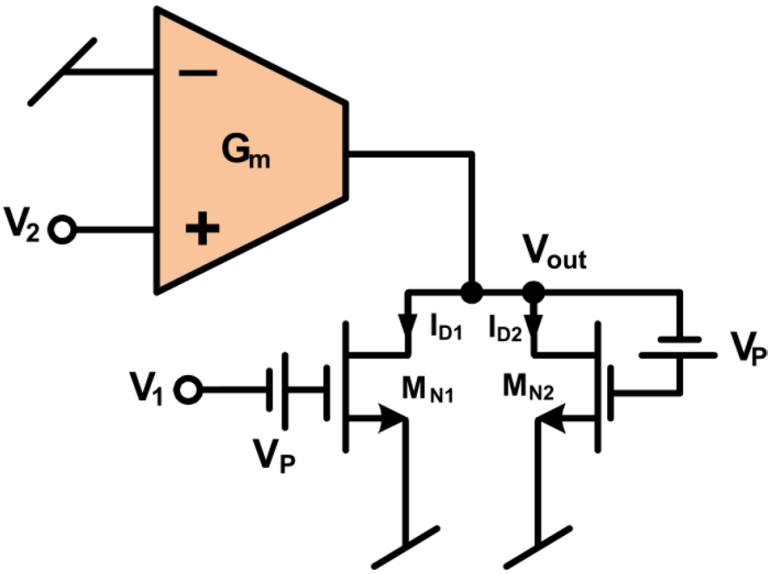


Fig. 1 Proposed new configuration of VM analog divider

$$I_{\mathrm{D2}} = K_n\left[\left\{\left(V_{\mathrm{out}} + V_P\right) - V_{\mathrm{TH}}\right\}V_{\mathrm{out}} - \frac{1}{2}V_{\mathrm{out}}^2\right] \tag{3}$$

From Fig. 1, the output current of the OTA can be expressed as:

$$G_m V_2 = I_{\mathrm{D1}} + I_{\mathrm{D2}} \tag{4}$$

where G_m is the transconductance of the OTA [13] and is given by:

$$G_m = \sqrt{2K_n I_{\mathrm{Bias}}\left(\frac{W}{L}\right)} \tag{5}$$

Substitution of Eqs. (2) and (3) in Eq. (4), the following expression can be obtained

$$G_m V_2 = K_n\left[\left\{V_1 + 2\left(V_P - V_{\mathrm{TH}}\right)\right\}V_{\mathrm{out}}\right] \tag{6}$$

Now if $V_P = V_{\mathrm{TH}}$, Eq. (6) reduces to:

$$V_{\mathrm{out}} = K\left(\frac{V_2}{V_1}\right) \text{ where } K = \frac{G_m}{K_n} \tag{7}$$

where V_1, V_2 are input voltage signals and V_P is the external DC voltage. The floating voltage V_P may be generated as shown in Fig. 2.

From Eq. (7), it can be seen that the output voltage of the proposed circuit in Fig. 1 provides a division of two analog input voltage signals. The output voltage can also be scaled electronically through the transconductance of OTA.

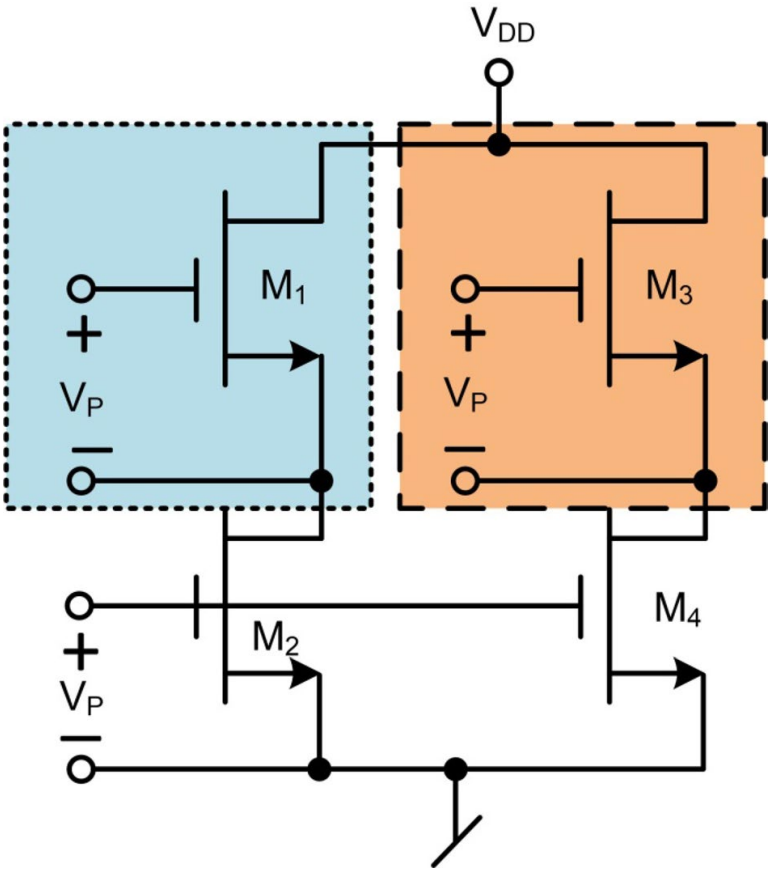


Fig. 2 Generation of floating voltage source [42]

2.2 Inverse Function

The proposed divider circuit can be modified to provide an inverse function. It can be obtained by keeping V_2 constant and making V_1 as an input signal, the output voltage of Eq. (7) reduces to:

$$V_{\text{out}} = K_1 \left(\frac{1}{V_1}\right); \text{where } K_1 = KV_2 \tag{8}$$

From Eq. (8), it may be noted that the output voltage V_{out} is proportional to the inverse of the input voltage signal V_1. In this case too, V_{out} can be electronically controlled through the transconductance G_m of OTA.

2.3 Proposed New Analog Square-Root Circuit

The proposed square-root circuit is shown in Fig. 3.

The proposed square-root circuit employs one single output OTA and one MOSFET. The drain current of MOSFET in the saturation region (neglecting channel length modulation) is given by [1]:

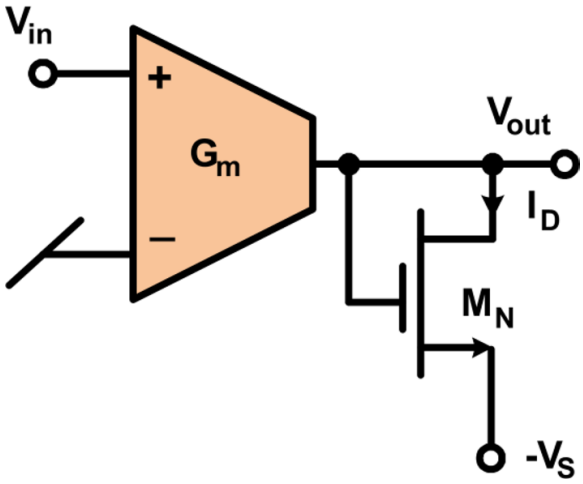


Fig. 3 Proposed square-root circuit

$$I_{\mathrm{D}} = \frac{K_n}{2}\left\{V_{\mathrm{GS}} - V_{\mathrm{TH}}\right\}^2 \tag{9}$$

where $V_{\mathrm{DS}} > \left(V_{\mathrm{GS}} - V_{\mathrm{TH}}\right)$ and $V_{\mathrm{GS}} > V_{\mathrm{TH}}$.

A straightforward analysis of the proposed circuit in Fig. 3 gives the following output voltage:

$$G_m V_{\mathrm{in}} = \frac{K_n}{2}\left(V_{\mathrm{out}} + V_s - V_{\mathrm{TH}}\right)^2 \tag{10}$$

Now if $V_s = V_{\mathrm{TH}}$, Eq. (10) becomes

$$V_{\mathrm{out}} = \sqrt{2\,K\,V_{\mathrm{in}}} \tag{11}$$

From Eq. (11), we can see that the output voltage is proportional to the square root of the input signal voltage which can also be scaled electronically through the transconductance of OTA.

If channel length modulation effect is taken into account, the drain current of the MOSFET will be expressed by [1]:

$$I_{\mathrm{D}} = \frac{K_n}{2}\left\{V_{\mathrm{GS}} - V_{\mathrm{TH}}\right\}^2\left(1 + \lambda V_{\mathrm{DS}}\right) \tag{12}$$

The modified output voltage of the square-root circuit with channel length modulation may be approximated (with $V_s = V_{\mathrm{TH}}$) as:

$$V_{\mathrm{out}} \cong \sqrt{2KV_{\mathrm{in}}}\left(1 - \frac{\lambda}{2}V_{\mathrm{TH}}\right)\left(1 + \frac{\lambda}{2}V_{\mathrm{in}}\right) \tag{13}$$

From Eq. (13), it is noted that the channel length modulation coefficient introduces some error in the output of the square-root circuit. However, the error introduced, because of the channel length modulation effect ($\lambda = 0.094$), when the input voltage V_{in} in Eqs. (11) and (13) was varied between 0 and 100 mV, has been found to be less than 1.7%, which is adequately low.

3 Mismatch Analysis

3.1 Mismatch Between Threshold Voltages of M_{N1} and M_{N2}

The MOSFETs M_{N1} and M_{N2} of Fig. 1 are assumed to be matched, but in practical case, it is very challenging to make threshold voltages of MOSFETs exactly equal. Assuming both MOSFETs M_{N1} and M_{N2} have different threshold voltages V_{TH1} and V_{TH2}, the output voltage of divider circuit may be expressed as:

$$V_{\text{out}} = K\left[\frac{V_2}{V_1 + 2V_p - (V_{TH1} + V_{TH2})}\right] \tag{14}$$

From Eq. (14), it may be noted that the effect of mismatch in threshold voltage of the MOSFETs may be eliminated by selecting appropriate value of the external DC voltage V_P.

3.2 Mismatch Between the Aspect Ratios of M_{N1} and M_{N2}

To consider the effect of mismatch in the aspect ratios of the MOSFETs on the output voltage of the divider circuit, we have defined $\left(\frac{W}{L}\right)_1 = \left(W/L\right) + \Delta\left(W/L\right)$ and, $\left(\frac{W}{L}\right)_2 = \left(W/L\right) - \Delta\left(W/L\right)$. Thus, reanalysing the divider circuit shown in Fig. 1, the following quadratic equation is obtained:

$$V_{\text{out}}^2\left(\frac{\Delta\left(\frac{W}{L}\right)}{\left(\frac{W}{L}\right)}\right) - V_{\text{out}}\left\{2\left(V_P - V_{TH}\right) + V_1\left(1 + \frac{\Delta\left(\frac{W}{L}\right)}{\left(\frac{W}{L}\right)}\right)\right\} + KV_2 = 0 \tag{15}$$

Solution of the quadratic equation with $V_P = V_{TH}$ gives the following output voltage:

$$V_{\text{out}} \cong \frac{V_1\left(1 + \frac{\Delta\left(\frac{W}{L}\right)}{\left(\frac{W}{L}\right)}\right) - V_1\left(1 + \frac{\Delta\left(\frac{W}{L}\right)}{\left(\frac{W}{L}\right)}\right)\left(1 - \left(\frac{4KV_2\left(\frac{\Delta\left(\frac{W}{L}\right)}{\left(\frac{W}{L}\right)}\right)}{V_1^2\left(1+\frac{\Delta\left(\frac{W}{L}\right)}{\left(\frac{W}{L}\right)}\right)^2}\right)\right)^{\frac{1}{2}}}{2\left(\frac{\Delta\left(\frac{W}{L}\right)}{\left(\frac{W}{L}\right)}\right)} \tag{16}$$

Using binomial expansion of the radical term in Eq. (16) and neglecting the higher-order terms in the expansion we get:

$$V_{\text{out}} \cong \frac{KV_2}{V_1}\left(\frac{1}{1+\frac{\Delta\left(W/L\right)}{\left(W/L\right)}}\right) \tag{17}$$

From Eq. (17), it may be noted that the effect of mismatch in the aspect ratios does not change the nature of the output voltage function.

3.3 Effect of Variation in the Supply Voltage on V_p

The effect of variation in supply voltage on V_P in the output voltage expression of the voltage divider circuit may be considered by assuming $V_P=(V_P\pm\Delta V_P)$. Hence, reanalysing the divider circuit shown in Fig. 1, the new output voltage of divider circuit is obtained as:

$$V_{\text{out}} = K\left[\frac{V_2}{V_1+2(V_P\pm\Delta V_P)-2V_{\text{TH}}}\right] \tag{18}$$

From Eq. (18), it may be noted that the voltage V_P must be selected in such a way that the term ΔV_p in parenthesis in the denominator of Eq. (18) is not significant. An exemplary way of minimizing ΔV_p has been presented in "Appendix" section. The effect of variation in V_P on the output has been evaluated through simulation results.

4 High-Frequency Response of Divider Circuit

The high-frequency performance of the divider circuit in Fig. 1 has been investigated by determining the poles associated with (1) the OTA and (2) the M_{N1}–M_{N2} network which is connected at the output of the OTA. The transconductance G_m of the OTA may be modeled by a one pole model [31] as:

$$G_m(s) = \frac{G_{m0}}{\left(1+\frac{s}{\omega_b}\right)} \tag{19}$$

On the other hand, the pole of the M_{N1}–M_{N2} network has been determined by carrying out a small-signal analysis of this network and has been found to be:

$$\omega'_{3\text{ dB}} = \left(\frac{1+g_2(r_{o1}\|r_{o2})}{(C_{\text{gd1}}+C_{\text{ds1}}+C_{\text{ds2}}+C_{\text{gs2}})(r_{o1}\|r_{o2})}\right) \cong \frac{g_2}{C_o} \text{ for } g_2r_o >> 1 \tag{20}$$

where $C_{\text{o}}=(C_{\text{gd1}}+C_{\text{ds1}}+C_{\text{ds2}}+C_{\text{gs2}})$, g_2 is the transconductance of the MOSFET M_{N2} and $r_{\text{o}}=r_{\text{o1}}\|r_{\text{o2}}$. From simulations, the values of $\omega'_{3\text{ dB}}$ and ω_b have been found to

be 53.7 GHz and 4.68 GHz, respectively. Thus, the high-frequency performance of the voltage divider circuit will be limited by the pole of the OTA circuit as $\omega_b < \frac{\omega'_{3\,\mathrm{dB}}}{10}$.

5 Simulation Results

The workability of all the presented circuits has been verified through simulations by Cadence virtuoso using 0.18 µm TSMC CMOS technology parameters. The CMOS realization of OTA is shown in Fig. 4 [13]. The *W/L* ratios used for the implementation of OTA were 3.6/0.36 for *M*1–*M*2, 1.44/0.36 for *M*3–*M*5, *M*8, 2.88/0.36 for *M*6–*M*7 and 5.4/0.36 for *M*9, respectively. The supply voltages used were ± 0.9 V. The transconductance (G_m) of OTA was set in such a way that the G_m/K_n is equal to 1 for both divider and square-root circuits. It may be noted that a voltage follower (buffer) will be required at the output of each divider/square-root circuit to prevent loading.

5.1 Voltage Division Mode

For voltage division mode of operation, the aspect ratios of M_{N1} and M_{N2} were taken as 1.44 µm/0.36 µm. The aspect ratios of all the four MOSFETs used in the level shifter circuit (Fig. 2), used for generating the floating voltage source of 380 mV (equal to V_{TH}), were taken as 3.6 µm/3.6 µm.

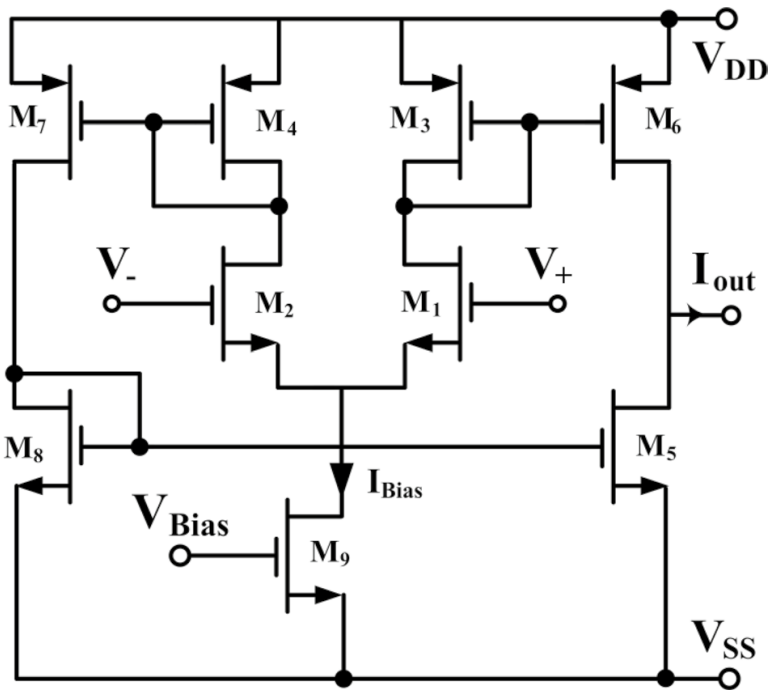


Fig. 4 CMOS implementation of OTA [13]

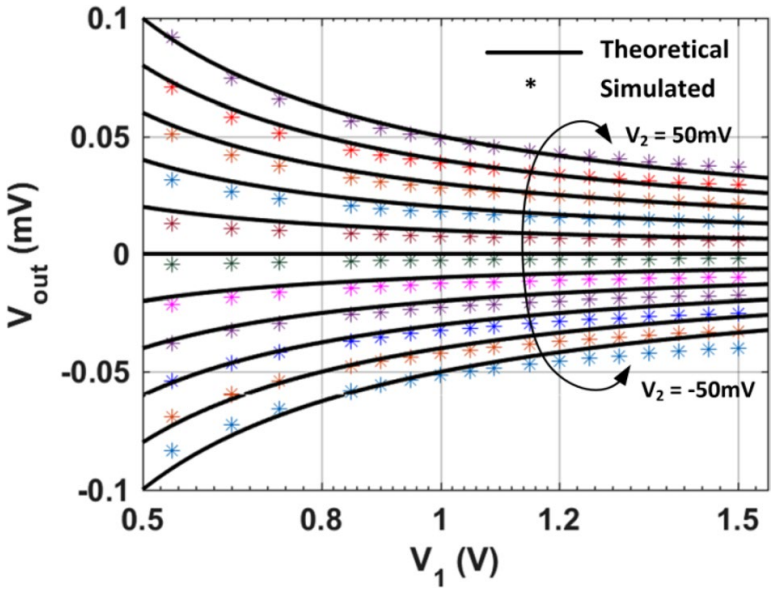


Fig. 5 DC transfer curve of V_{out} and V_1 with V_2 as a variable parameter

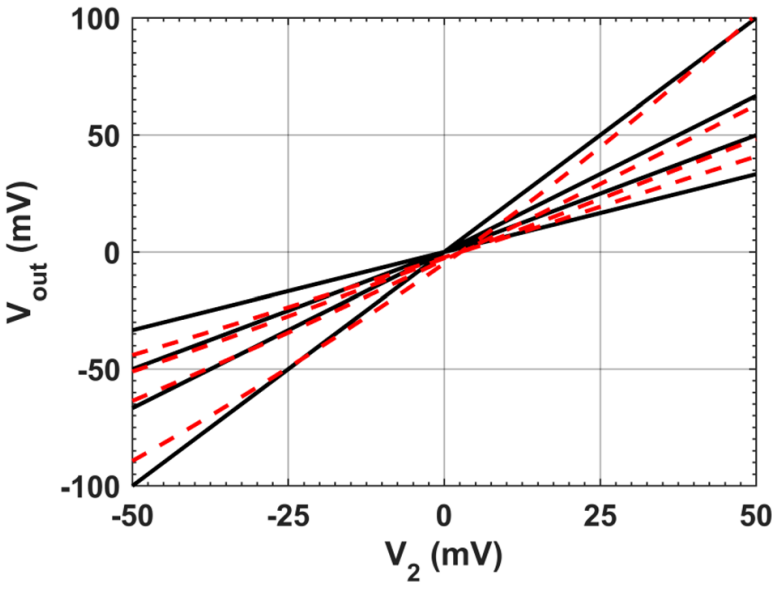


Fig. 6 DC transfer curve of V_{out} and V_2 with V_1 as a variable parameter

5.1.1 DC Characteristics

The DC transfer characteristics of the proposed divider circuit are illustrated in Fig. 5 in which the variation of output voltage with V_1 is shown when the input voltage V_2 was varied from −50 mV to +50 mV in steps of 10 mV, whereas in Fig. 6, the variation of output voltage with V_2 is shown when the input voltage V_1 is varied from 0.5 V to 1.5 V. An offset of −2.94 mV, −2.65 mV and −2.04 mV was observed in output voltage when V_2 was equal to zero and V_1 was equal to 0.5 V, 1 V and 1.5 V, respectively. We have computed the error between theoretical and simulated outputs when the input voltage V_1 was varied from 0.5 V to 1.5 V and V_2 was

varied between −50 mV and +50 mV and the same is presented in Table 1. As the input voltage V_2 is applied to the OTA directly, the maximum/minimum value of V_2 which can be applied is restricted by the linear range of the CMOS OTA (±50 mV). For the input voltage V_1, the minimum value is restricted by the saturation voltage (±0.9 V) of the OTA, while the maximum value of V_1 is restricted by the linearity error. After the repeated simulations, we have found that the divider circuit works satisfactorily when V_1 was kept in the range of 0.5 V to 1.5 V and V_2 was kept in the range of −50 mV to +50 mV.

5.1.2 Frequency and Transient Responses

The frequency response of the proposed analog divider circuit is displayed in Fig. 7, and −3 dB bandwidth is found to be 4.5 GHz. We have verified the transient performance of the divider circuit at 1 MHz, 100 MHz, 1 GHz and 4 GHz by applying a triangular input voltages, $V_2 = \pm 20\text{m}V_{\text{PP}}$ and $V_1 = \pm 100\text{m}V_{\text{PP}}$ with an offset of 150 mV. The simulated output voltage obtained by the divider circuit is shown in Fig. 8. In Fig. 9, we have shown the output voltage with 1%, 2%, 5% and 10% variation in V_P. Figure 10 displays the electronic controllability of output voltage when G_m was varied from 1 to 2mS.

5.1.3 Temperature Analysis

As the output voltage depends on G_m and K_n (both of which depend on temperature), we have evaluated the effect of variations in temperature on the output voltage when the temperature was varied between −10 and 50 °C and the simulated output voltages are shown in Fig. 11. The maximum power consumption for the divider circuit was found to be 821 µW.

5.2 Inverse Function Generation Mode

For the inverse function generator circuit, we have taken V_1 as a triangular wave with amplitude of $20\text{m}V_{\text{PP}}$ and frequency of 500 kHz, while V_2 was taken as 10 mV (DC). The transient responses of input and output voltage of the inverse function generator circuit are displayed in Fig. 12.

5.3 Square-Root Circuit

5.3.1 DC Characteristics

The DC transfer characteristic of the proposed square-root circuit when the input voltage was varied between 0 and 100 mV is illustrated in Fig. 13. The error between the theoretical and simulated output is given in Table 2.

Table 1 Theoretical and simulated output voltages with error for different values of V_1 and V_2

V_2 (mV)	$V_1 = 0.5$V			$V_1 = 1$ V			$V_1 = 1.5$ V		
	V_{out} (T) (mV)	V_{out} (S) (mV)	Error (%)	V_{out} (T) (mV)	V_{out} (S) (mV)	Error (%)	V_{out} (T) (mV)	V_{out} (S) (mV)	Error (%)
50	100	100.43	0.43	50	48.33	3.34	33.33	34.60	3.81
40	80	78.14	2.32	40	38.13	4.67	26.67	27.65	3.67
30	60	57.72	3.8	30	28.85	3.83	20	21.08	5.4
20	40	38.32	4.2	20	18.98	5.1	13.33	13.5	1.27
10	20	18.90	5.5	10	9.47	5.3	6.67	6.35	4.79
0	0	−2.94		0	−2.65		0	−2.04	
−10	−20	−21.05	5.2	−10	−10.52	5.2	−6.67	−7.08	6.14
−20	−40	−40.84	2.1	−20	−21.12	5.6	−13.33	−14.27	7.05
−30	−60	−58.78	2.03	−30	−31.26	4.2	−20	−21.86	9.3
−40	−80	−77.96	2.55	−40	−41.78	4.45	−26.67	−29.29	9.82
−50	−100	−95.25	4.75	−50	−51.02	2.04	−33.33	−36.71	10.14

T theoretical, *S* simulated

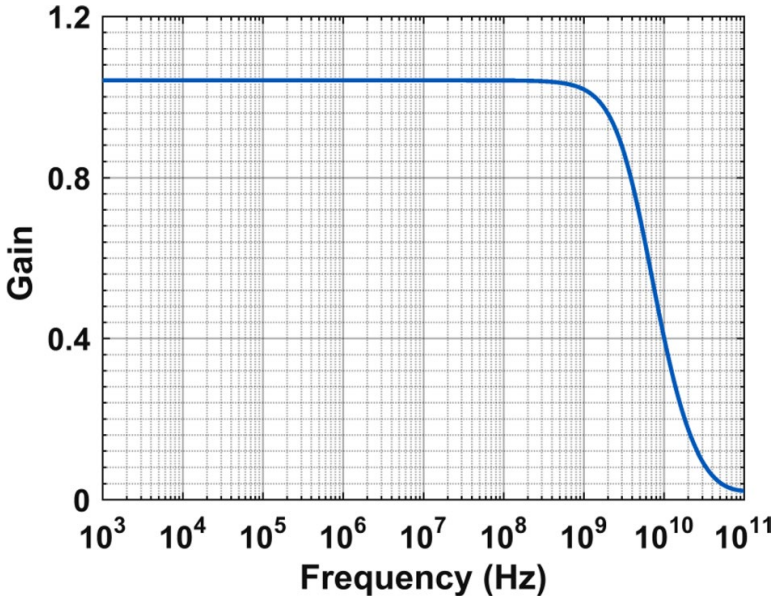


Fig. 7 Frequency response of the proposed analog divider circuit

5.3.2 Frequency and Transient Response

The frequency response of the presented square-root circuit is shown in Fig. 14, from which the bandwidth of this circuit is found to be 4.42 GHz. We have verified the transient performance of the square-root circuit at 1 MHz, 100 MHz and 1 GHz by applying a triangular input voltages $V_{\text{in}} = \pm 100$ m V_{PP} with an offset of 50 mV. The simulated output voltage obtained by the square-root circuit is shown in Fig. 15. A sinusoidal waveform of magnitude 100 m V_{PP} with an offset of 50 mV at a frequency of 1 MHz was also applied as the input, and the corresponding output voltages (V_{01}, V_{02} and V_{03}) for different values of bias current (682µA, 303µA and 75µA) are shown in Fig. 16.

5.3.3 Temperature Analysis

The output voltage of the square-root circuit depends on G_m and K_n; thus, we have evaluated the effect of variations in temperature on the output voltage when the temperature was varied between − 10 and 50 °C and the simulated responses are shown in Fig. 17. The maximum power consumption for the square-root circuit was found to 442 µW.

6 Experimental Results

To validate the simulation results, square-root circuit was bread-boarded using off-the-shelf available OTA IC LM13700 and CMOS transistor array HCF4007UBE. The experimental setup of proposed square-root circuit is shown in Fig. 18.

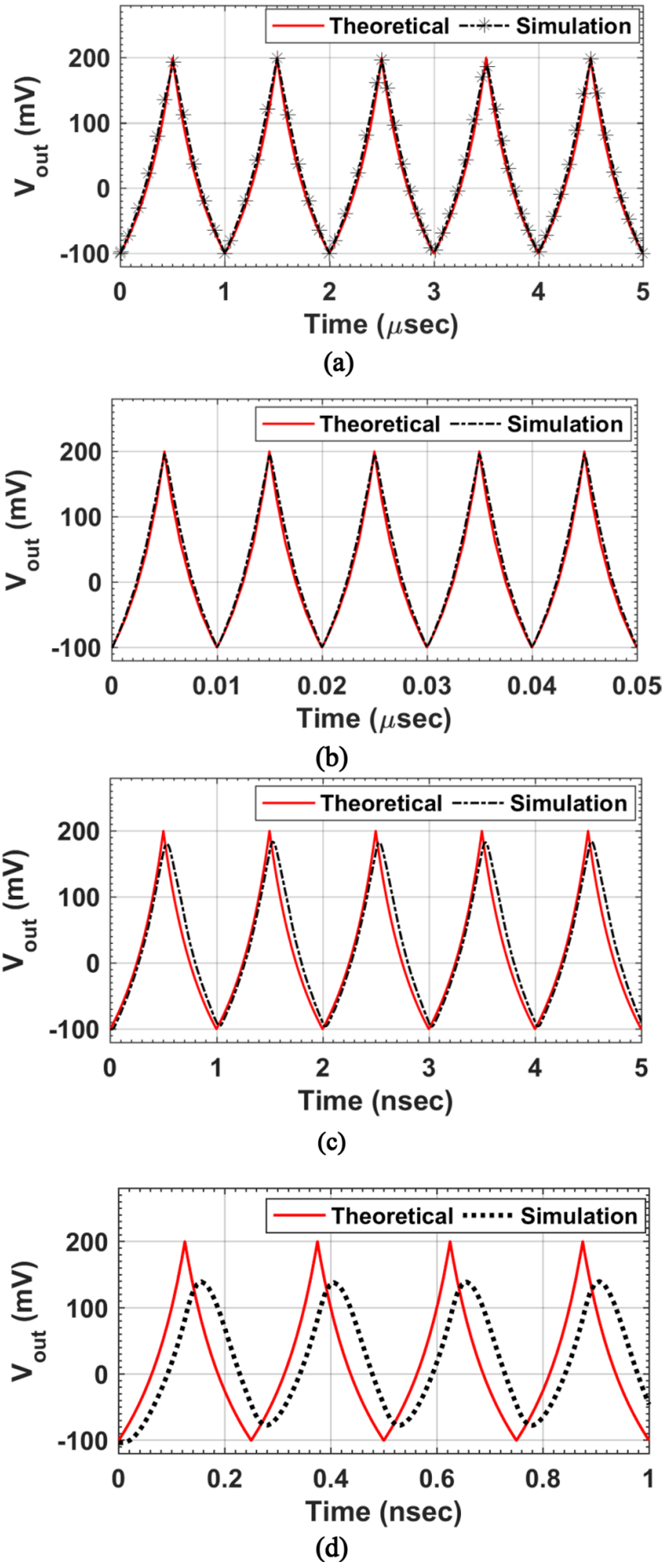


Fig. 8 Transient output responses of the proposed divider circuit at **a** 1 MHz, **b** 100 MHz, **c** 1 GHz, **d** 4 GHz

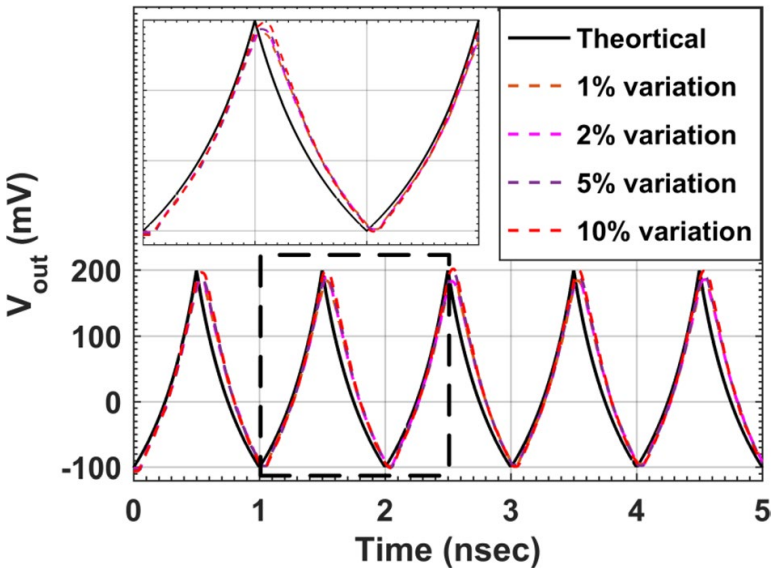


Fig. 9 Output voltage of voltage divider circuit with variation in V_P

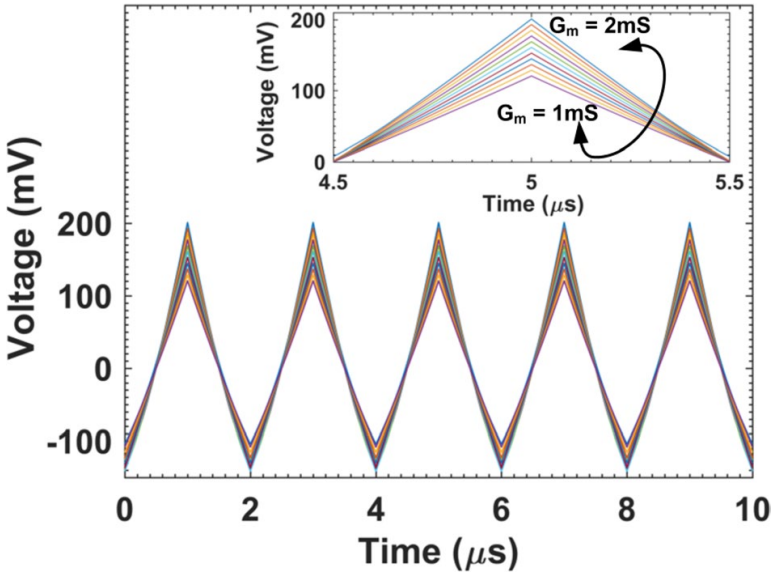


Fig. 10 Electronic tunability of output voltage of proposed divider circuit

The power supplies used to bias the OTA and the transistor array were ± 6 V. A resistor of 100 kΩ was used to set the transconductance value of OTA at 2.04 mS, while the source of NMOS transistor M_N was kept at − 1.63 V. The performance of the square-root circuit was verified by applying (1) triangular wave and (2) sinusoidal wave of 50mV_{PP} with 25 mV offset at 10 kHz, 50 kHz and 100 kHz. The corresponding output voltages are illustrated in Figs. 19, 20 and 21, respectiv ely.

The above simulation and experimental results, thus, establish the workability of the presented circuits.

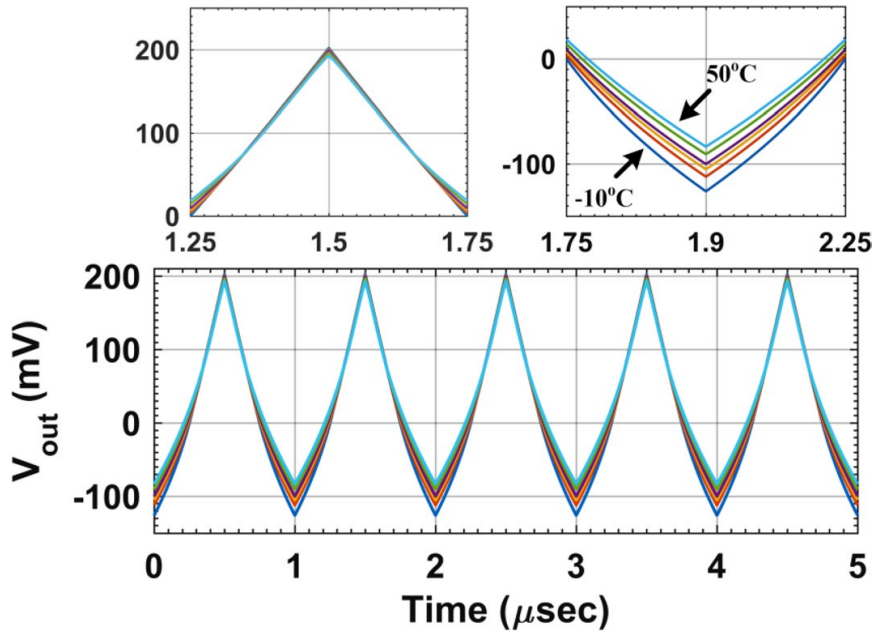


Fig. 11 Effect of temperature on output voltage of divider circuit

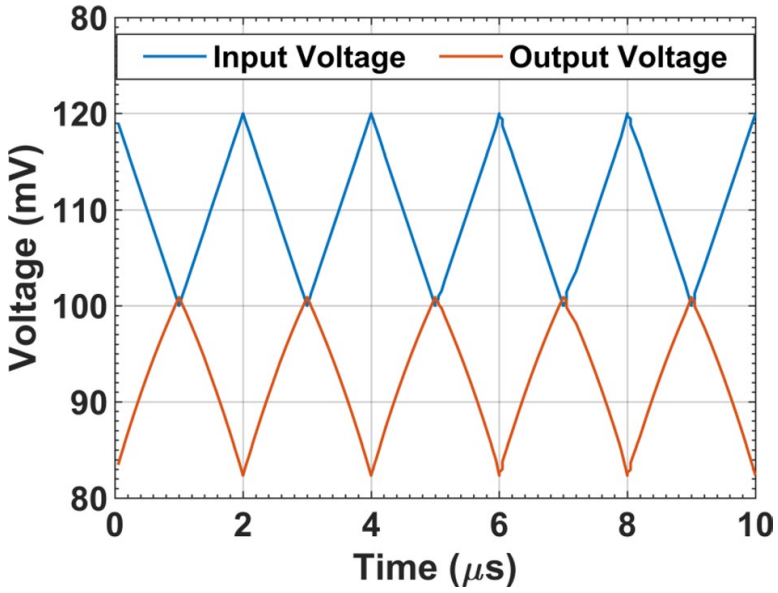


Fig. 12 Input and output transient responses of inverse function

7 Application Examples

7.1 Use of Divider Circuit to Linearize Transducer Output

The divider circuit presented in this paper may be used to control electronically the sensitivity of the linearized transducer output [12] as shown in Fig. 22.

The output voltage equation with the presented divider circuit (Fig. 1) may be expressed as:

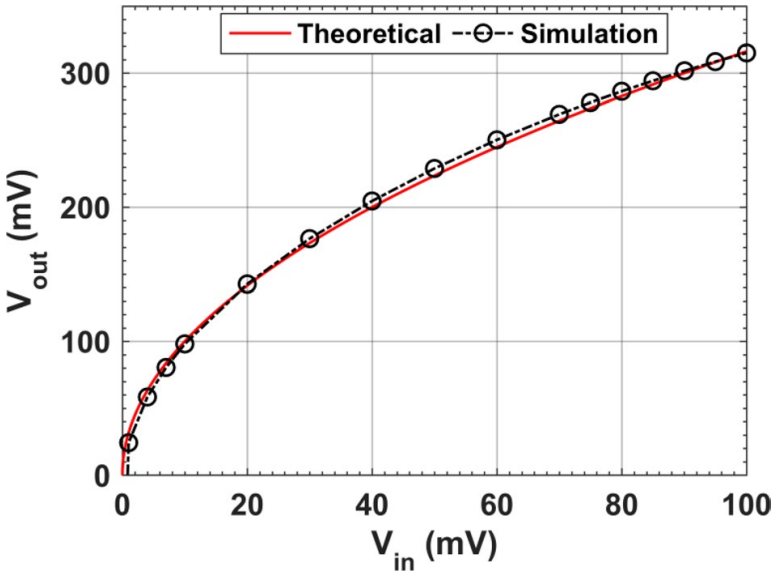


Fig. 13 DC transfer characteristics of output voltage for different bias currents

Table 2 Theoretical and simulated output voltage and % error of output voltage

Input voltage (V_{in})	Output voltage (V_{out})		% error (%)
	Theoretical	Simulated (mV)	
0	0	0.3	0.3
20 mV	141.421 mV	142.634	0.85
40 mV	200 mV	204.653	2.3
60 mV	245 mV	250	2.04
80 mV	282.843 mV	286.456	1.27
100 mV	316.228 mV	315.088	0.36

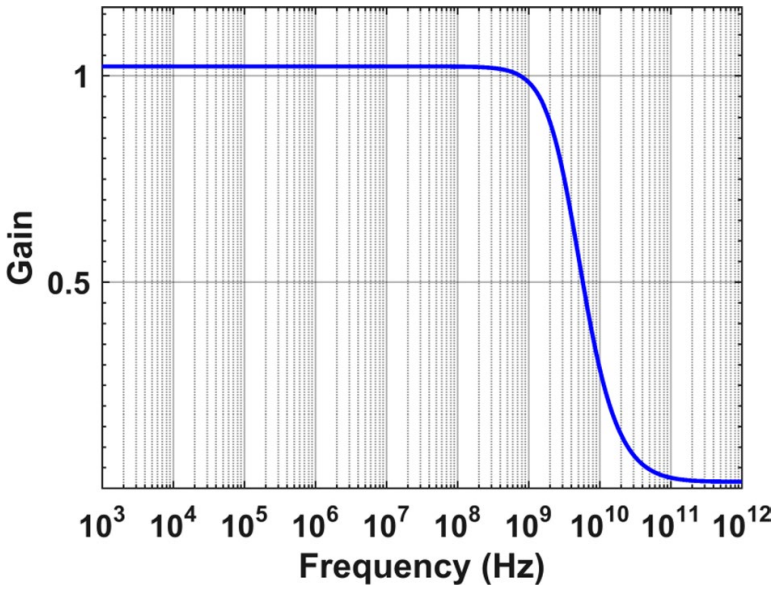


Fig. 14 Frequency response of the proposed square-root circuit

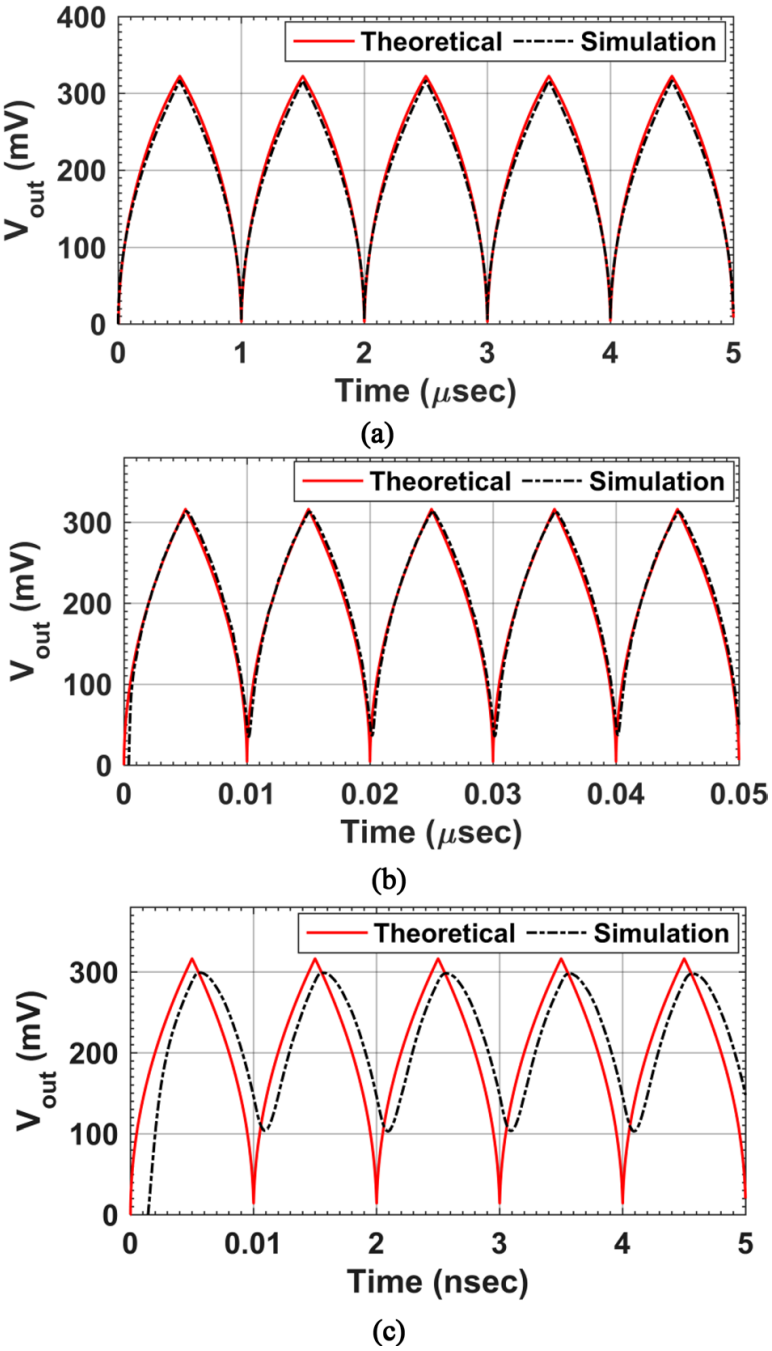


Fig. 15 Transient output responses of the proposed square-root circuit at **a** 1 MHz **b** 100 MHz **c** 1 GHz

$$V_o = \frac{VG_m}{K_n I R_s} + \frac{V}{K_n I K_3} G_m M \tag{21}$$

From Eq. (21), it may be noted that as G_m can be controlled by changing the bias voltage of the OTA (as shown in Fig. 9), the sensitivity of the transducer output may be controlled electronically.

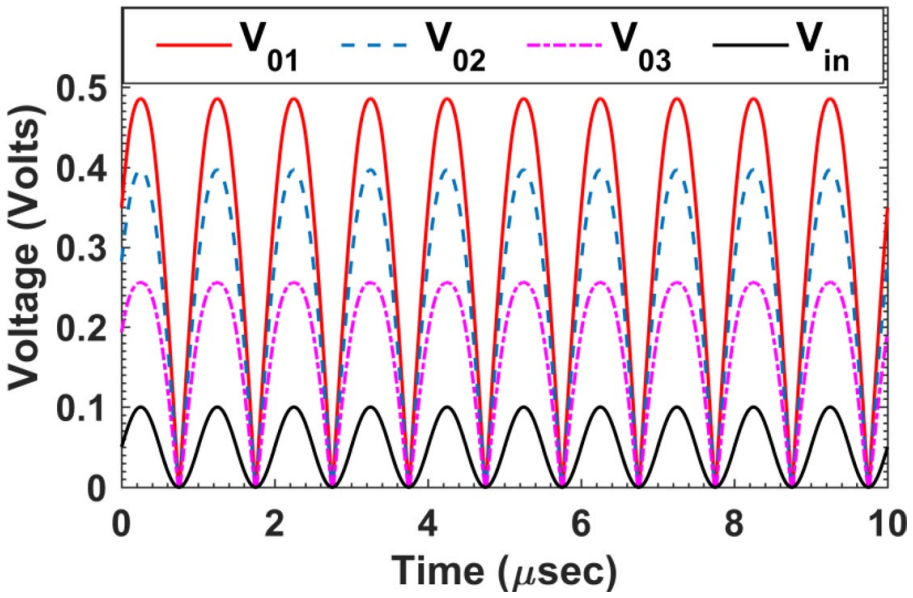


Fig. 16 Transient input and output responses for the sine wave signal of 1 MHz

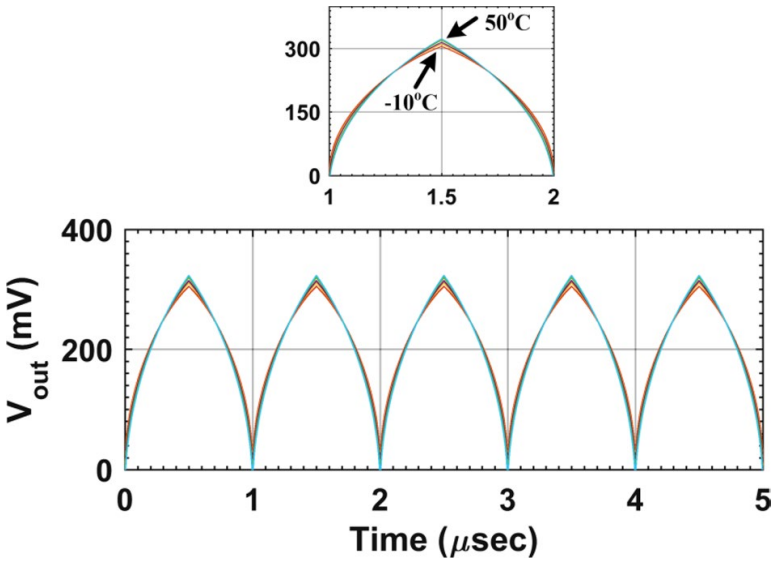


Fig. 17 Effect of temperature on output voltage of square-root circuit

7.2 Linear Voltage Controlled Resistor Using Analog Divider Circuits

The analog divider circuit presented in this paper can be used to modify the characteristics of the linearized voltage controlled resistor [42] by deriving the control voltages from the output of analog divider circuits as shown in Fig. 23. Assuming MOSFETs M_x and M_y to be matched and operating in saturation region, the input resistance R_{in} can be expressed as:

$$R_{\text{in}} = \frac{V_{\text{in}}}{I_{\text{in}}} = \frac{1}{2K_x(V_C - V_{\text{TH}})}; \text{ where } K_x = K_y = \frac{1}{2}\mu_n C_{\text{ox}}\left(\frac{W}{L}\right)_{x,y} \tag{22}$$

Since V_C is the output voltage of the divider circuit which is given by:

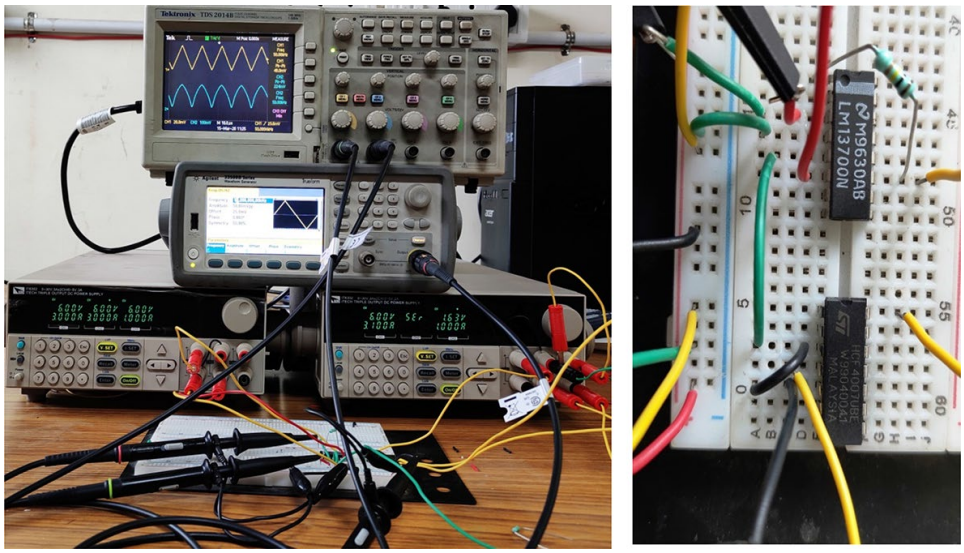


Fig. 18 Experimental setup for proposed square-root circuit in Fig. 3

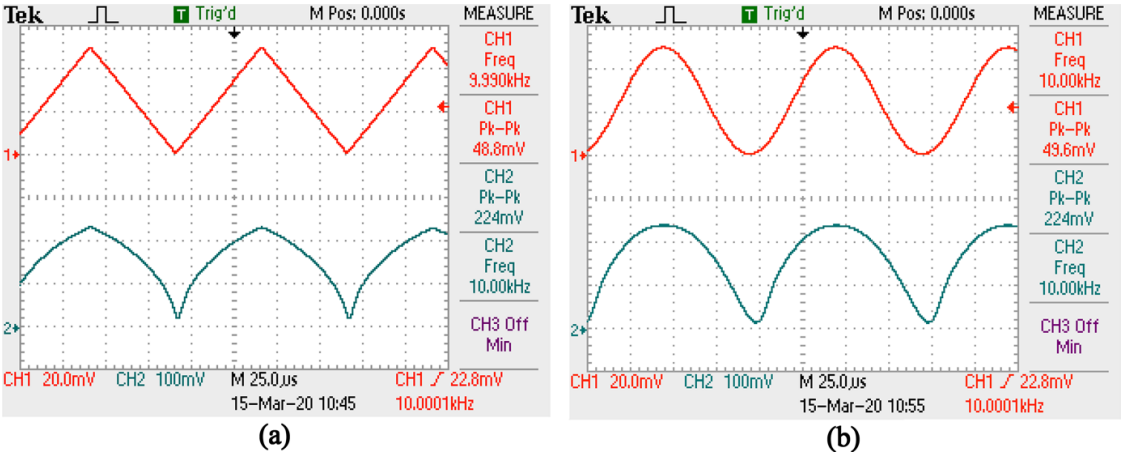


Fig. 19 Experimental input and output transient responses of the proposed square-root circuit at 10 kHz **a** triangular wave **b** sinusoidal input

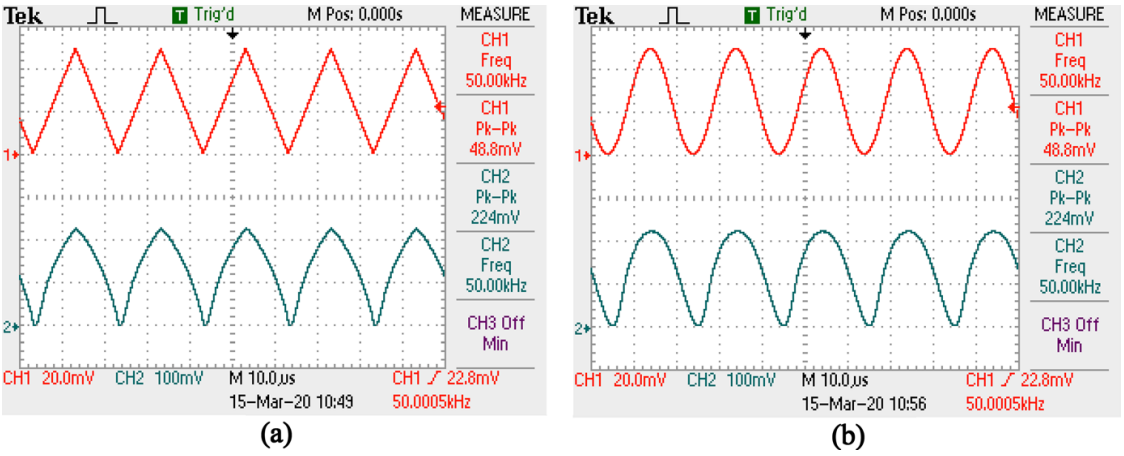


Fig. 20 Experimental input and output transient responses of the proposed square-root circuit at 50 kHz **a** triangular wave **b** sinusoidal input

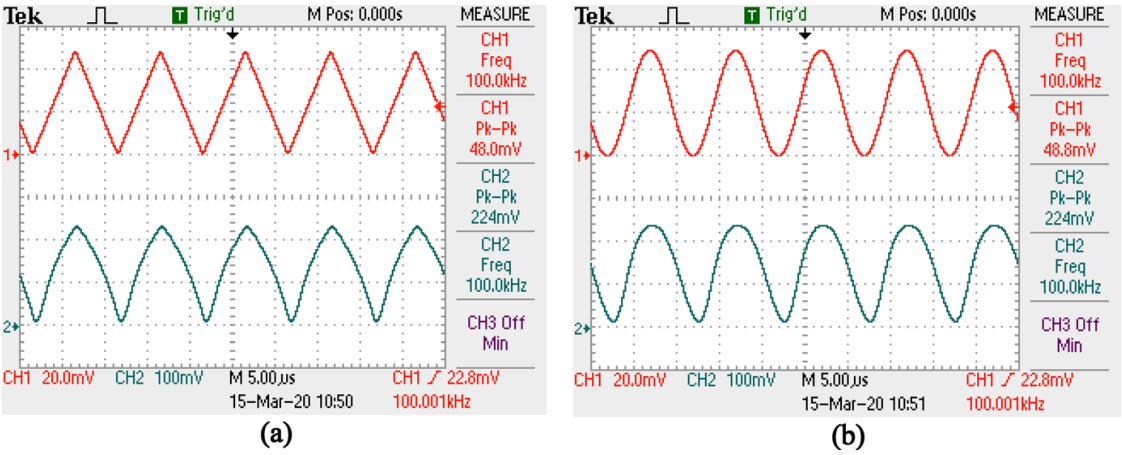


Fig. 21 Experimental input and output transient responses of the proposed square-root circuit at 100 kHz **a** triangular wave **b** sinusoidal input

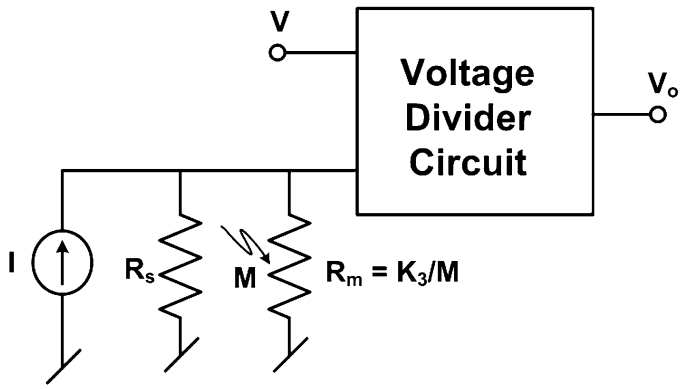


Fig. 22 Scheme to linearize transducer output [12]

$V_C = \frac{G_m}{K_n}\left(\frac{V_2}{V_1}\right)$, now if we keep V_2 constant and vary V_1, the simulated resistance will increase which is displayed in Fig. 24.

8 Comparison Between Previously Reported Circuits and Proposed Work

A comparison between the proposed voltage divider and square-root circuits with the earlier reported voltage mode divider and square-root circuits is presented in Tables 3 and 4, respectively.

When compared with the previously reported divider and square-root circuits, it may be noted that the proposed divider circuit uses a single OTA and two MOSFETs only, whereas the square-root circuit utilizes a single OTA and only one MOSFET while retaining all the functionalities of the earlier reported voltage mode divider and square-root circuits.

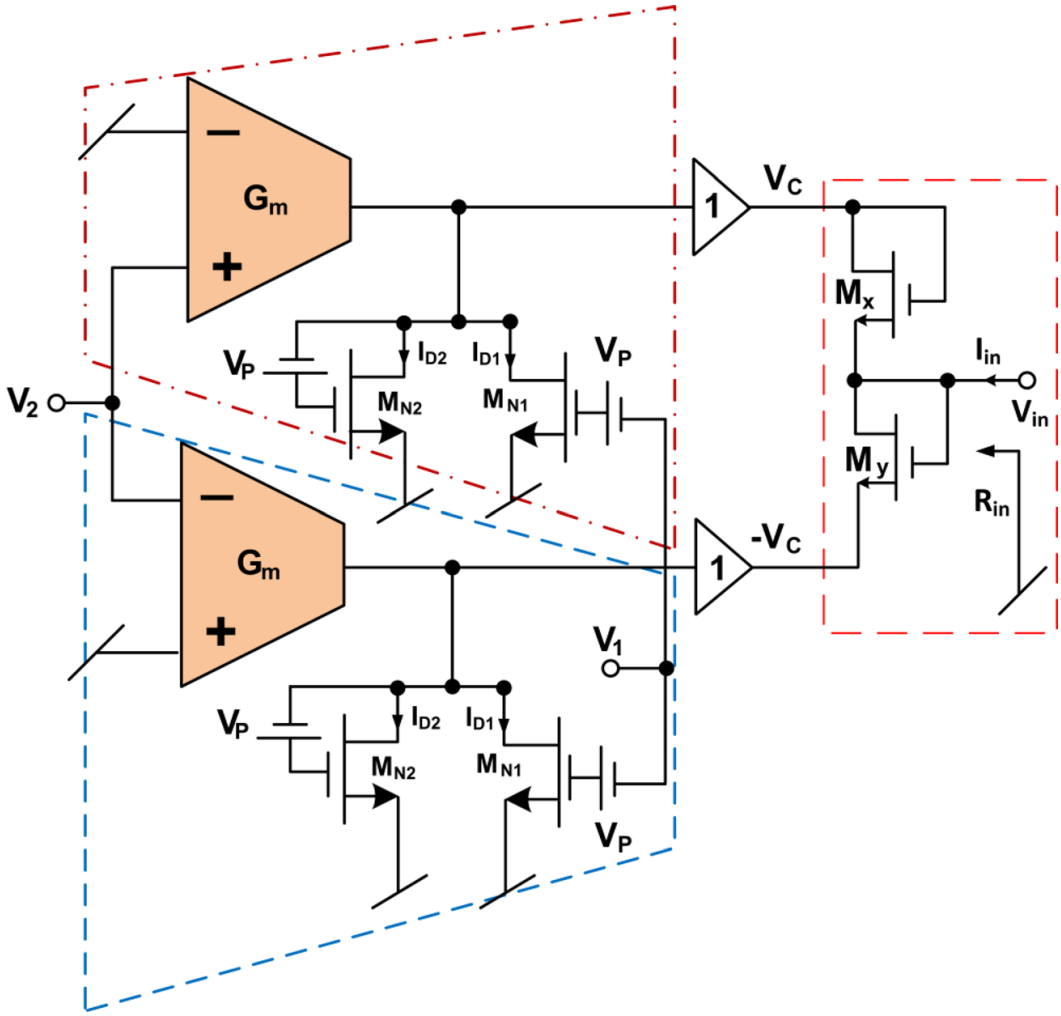


Fig. 23 Scheme to control resistor through output of the divider circuit

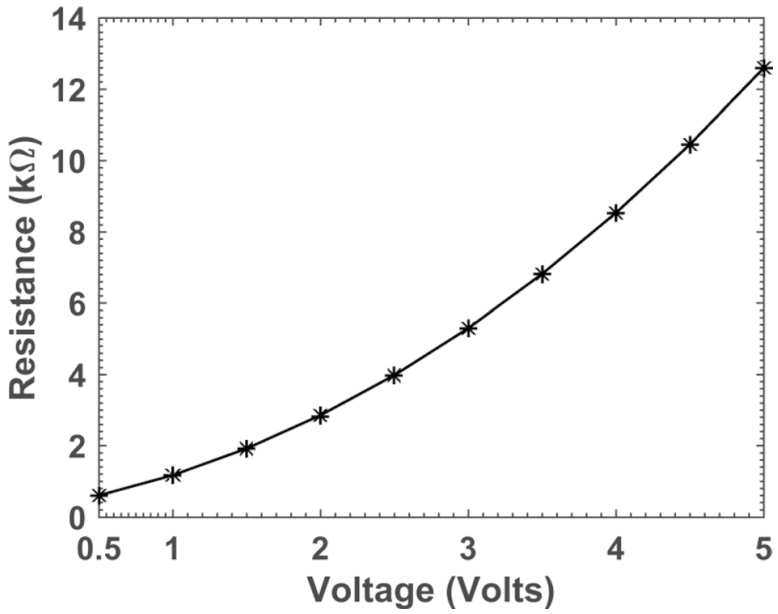


Fig. 24 Variation of R_{in} with V_1

Table 3 Performance comparison of the divider circuits with previously published works

References	ABBs	MOSFETs	*R*/*C*	Quadrant of operation	BW	PD (mW)	Output impedance	Technology	Linearity error	Input range
[6]	6 OTA	–	2R	Four	NA	3.8	High	0.13 µm	1.9%	
[7]	5 CFA	6	4C	–	NA	NA	Low	BIC	NA	NA
[12]	2 Op-amp	1	6R	Two	NA	NA	Low	BIC	NA	±100 mV
[12]	3 Op-amp	2	7R	Two	NA	NA	Low	BIC	NA	±100 mV
[14]	1CCII, 1MMCC	–	2R	Four	NA	1.08	High	0.35 µm	NA	±100 mV
[21]	2CFA	4	–	Two	2 MHz	NA	Low	BIC	NA	NA
[22]	2CCII+, 1CCII−	2	–	Two	NA	NA	High	BIC		±100 mV
[26]	1OTRA	4	–	Two	NA	1.05	Low	0.25 µm	0.84%	± 100 mV
[37]	3OTA	–	–	Two	NA	10.2	High	3 µm	NA	±1 V
Proposed	1 OTA, 1B	2	–	Two	4.5 GHz	0.821	High	0.18 µm	< 5%	$0.5 \le V_1 \le 1.5$ $-50\ \text{m} \le V_2 \le 50\ \text{m}$

NA- not available, *M* MOSFET, *R* resistor, *C* capacitor, *BW* bandwidth, *PD* power dissipation, *BIC* bipolar ICs

Table 4 Performance comparison of square-root circuits with previously published works

References	ABBs	MOSFETs	*R*	MOSFET region	Output impedance	BW	PD (mW)	Technology	Linearity error
[5]	1 OTRA	3	–	Triode	Low	500 kHz	NA	0.5 µm	NA
[8]	1DDCC, 1 *V–I* converter	2	–	triode	Low	NA	NA	CMOS	NA
[10]	2 CCCII	–	3	NA	High	60 MHz	50.4	NA	NA
[11]	1 Op-amp/*M*	2	–	Pinch-off/triode	Low	NA	NA	BIC	NA
[23]	1 CCII + 2 Op-amp	2	2	Triode	Low	400 kHz	NA	BIC	NA
[30] Fig. 3	1 CDBA,1B	2	1	Triode	High	3.46 MHz	NA	0.35 µm	<1%
[30] Fig. 4	1 CDBA	4	–	Triode	High	2.04 MHz	NA	0.35 µm	<2%
[32]	1 Op-amp, 1D	–	6	NA	Low	NA	NA	BIC	0.35%
[33]	4 OTA	–	1	NA	High	NA	NA	BIC	0.59%
[40]	3 OTA	–	2	NA	High	NA	NA	BIC	NA
Proposed	1 OTA, 1B	1	–	Saturation	High	4.42 GHz	0.442	0.18 µm	<2.4%

D diode, *B* buffer

9 Conclusions

Two structures of VM analog divider and square-root circuits employing an OTA have been presented. The presented divider circuit uses only one OTA and two MOSFETs (biased in triode region), while the square-root circuit employs one OTA along with one MOSFET only (biased in saturation region). The presented divider circuit can also be configured as an inverse voltage function generator. The bandwidth of the proposed divider and square-root circuits is relatively large (4.5 GHz and 4.42 GHz) with low power consumption (approximately 821 µW and 449 µW), respectively. A new application of divider circuit has been included; however, the problem of finding an entirely new application of the square-root circuit is open to investigation. The validity of the proposed circuits has been established through Cadence Virtuoso (Cadence EDA Tools) simulations using 0.18 µm CMOS technology parameters. Experimental results have also been included to validate the theoretical propositions.

This work has been performed at Linear Integrated Circuits Laboratory of Electrical Engineering Department, Delhi Technological University, New Delhi-110042, India.

Acknowledgements The authors gratefully acknowledge Professor Raj Senani, Department of Electronics and Communication Engineering, Netaji Subhas University of Technology, New Delhi, India, for his help and valuable suggestions in the preparation of this manuscript. The authors also wish to thank the anonymous reviewers for their constructive comments and suggestions in improving the quality of the manuscript.

Appendix

Variation in V_P Due to Changes in V_{DD}

For generation of external voltage V_P, we have used a voltage divider circuit with three matched MOSFETs (operating in saturation region) (Fig. 25).

The current flowing through each MOSFET is same and is equal to:

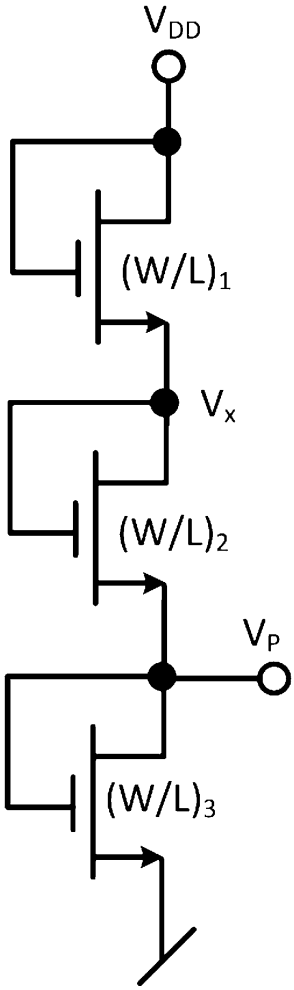
$$I_{\mathrm{D1}} = \frac{1}{2}\mu_n C_{\mathrm{ox}}\left(\frac{W}{L}\right)_1\left(V_{\mathrm{GS1}} - V_{\mathrm{TH1}}\right)^2 = \frac{1}{2}\mu_n C_{\mathrm{ox}}\left(\frac{W}{L}\right)_1\left(V_{\mathrm{DD}} - V_x - V_{\mathrm{TH1}}\right)^2 \tag{23}$$

$$I_{\mathrm{D2}} = \frac{1}{2}\mu_n C_{\mathrm{ox}}\left(\frac{W}{L}\right)_2\left(V_{\mathrm{GS2}} - V_{\mathrm{TH2}}\right)^2 = \frac{1}{2}\mu_n C_{\mathrm{ox}}\left(\frac{W}{L}\right)_2\left(V_x - V_P - V_{\mathrm{TH2}}\right)^2 \tag{24}$$

$$I_{\mathrm{D3}} = \frac{1}{2}\mu_n C_{\mathrm{ox}}\left(\frac{W}{L}\right)_3\left(V_{\mathrm{GS3}} - V_{\mathrm{TH3}}\right)^2 = \frac{1}{2}\mu_n C_{\mathrm{ox}}\left(\frac{W}{L}\right)_3\left(V_P - V_{\mathrm{TH3}}\right)^2 \tag{25}$$

Solving Eqs. (23–25), assuming equal threshold voltages, we get:

Fig. 25 Voltage divider circuit



$$V_P = \frac{\sqrt{\frac{\left(W/L\right)_1}{\left(W/L\right)_2}}\left(V_{\mathrm{DD}} - V_{\mathrm{TH}}\right)}{1 + \sqrt{\frac{\left(W/L\right)_1}{\left(W/L\right)_2}} + \sqrt{\frac{\left(W/L\right)_1}{\left(W/L\right)_3}}} \cong \sqrt{\frac{\left(W/L\right)_1}{\left(W/L\right)_2}}\left(V_{\mathrm{DD}} - V_{\mathrm{TH}}\right)\left(1 - \sqrt{\frac{\left(W/L\right)_1}{\left(W/L\right)_2}} - \sqrt{\frac{\left(W/L\right)_1}{\left(W/L\right)_3}}\right) \tag{26}$$

Now, in Eq. (26), if a deviation is introduced in V_{DD} by a value of $\pm\Delta V_{\mathrm{DD}}$, then Eq. (26) becomes:

$$V_P \cong V_P + V_P' \tag{26}$$

$$\text{where } V_P' = \pm\frac{\Delta V_{\mathrm{DD}}}{\left(1 + \sqrt{\frac{\left(W/L\right)_2}{\left(W/L\right)_1}} + \sqrt{\frac{\left(W/L\right)_2}{\left(W/L\right)_3}}\right)} \tag{27}$$

Therefore, from Eq. (27), the effect of changes in V_{DD} may be made negligible small, if $(W/L)_2 \gg (W/L)_1$ and $(W/L)_2 \gg (W/L)_3$.

References

1. P.E. Allen, D.R. Holberg, *CMOS Analog Circuit Design* (Oxford University Press, New York, 1987)
2. M.T. Abuelma'atti, M.L. Al-Qahtani, A current-mode current-controlled current conveyor-based analogue multiplier/divider. Int. J. Electron. **85**(1), 71–77 (1998)
3. M.A. Al-Absi, A. Hussein, M.T. Abulma'atti, A low voltage and low power current-mode analog computational circuit. Circuits, Syst., Sig. Process. **32**(1), 321–331 (2013)
4. E.S. Al-Suhaibani, M.A. Al-Absi, A new CMOS current-mode controllable-gain square rooting circuit using MOSFET in subthreshold. Analog Integr. Circuits Sig. Process **82**(2), 431–434 (2015)
5. G. Aggarwal, H. Garg, N. Bansal, P. Gangwar, R. Pandey, Single OTRA based Low voltage square root circuit. Int. J. Adv. Res. Innov. **5**(4), 431–432 (2017)
6. M. Bhanja, B.N. Ray, OTA-based logarithmic circuit for arbitrary input signal and its application. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **24**(2), 638–649 (2016)
7. N. Bajaj, J. Govil, Realisation of a differential multiplier-divider based on current feedback amplifiers, in *International Conference on Semiconductor Electronics* (IEEE, 2006), pp. 708–712
8. W. Chiu, S.-I. Liu, H.-W. Tsao, J.-J. Chen, CMOS differential difference current conveyors and their applications. IEE Proc. Circuits Dev. Syst. **143**(2), 91–96 (1996)
9. E.O. Doebelin, *Measurement Systems: Application and Design* (McGraw Hill, New York, 2004)
10. K. Dejhan, C. Netbut, New simple square-rooting circuits based on translinear current conveyors. Int. J. Electron. **94**(7), 707–723 (2007)
11. I.M. Filanovsky, H.P. Baltes, Simple CMOS analog square-rooting and squaring circuits. IEEE Trans. Circuits Syst. **39**(4), 312–315 (1992)
12. D. Ghosh, D. Patranabis, A simple analog divider having independent control of sensitivity and design conditions. IEEE Trans. Instrum. Meas. **39**(3), 522–526 (1990)
13. E.W. Greeneich, *Analog Integrated Circuits* (International Thomson Publishing, Stamford, 1997)
14. Y.S. Hwang, W.H. Liu, S.H. Tu, J.J. Chen, New building block: multiplication-mode current conveyor. IET Circuits Dev. Syst. **3**(1), 41–48 (2009)
15. W. Jaikla, M. Siripruchyanun, A novel current-mode multiplier/divider employing only single dual-output current controlled CDTA, in *2007 International Symposium on Communications and Information Technologies* (IEEE, 2007), pp. 106–109
16. K. Kaewdang, C. Fongsamut, W. Surakampontorn, A wide band current-mode OTA based analog multiplier-divider, in *Proceedings of the 2003 International Symposium on Circuits and Systems* (2003), pp. 349–352
17. M. Kumngern, U. Torteanchai, A CMOS current-mode multiplier/divider using a current amplifier, in *7th International Power Engineering and Optimization Conference (PEOCO)* (IEEE, 2013), pp. 742–745
18. M. Kumngern, U. Torteanchai, OTA-based current square rooting circuit. Far East J. Electron. Commun. **17**(1), 127–140 (2017). https://doi.org/10.17654/ec017010127
19. A.J. Lopez-Martin, A. Carlos, J. Ramirez-Angulo, R.G. Carvajal, Compact low-voltage CMOS current-mode multiplier/divider, in *International Symposium on Circuits and Systems* (IEEE, 2010), pp. 1583–1586
20. W. Liu, S.-I. Liu, S.K. Wei, CMOS current-mode divider and its applications. IEEE Trans. Circuits Syst.-II: Express Briefs **52**(3), 145–148 (2005)
21. S.-I. Liu, J.J. Chen, Realisation of analogue divider using current feedback amplifiers. IEE Proc. Circuits, Dev. Syst. **142**(1), 45–48 (1995)
22. S.-I. Liu, D.S. Wu, H.W. Tsao, J. Wu, J.H. Tsay, Nonlinear circuit applications with current conveyors. IEE Proc. G-Circuits, Dev. Syst. **140**(1), 1–6 (1993)
23. S.-I. Liu, Square-rooting and vector summation circuits using current conveyors. IEE Proc. Circuit, Dev. Syst. **142**(4), 223–226 (1995)

24. P. Mongkolwai, W. Tangsrirat, CFTA-based current multiplier/divider circuit, in *2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)* (IEEE, 2011), pp. 1–4
25. W. Narksarp, P. Pawarangkoon, W. Kiranon, P. Wadkien, A four-quadrant current-mode multiplier/divider building block, in *2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (IEEE, 2009), pp. 574–577
26. B.C. Nagar, S.K. Paul, Single OTRA based two quadrant analog voltage divider. Analog Integr. Circuits Sig. Process **94**(1), 161–169 (2018)
27. S. Parsnejad, Y. Hu, H. Wan, E. Ashoori, A.J. Mason, Wide dynamic range multi-channel electrochemical instrument for in-field measurements, in *IEEE Sensors* (2016), pp. 1–3
28. W. Petchakit, W. Kiranon, P. Wardkien, S. Petchakit, A current-mode CCCII-based analog multiplier/divider, in *ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (IEEE, 2010), pp. 221–224
29. N. Pisutthipong, M. Siripruchyanun, A novel simple current-mode multiplier/divider employing only single multiple-output current controlled CTTA, in *TENCON 2009 IEEE Region 10 Conference* (IEEE, 2009), pp. 1–4
30. J.K. Pathak, A.K. Singh, R. Senani, New squaring and square-rooting circuits using CDBA. Am. J. Electr. Electron. Eng. **2**(6), 175–179 (2014)
31. K.D. Peterson, A. Nedungadi, R.L. Geiger, Amplifier design considerations for high frequency monolithic filters, in *Proceedings of European Conference on Circuit Theory and Design (ECCTD)* (1987), pp. 321–326
32. V. Riewruja, T. Kamsri, Square-rooting and absolute function circuits using operational amplifiers. IET Circuits Dev. Syst. **3**(2), 57–63 (2009)
33. V. Riewruja, Simple square-rooting circuit using OTAs. Electron. Lett. **44**(17), 1000–1002 (2008)
34. W. Surakampontorn, K. Khanittha, F. Chalermpan, A simple current-mode analog multiplier-divider circuit using OTAs, in *Proceedings of the IEEE Conference* (2002), pp. 658–661
35. M. Siripruchyanun, A design of analog multiplier and divider using current controlled current differencing buffered amplifiers, in *Proceedings of the International Symposium on Integrated Circuits (ISIC'07)* (2007), pp. 568–571
36. M. Siripruchyanun, W. Jakila, A current-mode multiplier/divider based on CCCDTA. Int. J. Electron. Commun. (AEU) **62**(3), 223–227 (2008)
37. E.S. Sinencio, J.R. Angulo, B.L. Barranco, A.R. Vazquez, Operational transconductance amplifier-based nonlinear function syntheses. IEEE J. Solid-State Circuits **24**(6), 1576–1586 (1989)
38. N. Tadic, D. Gobovic, Smart sensor interfacing circuit using square-rooting current-to-frequency conversion. Int. J. Electron. **94**(12), 1075–1098 (2007)
39. W. Tangsrirat, T. Pukkalanun, P. Mongkolwai, W. Surakampontorn, Simple current-mode analog multiplier, divider, square-rooter and squarer based on CDTAs. Int. J. Electron. Commun. (AEU) **65**(3), 198–203 (2011)
40. K.S. Tarun, R. Arabinda, N.R. Baidyanath, An arbitrary power law device based on operational transconductance amplifiers. IEEE Trans. Instrum. Meas. **42**(5), 948–952 (1993)
41. H. Wasaki, Y. Horio, S. Nakamura, Current multiplier/divider circuit. Electron. Lett. **27**(6), 504–506 (1991)
42. Z. Wang, A CMOS four-quadrant analog multiplier with single-ended voltage output and improved temperature performance. IEEE J. Solid-State Circuits **26**(9), 1293–1301 (1991)

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

International Journal of Peptide Research and Therapeutics
https://doi.org/10.1007/s10989-020-10073-z

Venom-Derived Bioactive Compounds as Potential Anticancer Agents: A Review

Arpita Roy[1] · **Navneeta Bharadvaja**[1]

Accepted: 8 May 2020
© Springer Nature B.V. 2020

Abstract
Biologically active compounds with potential anticancer activity have been isolated from various animal sources. Recent research is focused on venoms from animal sources, since they specifically target cancer cells and are less toxic to normal tissues, thus making them a better alternative for prevention and management of cancer. Animals produce pharmacologically active compounds that are capable of protein synthesis inhibition, induction of angiogenesis and apoptosis. Over the past decades, venom has been identified as a potential source of therapeutic agents, including anti-cancer agents. Venoms are isolated from various animal sources such as wasp, honey bee and scorpion are composite mixtures of various peptides, protein, enzymes, toxins and non-proteinaceous secretions. They are natural compounds that act as a template and can be utilized for the development of new therapeutic agents against cancer. The objective of this review is to create a comprehensive resource of anticancer agents from venoms which would foster pharmacogenomic studies and the development of cost-effective cancer treatment. Information about different animal venoms has been compiled after a panoramic literature survey.

Keywords Cancer · Natural agents · Venoms · Clinical trials

Introduction

Cancer is one of the leading causes of death in both developing and developed countries. Cell division is a physiological process that occurs in tissues. Balance between proliferation and programmed cell death is maintained under normal circumstances, usually in the form of apoptosis (Siegel et al. 2019). According to Cancer Key Facts (2018) National Institute of Health estimated 1,735,350 new cases of cancer and 609,640 cancer linked deaths in the United States. According to World Health Organization fact sheet (September 2018), cancer is the second major reason for death worldwide and approximately one in six deaths occurs due to cancer. Intake of fewer vegetables and fruits, reduced physical activities, consumption of tobacco and alcohol are some of the major causes of cancer development. Among these tobaccos is the most fatal, which accounts for 22% of global deaths and 70% of total lung cancer deaths (WHO factsheet). According to National Cancer Institute, the most widely observed cancers include lung, prostate, rectum, breast, colon, bronchus, melanoma, kidney, endometrial, leukemia, pancreatic, thyroid and liver in United States. New cancer cases are expected to increase to 23.6 million per year by 2030 in United States (Siegel et al. 2018). In 2016, 3.9 million cancer cases were estimated in India (Indian Council of Medical Research). In the current scenario, approximately 2.5 million cancer cases are estimated to be presented in India and every year over 7 lakh new cancer patients registered (https://cancerindia.org.in/statistics/).

Various treatments have been used to treat cancer, including chemotherapy, radiation therapy, immunotherapy and gene therapy (Kumar et al. 2017a, b; Roy et al. 2017). Out of which chemotherapy remains the principal choice but it has serious side effects. Enormous investigations have been made globally during the past few decades to discover new drugs against cancer. Traditional approaches for discovery of anticancer drug is time-consuming and expensive (Tomar et al. 2019). Nature contains a vast diversity of bioactive molecules which have therapeutic activity. Utilization of natural anticancer agents has great potential to overcome cancer and yield promising results (Roy et al. 2018). Venomous animals play a vital role in the discovery of novel

✉ Navneeta Bharadvaja
navneetab@dce.ac.in

[1] Plant Biotechnology Laboratory, Department of Biotechnology, Delhi Technological University, Delhi, India

Published online: 16 May 2020

therapeutic candidates. They produce various venoms to prevent themselves from external surroundings (Abidin et al. 2019). Venom inhibits proliferation of cancer cells and promotes cell death by inducing apoptosis, cytochrome C release, decreasing or increasing proteins expression which controls cell cycle (Li et al. 2018a, b). Venom diversity is an exceptional way to develop new therapeutic agents as they offer a formidable array of biological properties. Venom from various organisms such as scorpion, beetles, honeybee, etc. are in the experimental stage and have shown biotechnological or pharmacological applications. Some of the examples of bioactive compounds produced by animals which have potential to treat cancer include: chlorotoxin from scorpion which inhibits the cancer cell invasion, raventoxin from spider which inhibit HeLa cells, BJcuL from snake which inhibit tumor cell development and growth, etc. (Carvalho et al. 2001; Gao et al. 2005; Ojeda et al. 2016). A compound TM-601 which is a modified form of chlorotoxin isolated from scorpion venom able to bind specifically to glioma cell surfaces and blockes chloride channel is currently in phase II of human trials (Cohen et al. 2018). These examples show that animal toxins have extended the field of drug development. This review gives a summary of different scientific studies that reinforce the use of animal toxins/venoms as an anticancer agent. Further venom-based drugs that are under clinical trials have also been discussed.

Animal Derived Venoms as Anticancer Agents

Natural products are a potential source of therapeutic agents because they display huge chemical diversity (Roy and Bharadvaja 2017; Sachdeva et al. 2020). Identification of potential anticancer compounds from animal source grow continuously (Wang et al. 2017; Ejaz et al. 2018). Animals contain various toxins/venoms which they use for their protection from environmental conditions. Venoms are secreted by venomous animals and have various pharmaceutical applications, including treatment of cardiovascular diseases, autoimmune diseases, neurodegenerative diseases, cancer, etc. (Klupczynska et al. 2018). Many venom proteins and peptides have been reported as potential anticancer agents. Venoms from different animals hold a promising result to cure various types of malignancies, therefore they can be utilized as a potential anticancer agent (Aaghaz et al. 2019). Anticancer properties of venoms obtained from various animal sources such as snake, bee, scorpion, beetle, wasp, ant, caterpillar, and spider have been discussed in detail (Fig. 1). Venoms from various animal sources and its mode of action are summarised in Table 1.

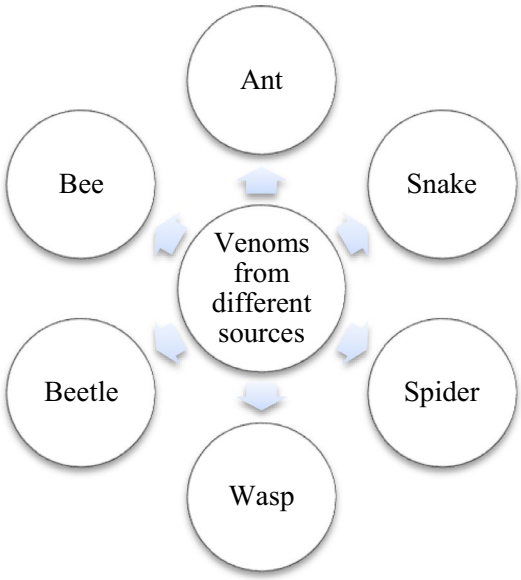


Fig. 1 Venoms from different sources

Ant Venom

Ant venoms have been found to contain an extraordinary diversity of toxins which have great therapeutic value. Investigators found that solenopsin A (from *Solenopsis invicta*) possess anti-angiogenic activity. Arbiser et al. (2007) reported that synthetic solenopsin A possesses strong inhibitory effect against class-1 phosphatidylinositol-3-kinase signaling and angiogenesis in zebra fish and mice embryos, which makes this compound a potential lead for cancer treatment. Al-Tamimi et al. (2018) reported that some ant venom possessess antineoplastic activity in different cell lines (HepG2, MCF-7, and LoVo). The cell lines showed different dose-dependent antineoplastic effects with an improved level of cytokines, which includes interleukin (IL)-1β, IL-6, IL-8, transcription factor and NF-κB. Therefore, this venom has an exceptional antineoplastic activity that can be directly utilized for the treatment of certain types of cancer.

Bee Venom

Bee venom comprises of a wide range of proteins, peptides, bioactive amines, and low molecular weight compounds with therapeutic importance such as apamin, melittin, adolapin, mast-cell-degranulating peptide, amine, enzymes, and non-peptide components (Son et al. 2007). Venom from bee has been used for treatments of various diseases like rheumatism, arthritis, pain, tumors, and skin diseases (Rady et al. 2017). Various studies have reported the anticancer activity of bee venom, which includes apoptosis induction, cytotoxicity, necrosis, and inhibition of cancer cell proliferation

Table 1 Anticancer agents from animal source

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Cathelicidin-BF	*Bungarus fasciatus*	Banded krait	Snake	Indian subcontinent and Southeast Asia	Venom	Peptide	–	Inhibits B16F10 and B16 proliferation	Marcinkiewicz et al. (2000)
BJcuL (lectin)	*Bothrops jararacussu*	Golden lancehead	Snake	Eastern Brazil	Venom	Protein	10 µg	Prevents tumor cell development and growth of endothelical cell and induces agglutination of erythrocytes	Carvalho et al. (2001)
Melittin	*Apis milefera*	European honeybee	Honeybee	Europe and South East Asia	Venom	Peptide	5–10 µg/mL	Induces apoptosis and possess anti-tumor propertie	Jang et al. (2003)
Kahalalide F	*Elysia rufescens*	–	Marine mollusc	Pacific Ocean	Skin	Depsipeptide	0.07 mM to 0.28 mM	Affects multiplication of cancer cells in a dose dependent manner by inhibition of DNA synthesis	Suárez et al. (2003)
Obtustatin	*Vipera lebetina obtuse*	West Asian blunt-nosed viper	Sanke	Asisa	Venom	Peptide	–	It possesses antiangiogenesis in cancer cells	Marcinkiewicz et al. (2003a, b)
Magi-5	*Macrothele gigas*	Japanease funnel web spider	Spider	Okinawa	Venom	Peptide	267 nM	Blocking the ion channels principle	Corzo et al. (2003)
Oxyopinins	*Oxyopes kitabensis*	wolf spider	Spider	Central Asia	Venom	Peptide	–	It forms pores in lipid membrane and possess anti-tumor action of other pore-forming peptides	Belokoneva et al. (2003)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Norcantharidin	*Mylabris phalerata pallas*	Chinese blister beetle	Beetle	China	Venom	Fatty substance of terpenoid	60–480 µmol/l	Regulates activity of caspase-3,-8 and -9 which are essential pathways for the induction of apoptosis	An et al. (2004)
Raventoxin	*Macrothele raven*	Common raven	Spider	Australia Asia and Europe	Venom	Peptide	20–30 mg/l	Inhibit HeLa cells by induction of apoptosis, necrosis of toxicity damage and direct lysis	Gao et al. (2005)
Lebestatin	*Macrovipera lebetina*	Blunt-nosed viper	Snake	North Africa	Venom	Small protein	0.20 nM	Interacts precisely with the α1β1 integrin that is present on leminin1 and collagen. Interaction amongst integrin and ECM is essential for cancer adhesion and migration of the cancer cells	Olfa et al. (2005)
Rhodostomin	*Agkistrodon rhodostoma*	Malayan pit viper	Snake	South East Asia	Venom	Protein	10 µg/mL	Hinders the interaction between intergrins and ECM proteins which is necessary for proliferation of the cancer cell	Yang et al. (2005)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Chlorotoxin	*Leiurus quinquestriatus*	Death stalkwer-scorpion	Scorpian	Desearts of North Africa	Venom	Peptide	3.7 gbq	Inhibits the cancer cell invasion by disabling the MMP-2 and induced amplified internalization of lipid rafts	Shen et al. (2005)
Batroxobin	*Bothrops atrox*	Common lancehead	Snake	Northern South America	Venom	Protein	–	Prevents cancer cell propagation and metastasis	Sahni et al. (2007)
Snake venom toxin	*Vipera lebtina turanica*	Turan blunt-nosed viper	Snake	West Asia	Venom	–	1.7 and 1.8 µg/mL	Constrains cell growth of cancer cells through arrestin in the G_2-M phase. Prevents NF-κβ, an anti-apoptotic transcriptional factor and a substantial inhibition of p50 nucleus translocation	Son et al. (2007)
Saxatilin	–	–	Snake	–	Venom	Polypeptide (73 AA)	–	Decreased cell invasion through MMP-9 regulation activity in MDAH 2774	Kim et al. (2007)
Lebectin	*Macrovipera lebetina*	Blunt-nosed viper	Snake	North Africa	Venom	Protein	–	Hinders adhesion, migration, and tumor cells invasion and angiogenesis inhibition	Sarray et al. (2007)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Eristostatin	*Eristicophis macmahoni*	Asian sand viper	Snake	Balochistan Iran, Pakistan, and Afghanistan	Venom	Small protein	–	Inhibition of melanoma cells colonization in the lung and liver	Tian et al. (2007)
Ammodytoxin- C	*Vipera ammodytes*	Horned viper	Snake	Southern Europe	Venom	Protein	–	Activates autophagy and rise formation of autophagosomes as well as apoptosis	Premzl et al. (2008)
Crotatroxin 2	*Crotalus atrox*	Western diamondback rattlesnake	Snake	America and Mexico	Venom	Protein	17.5 nM	Inhibits fibrinogen binding to GP IIb/IIIa.), migration of cells and tumor colonization	Galán et al. (2008)
A-Bungarotoxin	*Bungarus multicinctus*	Twaiwanen krait	Snake	China and Taiwan	Venom	Protein	10 µm	Ability to block cancer cell proliferation as it causes G_1 arrest and abrogated nicotine NK-stimulated cell proliferation	Shin et al. (2008)
Hyaluronidase	*Mesobuthus martensi*	Chinese scorpionn	Scorpian	China, Mongolia, Japan and Korea	Enzyme	Enzyme	–	Removes hyaluronan and modulates CD44 variant expression in breast cancer cells	Feng et al. (2008)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Viperistatin	*Vipera palEastinae*	Palestine viper	Snake	Western Syria Northern and Central America	Venom	Protein	5 mg/kg	Interaction between intregrins and extra cellular matrix is highly essential for cancer cell migration and invasion. It prevents both adhesion and relocation of cancer cells	Staniszewska et al. (2009)
Resibufogenin	*Bufo melanosticus schneider*	Spectacled toad	Toad	East Asia	Skin	Protein	17.7 to 19.0 ng/mL	Inhibition of tumor cell proliferation and stimulates apoptosis by alterating expression of c-Myc and Bcl-2	Meng et al. (2009)
Bufalin	*Bufo bufo gargarizans*	Asiatic toad	Toad	East Asia	Skin	Steroid	18.0 to 19.5 ng/mL	Prevents tumor cell proliferation and promote apoptosis by changing c-Myc and Bcl-2 expression. Also induces changes in cell cycle by reduction in topoisomerase II level	Meng et al. (2009)
Promelittin protoxin	*Apis milefera*	European honeybee	Honeybee	Europe and South East Asia	Venom	Protease	5.2 µmol/L	Possess significant inhibition of human prostate and breast cancer	Lebeau et al. (2009)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Polypeptide extract	*Buthus martensii kirsch*	Chinese golden scorpian	Scorpian	China Japan and Korea	Venom	–	40 µg/mL	Moderates expression of proteins related to cell cycle and apoptosis subsequently induces growth suppression and apoptosis in cancer cells	Zhang et al. (2009)
Cromatine	*Agkistrodon acutus*	South American rattlesnake	Snake	South America	Venom	–	–	Inhibit skin melanoma and human urinary cancer cell attached to fibronectin	Sánchez et al. (2009)
Viperistatin	*Vipera palEastinae*	Palestine viper	Snake	Western Syria Northern and Central America	Venom	Protein	5 mg/kg	Possess inhibitory properties against collagen receptors and integrins	Staniszewska et al. (2009)
Cinobufalin	*Bufo bufo gargarizans*	Asiatic toad	Toad	East Asia	Skin	Steroid	18.0 to 19.5 ng/mL	Prevent proliferation of tumor cell and induce apoptosis by modifing expression of Bcl-2 and c-Myc. It blocks proliferation of cancer cells by MAPK activation and also promotes cell cycle alterations by decreasing levels of topoisomerase II	Meng et al. (2009)
Colombistatin	*Bothrops colombiensis*	Common lancehead	Snake	Northern south America	Venom	Protein	–	Inhibition of ADP-induced platelet aggregation	Sánchez et al. (2009)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
MVL-PLA2	*Macrovipera lebetina*	Blunt-nosed viper	Snake	North Africa	Venom	Phospholipase	–	Possess angiogenesis inhibition and induces actin cytoskeleton changes	Kessentini-Zouari et al. (2010)
Dendroptoxin-κ	*Dendroaspis polylepis*	Black mamba	Snake	Africa	Venom	Small polypeptide	5 nM	It has anti-tumor properties in cancer cells through the pathway leading G_1-S transition. Increased expression of p27Kip1, p21Waf1/Cip1, and p15INK4B protein and prevent activity of cyclin-dependent kinases (cdk) which regulate the cell growth	Jang et al. (2010)
Contortrostatin	*Agkistrodon controtrix*	Southern copperhead snake	Snake	North America	Venom	Peptide	13.8 nM	Prevent platelet aggregation, obstruct growth of cancer cell, adhesion, relocation and angiogenesis	Lucena et al. (2011)
Polybia-cp	*Polybia paulista*	Neotropical wasp	Wasp	Almost found everywhere	Venom	Peptide	15 µm	Possess cytotoxic effect on cancer cells in dose and time dependent manner through disturbing cell membrane integrity. It prevents cancer cell metastasis	Wang et al. (2011)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Mojastin1	*Crotalus scutulatus*	Mohave rattlesnake	Snake	Deaserts of Southern West US and Central Mexico	Venom	Peptide	13.8 nM	Prevent ADP-induced platelet aggregation, cell migration, cell invasion and tumor colonization	Lucena et al. (2011)
Leucurogin	*Bothrops leucurus*	Whitetail lancehead	Snake	Brazil	Venom	Peptide	–	It possesses antiangiogenesiss in cancer cells	Higuchi et al. (2011)
Leucurolysin-B	*Bothrops leucurus*	Whitetail lancehead	Snake	Brazil	Venom	Metalloproteinase	–	Possess cytotoxicity against U87, T98, MCF7, RT2, EAC, and UACC cancer cell lines	Naumann et al. (2011)
Vicrostatin	*Echis carinatus*	Indian saw-scaled viper	Snake	Middle East and Central Asia	Venom	Peptide	–	Inhibits migration of human umbilical vein endothelium cells	Minea et al. (2012)
Disintegrin	*Vipera xanthine*	–	Snake	Most parts of the world	Venom	Small protein (45-84AA)	–	Inhibits tumor invasion and blocks OVCAR-5 adhesion to extracellular matrix proteins	Markland and Swenson (2013)
Phospholipases A_2	*Cerastes cerastes*	Saharan horned viper	Snake	Northern Africa	Venom	Enzyme	–	It possesses antiangiogenic and antitumor properties and act particularly on integrins αβ and α5β1	Zouari-Kessentini et al. (2013)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
Pooled venom	*Montivipera xanthine*	Ottoman viper	Snake	Northeastern Greece and Turkey	Venom	–	–	Possess cytotoxic effect and inhibits cell proliferation in dose dependent way	Yalcın et al. (2014)
Viridistatin 2	*Crotalus viridis*	Prairie rattlesnake	Snake	Western United States, northern Mexico and southwestern Canada,	Venom	Peptide	–	Inhibits proliferation, migration, adhesion, and survival through apoptosis in human pancreatic carcinoma (BXPC-3)cells	Lucena et al. (2015)
Tzabcanin	*Crotalus tzab can*	Yucatan Neotropical Rattlesnake	Snake	-	Venom	Protein	–	Inhibits melanoma (A-375) cells adhesion and lung cancer (A-549) to vitronectin	Saviola et al. (2016)
Chlorotoxin	*Leiurus quinquestriatus*	Death stalkwer-scorpion	Scorpian	Deserts of North Africa	Venom	Peptide	–	Inhibits cancer cell invasion through disabling the membrane-bound matrix metalloproteinase 2	Ojeda et al. (2016)
Vidatox 30 CH	*Rhopalurus junceus*	Blue scorpionn	Scorpionn	Cuba, Dominican Republic and Central America	Venom	Protein	–	Enhanced cell proliferation and expression of cell growth, p53, p21 p16 and p27	Giovannini et al. (2017)
Ophiophagus hannah venom	*Ophiophagus Hannah*	King cobra	Snake	Southeast Asia	Venom	–	0.02 ng/mL	Showed inhibitory properties on tumour cell induced angiogenesis	Kerkkamp et al. (2018)

Table 1 (continued)

Compounds	Scientific name	Common name	Organism	Geographic location	Component used	Type of molecule	Dosage	Mechanism of action	References
L-amino acid oxidase	*Ophiophagus Hannah*	King Cobra	Snake	Southeast Asia	Venom	Enzyme	–	Declines thymidine uptake in murine fibrosarcoma, melanoma, colorectal cancer as well as reduce proliferation of cell	Hiu and Yap (2020)

(Heinen and da Veiga 2011; Oršolić 2012). Additionally, bee venom has also been used to manage adverse effects of cancer chemotherapy (Rady et al. 2017). Melittin, a main compound of bee venom accounts for 40–50% of venom's total dry weight. Melittin has been extensively studied for its anticancer activity and regarded as an alternative for cancer chemotherapy. Phospholipase A2 and melittin can target various cancer cells such as lung, renal, prostate, liver, mammary gland, etc. (Pal et al. 2015). Moon et al. (2006) reported that venom from bees was capable of apoptosis induction in human leukemic cells. Choi et al. (2014) reported that bee venom causes cell death by apoptotsis in lung cancer cells (A549 and NCI-H460) by increasing the death receptor 3 expression and NF-κB pathway inhibition. Anticancer effect of bee venom on colon cancer cells was also observed where activation of death receptors and inhibition of NF-κB regulates cancer cell death (Zheng et al. 2015). Jung et al. (2018) reported apoptotic effect of bee venom on human breast cancer cells (MDA-MB-231) and observed protein denaturation, degradation and DNA fragmentation in cancer cells. Ceremuga et al. (2020) reported that melittin was able to induce apoptosis in human leukaemia cells and strongly reduced cell viability.

Beetle Venom

Various blister beetles produce toxic substances which when exposed to, skin causes blistering. Cantharidin is one such compound taken from *Mylabris caragnae,* and its dried bodies have been utilized for cancer treatment. Rauh et al. (2007) reported that cantharadin can inhibit human leukemic cells growth and acts as a leukemia progenitor (Dorn et al. 2009). Lee et al. (2010) reported that cantharidin is phosphoprotein phosphatase 1 and 2A inhibitor, which leads to damage of DNA and apoptosis. Verma and Prasad (2013) reported that cantharidin from blister beetles induces severe apoptosis in Ehrlich ascites carcinoma cells, which were involved in mitochondrial intrinsic pathway. A study reported the effect of cantharidin on human bladder carcinoma cell metastasis, and observed that cantharidin treatment blocked expression of genes, protein levels, and matrix metalloproteinase -2 (MMP-2) and MMP-9 activity (Huang et al. 2013). Another study reported that cantharidin led to apoptosis, reduction in growth and migration of human breast cancer cells (Shou et al. 2013). Various cantharidin derivatives inhibit growth of oral, cervical, prostate, colon, gall bladder cancer cell lines (Pal et al. 2015).

Caterpillar Venom

Few studies are available which report anticancer activity of caterpillar venom. Cecropins (from *Hyalophora cecropia*) are a group of peptides which have been utilized as a

potential anticancer agent against various tumour cell lines (Suttmann et al. 2008). Mechanism of action of cecropins against cancer cells involves pore formation in the cell membrane (Chen et al. 1997). A study reported that cecropins possess anticancerous activity against various leukemias and lymphomas (Moore et al. 1994). Studies revealed that cecropin B enhances lifespan of infected mice having colon adenocarcinoma cells. Suttmann et al. (2008) reported that cecropin A and B prevent viability of bladder cancer cell proliferation with no effect on fibroblasts. Anticancer activity of these peptides relies on target cell membrane disruption. Heinen et al. (2014) reported that exposure of *L. obliqua* venom to glioma cell lines decreases the nitric oxide production and activates cAMP signaling, which indicates the potential role of venom in glioma cell line inhibition.

Scorpion Venom

Scorpion venoms are a composite mixture of wide range of molecules which play a vital role in defense mechanism. Scorpion venoms generally show lower levels of enzymatic activity. They comprise of phospholipases, hyaluronidases, mucopolysaccharides, low molecular weight compounds, mucus, inorganic salts, etc. (Fernandes-Pedrosa et al. 2013). Scorpion venom can bind to a specific type of cells such as cancer cells (Pal et al. 2015). Approximately, 250 biologically active proteinous compounds have been identified from around 1500 scorpion species that have shown anticancer activity (Possani et al. 2000). Long chain toxins are the most studied venoms which primarily trigger sodium channels (Goudet et al. 2002) whereas short chain toxins trigger potassium or chloride channels which is essential for protection from predators (Goudet et al. 2002). Li et al. (2014) reported that venom from *Buthus matensii karsch* was able to inhibit human breast cancer cell growth by inducing apoptosis, blocking cell cycle progression and down-modulating protein levels of cyclin D1. Li et al. (2018a) reported that gonearrestide showed potent anticancer activity and has an impactful role in wide range of human cancers. They observed that gonearrestide inhibits primary colon cancer cell and solid tumour growth by triggering cell cycle arrest in G1 phase through cyclin dependent kinases 4 inhibition and upregulates cyclin D3, p27, and p21 expression.

Snake Venom

Snake venom is an amalgamation of peptides, enzymes, minerals, carbohydrates and proteins of low molecular mass (Calderon et al. 2014). Venoms are further classified into three groups i.e. cardiotoxins, hemotoxins and neurotoxins. Cardiotoxins bind to specific sites in heart muscle cells and prevent contraction of muscles (Dong and Chen 2018). Neurotoxins have a detrimental impact on central nervous system, resulting in breathing problems or cardiac arrest. They are capable of blocking the movement of ions across cell membrane (Bradbury and Deane 1993). Toxins that lead to the destruction of RBC's (Red Blood Cells) are known as hemotoxins. Cobras, mambas and coral snakes possess neurotoxic venom whereas copper heads, rattle snake, and cotton heads contain hemotoxic venoms. Some snakes contain both hemotoxins and neurotoxins. Calmette et al., (1993) first reported the use of venom for cancer treatment. Da Silva et al. (1996) reported that *Crotalus durissus teri ficus* and *Bothrops jararaca* venoms act directly on the cancer cells and the anticancer activity is possibly due to indirect response of inflammation facilitated by interleukin 2, interleukin 8, and Tumor necrosis factor-α. Anticancer activities of crude venom of *Naja kaouthia* and *Vipera russelli* were investigated on sarcoma, carcinoma, and leukemia models. Debnath et al. (2007) reported that there was an improved lifespan with increased anti-oxidant activity when monocellate cobra and Russell's viper venoms were used for the treatment of Ehrlich ascites carcinoma mice. Debnath et al. (2010) also reported that cardiotoxic-cytotoxic protein from *Naja kaouthia* shows antiproliferative activity and induces apoptosis in human leukemic cells. *Vipera lebentina turnica* venom induces apoptosis in ovarian cancer cells by STAT3 and NF-kB signal inhibition, followed by p65 and p50 inhibition, it also enhanced the Bax and Caspase-3 gene expression and downregulates Bcl-2 protein (Song et al. 2012). Kerkkamp et al. (2018) reported that crude venom from *Ophiophagus hanna* showed cytotoxic activity on pancreatic cancer cells where apoptosis was induced and migration activity was reduced. Fung et al. (2018) found that LAAO purified from *Ophiophagus hanna* possessess antiproliferative activity against human breast adenocarcinoma and human lung adenocarcinoma cell lines with EC50 values of 0.04 µg/mL and 0.05 µg/mL, respectively. Malekara et al. (2020) reported effect of *Vipera raddei kurdistanica* venom on breast cancer cells and suggested that venom possess cytotoxicity and leads to apoptosis in cancer cells. Conlon et al. (2020) isolated three toxins from *M. nigricollis* and evaluated its cytotoxicity against human tumor cells.

Spider Venom

Venoms from spiders are a good source of biologically active therapeutic compounds. It is a composite mixture of polypeptides, proteins, free amino acids, nucleic acids, neurotoxins, inorganic salts and monoamines that showed various impacts in vertebrates and invertebrates (Koua and Kuhn-Nentwig 2017). *Macrothele raveni* venom effectively suppresses the growth of cells in myelogenous leukaemia in a dose and time dependent way with an IC_{50} value of 5.1 µg/mL. Results showed that venom specifically suppresses growth of K562 cell by apoptosis induction through caspase

8 and caspase 3 associated signalling pathways (Liu et al. 2012). Davies et al. (2002) reported antitumor activity of *Macrothele raveni* venom on human breast cancer cell line (MCF-7) and observed tumor growth inhibition in nude mice, size of tumor remarkably decline compared to control group by twenty-one days of treatment. Some spiders contain hyaluronidases in their venom, which could be utilized for enhanced tissue permeability, facilitate drug penetration, or even used directly as anticancer agents (Girisk and Kemparaju 2007). Lian et al. (2018) reported that *Haplopelma hainanum* venom showed strong inhibitory effects on HepG2 cells via growth suppression, mitochondrial membrane potential reduction, caspase-3 and caspase-9 activation, and apoptosis induction through a mitochondrial dependent pathway. Okada et al. (2019) reported that a pore forming peptide (LaFr26) isolated from *Lachesanasp* venom was able to possess specific cytotoxicity to lung cancer cell lines.

Wasp Venom

Wasp's sting produces serious pain and damages tissue which may lead to death. Numerous investigations have been done on bioactive compounds present in the venom of wasp. Venom of wasp contains various peptides which can produce pores in cell membrane, enter the cell and leads to death either by apoptosis or necrosis. Though, wasp venom not only harms cancer cells but also affects normal cells. Therefore, investigators are finding a way to carrying the peptide to cancer location so that it accumulates in a specific site and released in a controlled manner. The system contains a carrier polymer, with two components: a peptide part that is attached to a cancer cell receptor, and another is cytotoxic peptide of wasp venom. Moreno et al. (2014) reported that substance is sufficiently diffused within the tumor cells, leading to cell death, whereas healthy cells are not influenced. Fujiwara et al. (2008) extracted and resolved the structure of an anticancer molecule from exterior cover of *Vespa simillima* which showed growth inhibition of liver cancer cells in rats. Cytotoxicity is linked to morphological variations that induce cell apoptosis. Polybia MPI (from *Polybia paulista*) is a wasp venom which also possess anticancer activity. de Souza et al. (2012) reported that Polybia MPI possess cytotoxic against leukemic T lymphocytes and is very selective in nature. Polybia MPI is capable of targeting non-polar lipid cell membranes by creating ion permeable channels which leads to permanent cytolysis, depolarization and death of the cell (Pal et al. 2015).

Clinical Study of Anticancer Drugs from Animal Source

Investigation concerning venoms have turned out to be an attractive area due to the current advancement in genomic and proteomic techniques (Munawar et al. 2018). These techniques allow a better way to study biologically active substances that are found in animal venom. Clinical trials are significant for finding of new methods to identify, detect, and lessen the risk of disease. It provides information about what can work or not in a human being. On an average it takes 10–12 years for a drug to enter from design to clinical trials. Clinical trials help in the identification of possible risks and benefits of a treatment. Peptides from venom sources have been under clinical trials. One of the compounds from venom that has been approved by FDA for use in cancer is BLZ100, which is being developed by Blaze Bioscience to image brain tumours and receives Orphan drug designation (Blaze-Bioscience). It is a drug modified from chlorotoxin and isolated from *Leiurus quinquestriatus* venom. Venoms based anticancer drugs that have undergone clinical trials are summarised in Table 2.

Future Prospect

So far researchers have isolated venoms from wasp, snake, honey bee, scorpionns etc. and characterized several of their biological actions. Animal venom is a heterogeneous mixture, responsible for various pharmacological activities. Venom contains different amino acid sequence; alteration can lead to structural changes which eventually changes immunological property of toxin. Studies need to be done to eliminate toxicity of venoms as different toxin molecules exhibit different toxicity mechanisms at various molecular, biochemical and cellular levels with some specific attention for nervous system. Additionally, anticancer activity of venoms can be improved by enhancing its cytolytic activity. To accomplish therapeutic requirement of drugs and chemicals, recombinant synthesis can be used for finding peptide toxins. One of the possible approaches for development of anti-cancer drugs from cytotoxins is by conjugating drugs to monoclonal antibodies which can identify and bind to precise cancerous cell epitopes. Another method could be combining venom with nanoparticles and delivering them to a specific site for better treatment. Alteration of amino acid sequence of venom peptide gives rise to new structural compounds that can be utilized for new toxin development to treat cancer. Animal venoms should be put under scientific scrutiny for its pharmaceutical potential.

Conclusion

Cancer is one of the leading causes of death worldwide, and there is an urgent need to find better treatment. For pharmacology and medicine, finding a cancer cure is one of the major challenges. Conventional methods possess several adverse, therefore it is required to find new compounds that

Table 2 Venom based anticancer drug in clinical trials

Drug	Mode of action	Target	Clinical phase	References
Chlorotoxin (^{131}I-TM601)	Cancer cell proliferation and invasion	Chloride channels	Phase III	Mamelak and Jacoby (2007)
Cilengtide	Cancer cell invasion, migration, angiogenesis and metastasis	Integrins $\alpha_v\beta_3$, $\alpha_v\beta_5$	Phase II and Phase III	Reardon et al. (2011)
Ergtoxin	Cancer cell proliferation and invasion	Potassium channels	Preclinical studies	Asher et al. (2011)
Analgesic-antitumor peptide	Cancer cell proliferation and invasion	Sodium channels	Preclinical studies	Gu et al. (2013)
BAY86-7548	Cancer cell metastasis	Gastrin relasing peptide receptor	Phase II/III	Kähkönen et al. (2013)
Vicrostatin	Cancer cell invasion, migration, angiogenesis and metastasis	Integrins $\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_5\beta_1$	Preclinical studies	Swenson et al. (2015)
Mastoparan	Cancer cell membrane disruption	Sialic acid rich glycoproteins, phodphatidlserine, phosphatidylethanolamine heparamsulfate	Preclinical studies	de Azevedo et al. (2015)
Hemilipin	Cancer cell membrane disruption	Phospholipids	Preclinical studies	Jridi et al. (2015)
Melittin	Cancer cell membrane disruption	Sialic acid rich glycoproteins, phodphatidlserine, phosphatidylethanolamine heparamsulfate	Preclinical studies	Gajski et al. (2016)
BLZ-100	Cancer cell proliferation and invasion	Chloride channels	Phase I	Miller et al. (2016)
Shrew peptide (SOR-C13)	Proliferation of cancer cell and invasion	Transient receptor potential channels	Phase I	Fu et al. (2017)

have the ability to treat cancer and also harmless in nature. Anticancer agents from natural sources, especially animal venom have confirmed their potential in cancer therapy. Extensive research has been done to obtain an effective compound from natural source. Venomous animals such as scorpions, snakes, etc., were widely studied for their therapeutic values with some advantages and some challenges. These compounds possess cytotoxic properties with numerous modes of action which include cancer cell growth inhibition, apoptosis induction, DNA damage and others. Several of these substances are undergoing preclinical trials or are in the successive clinical trial stages. Studies on venoms obtained from animals have shown that these compounds have antiproliferative, antioxidant, and antimutagenic activities which could help in cancer management. However, more research in compounds from animal sources results in more potent discovery of anticancer agents.

Compliance with Ethical Standards

Conflict of interest All 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

Aaghaz S, Gohel V, Kamal A (2019) Peptides as potential anticancer agents. Curr Top Med Chem 19(17):1491–1511

Abidin Z, Asnawi S, Lee YQ, Othman I, Naidu R (2019) Malaysian cobra venom: a potential source of anti-cancer therapeutic agents. Toxins 11(2):75

Al-Tamimi J, Semlali A, Hassan I, Ebaid H, Alhazza IM, Mehdi SH, Al-Khalifa M, Alanazi MS (2018) Samsum ant venom exerts anticancer activity through immunomodulation in vitro and in vivo. Cancer Biother Radiopharm 33(2):65

An WW, Gong XF, Wang MW, Tashiro S, Onodera S, Ikejima T (2004) Norcantharidin induces apoptosis in HeLa cells through caspase, MAPK and mitochondrial pathways. Acta Pharmacol Sin 25(11):1502–1508

Arbiser JL, Kau T, Konar M, Narra K, Ramchandran R, Summers SA, Vlahos CJ, Ye K, Perry BN, Matter W (2007) Solenopsin, the alkaloidal component of the fire ant (Solenopsisinvicta), is a naturally occurring inhibitor of phosphatidylinositol-3-kinase signaling and angiogenesis. Blood 109:560–565

Asher V, Warren A, Shaw R, Sowter H, Bali A, Khan R (2011) The role of Eag and HERG channels in cell proliferation and apoptotic

cell death in SK-OV-3 ovarian cancer cell line. Cancer Cell Int. 11(6):6
Belokoneva OS, Villegas E, Corzo G, Dai L, Nakajima T (2003) The hemolytic activity of six arachnid cationic peptides is affected by the phosphatidylcholine-to-sphingomyelin ratio in lipid bilayers. Biochim Biophys Acta 1617(1–2):22–30
Bradbury MW, Deane R (1993) Permeability of the blood-brain barrier to lead. Neurotoxicology 14:131–136
Blaze-Bioscience https://www.businesswire.com/news/home/20150707005437/en/Blaze-Bioscience%E2%80%99s-BLZ-100-Receives-Orphan-Drug-Designation
Calderon LA, Sobrinho JC, Zaqueo KD, de Moura AA, Grabner AN, Mazzi MV, Marcussi S, Nomizo A, Fernandes CF, Zuliani JP, Carvalho BM, da Silva SL, Stábeli RG, Soares AM (2014) Antitumoral activity of snake venom proteins: new trends in cancer therapy. Biomed Res Int 2014:203639
Calmette A, Saenz A, Costil L (1993) Effects of du venimde cobra sur les greffescancereuses et sur le cancer spontanea (adeno carcinoma) de la souris. C R Acad Sci 197:205
Ceremuga M, Stela M, Janik E, Gorniak L, Synowiec E, Sliwinski T, Bijak M (2020) Melittin-A natural peptide from bee venom which induces apoptosis in human leukaemia cells. Biomolecules 10(2):247
Chen HM, Wang W, Smith D, Chan SC (1997) Effects of antibacterial peptide cecropin B and its analogues, cecropin B-1 and cecropin B-2, onliposome, bacteria and cancer cells. Biochim Biophys Acta 1336:171–179
Choi KE, Hwang CJ, Gu SM, Park MH, Kim JH, Park JH, Ahn YJ, Kim JY, Song MJ, Song HS, Han SB, Hong JT (2014) Cancer cell growth inhibitory effect of bee venom via increase of death receptor 3 expression and inactivation of NF-kappa B in NSCLC cells. Toxins (Basel) 6:2210–2228
Cohen G, Burks SR, Frank JA (2018) Chlorotoxin—a multimodal imaging platform for targeting glioma tumors. Toxins 10(12):496
Conlon JM, Attoub S, Musale V, Leprince J, Casewell NR, Sanz L, Calvete JJ (2020) Isolation and characterization of cytotoxic and insulin-releasing components from the venom of the black-necked spitting cobra *Naja nigricollis* (Elapidae). Toxicon X 6:100030
Corzo G, Gilles N, Satake H, Villegas E, Dai L, Nakajima T, Haupt J (2003) Distinct primary structure of the major peptides toxins from the venom of the spider *Macrothelegigas* that bind to sites 3 and 4 in the sodium channel. FEBS Lett 547(1–3):43–50
Davies G, Martin LA, Sacks N, Dowsett M (2002) Cyclooxygenase-2 (COX-2), aromatase and breast cancer: a possible role for COX 2 inhibitor in breast cancer chemoprevention. Ann Oncol 13:669–678
Da Silva RJ, Fecchio D, Barraviera B (1996) Antitumor effect of snake venom. J Venom Anim Toxins 2(2):79–90
De Carvalho DD, Schmitmeier S, Novello JC, Markland FS (2001) Effect of BJcuL (a lectin from the venom of the snake *Bothropsjararacussu*) on adhesion and growth of tumor and endothelial cells. Toxicon 39:1471–1476
de Azevedo RA, Figueiredo CR, Ferreira AK, Matsuo AL, Massaoka MH, Girola N, Auada AV, Farias CF, Pasqualoto KF, Rodrigues CP, Barbuto JA, Levy D, Bydlowski SP, de Sá-Junior PL, Travassos LR, Lebrun I (2015) Mastoparan induces apoptosis in B16F10-Nex2 melanoma cells via the intrinsic mitochondrial pathway and displays antitumor activity *in vivo*. Peptides 68:113–119
de Souza BM, da Silva AV, Resende VM, Arcuri HA, Dos Santos Cabrera MP, Ruggiero Neto J, Palma MS (2012) Characterization of two novel polyfunctional mastoparan peptides from the venom of the social wasp *Polybia paulista*. Peptides 30(8):1387–1395
Debnath A, Saha A, Gomes A, Biswas S, Chakrabarti P, Giri B, Biswas AK, Gupta SD, Gomes AA (2010) lethal cardiotoxic-cytotoxic protein from the Indian monocellate cobra (*Naja kaouthia*) venom. Toxicon 56:569–579
Debnath A, Chatterjee U, Das M, Vedasiromoni JR, Gomes A (2007) Venom of Indian monocellate cobra and Russell"s viper show anticancer activity in experimental models. J Ethnopharmacol 111:681
Dong J, Chen H (2018) Cardiotoxicity of anticancer therapeutics. Front Cardiovasc Med 5:9
Dorn DC, Kou CA, Png KJ, Moore MA (2009) The effect of cantharidins on leukemic stem cells. Int J Cancer 124:2186–2199
Ejaz S, Hashmi FB, Malik WN, Ashraf M, Iqbal M (2018) Applications of venom proteins as potential anticancer agents. Protein Peptide lett 25(7):688–701
Feng L, Gao R, Gopalakrishnakone P (2008) Isolation and characterization of a hyaluronidase from the venom of Chinese red scorpionn *Buthusmartensi*. Comp Biochem Physiol C Toxicol Pharmacol 148(3):250–257
Fernandes-Pedrosa MF, Félix-Silva J, Menezes Y (2013) Toxins from venomous animals: gene cloning, protein expression and biotechnological applications. Integr View Mol Recognit Toxinologu Anal Proced Biomed Appl 2:23–72
Fu S, Hirte H, Welch S, Ilenchuk TT, Lutes T, Rice C, Fields N, Nemet A, Dugourd D, Piha-Paul S, Subbiah V, Liu L, Gong J, Hong D, Stewart JM (2017) First-in-human phase I study of SOR-C13, a TRPV6 calcium channel inhibitor, in patients with advanced solid tumors. Invest New Drugs 35:324–333
Fujiwara Y, Mangetsu M, Yang P, Kofujita H, Suzuki K, Ohfune Y, Shinada T (2008) A quinone isolated from the nest of *Vespa simillima* and its growth-inhibitory effect on rat liver cancer cells. Biol Pharm Bull 31:722–725
Fung SY, Lee ML, Tan NH (2018) Proteomic investigation of the molecular mechanism of king cobra venom L-amino acid oxidase induced apoptosis of human breast cancer (MCF-7) cell line. Indian J Exp Biol 56:101–111
Gajski G, Domijan AM, Žegura B, Štern A, Gerić M, Novak Jovanović I, Vrhovac I, Madunić J, Breljak D, Filipič M, Garaj-Vrhovac V (2016) Melittin induced cytogenetic damage, oxidative stress and changes in gene expression in human peripheral blood lymphocytes. Toxicon 110:56–67
Galán JA, Sánchez EE, Rodríguez-Acosta A, Soto JG, Bashir S, McLane MA, Paquette-Straub C, Pérez JC (2008) Inhibition of lung tumor colonization and cell migration with the disintegrin crotatroxin 2 isolated from the venom of *Crotalus atrox*. Toxicon 51(7):1186–1196
Gao L, Shan BE, Chen J, Liu JH, Song DX, Zhu BC (2005) Effects of spider *Macrothele raven* venom on cell proliferation and cytotoxicity in HeLa cells. Acta Pharmacol Sin 26(3):369–376
Giovannini C, Baglioni M, Baron Toaldo M, Cescon M, Bolondi L, Gramantieri L (2017) Venom from Cuban Blue Scorpionn has tumor activating effect in hepatocellular carcinoma. Sci Rep 21(7):44685
Girisk KS, Kemparaju K (2007) The magic glue hyaluronan and its eraser hyaluronidases: a biologically overview. Life Sci 80:921–943
Goudet C, Chi CW, Tytgat J (2002) An overview of toxins and genes from the venom of the Asian scorpionn *Buthusmartensi*Karsch. Toxicon 40:403–410
Gu Y, Liu SL, Ju WZ, Li CY, Cao P (2013) Analgesic-antitumor peptide induces apoptosis and inhibits the proliferation of SW480 human colon cancer cells. Oncol Lett 5:483–488
Heinen TE, da Veiga AB (2011) Arthropod venoms and cancer. Toxicon 57(4):497–511
Heinen TE, de Farias CB, Abujamra AL, Mendonça RZ, Roesler R, da Veiga AB (2014) Effects of *Lonomiaobliqua* caterpillar venom upon the proliferation and viability of cell lines. Cytotechnology 66(1):63–74

Higuchi DA, Almeida MC, Barros CC, Sanchez EF, Pesquero PR, Lang EAS, Samaan M, Araujo RC, Pesquero JB, Pesquero JL (2011) Leucurogin, a new recombinant disintegrin cloned from *Bothrops leucurus* (white-tailed-jararaca) with potent activity upon platelet aggregation and tumor growth. Toxicon 58:123–129

Huang YP, Ni CH, Lu CC, Chiang JH, Yang JS, Ko YC, Lin JP, Kuo JH, Chang SJ, Chung JG (2013) Suppressions of migration and invasion by cantharidin in TSGH-8301 human bladder carcinoma cells through the inhibitions of matrix metalloproteinase-2/-9 signalling. Evid Based Complement Alternat Med 2013:190281

Hiu JJ, Yap MKK (2020) Cytotoxicity of snake venom enzymatic toxins: phospholipase A2 and l-amino acid oxidase. Biochem Soc Trans 48(2):719–731

Jang MH, Shin MC, Lim S, Han SM, Park HJ, Shin I, Lee JS, Kim KA, Kim EH, Kim CJ (2003) Bee venom induces apoptosis and inhibits expression of cyclooxygenase-2 mRNA in human lung cancer cell line NCI-H1299. J Pharmacol Sci 91:95–104

Jang SH, Ryu PD, Lee SY (2010) Dendrotoxin-κ suppresses tumor growth induced by human lung adenocarcinoma A549 cells in nude mice. J Vet Sci 12(1):35–40

Jridi I, Catacchio I, Majdoub H, Shahbazeddah D, El Ayeb M, Frassanito MA, Ribatti D, Vacca A, Borchani L (2015) Hemilipin, a novel *Hemiscorpius lepturus* venom heterodimeric phospholipase A2, which inhibits angiogenesis in vitro and in vivo. Toxicon 105:34–44

Jung GB, Huh JE, Lee HJ, Kim D, Lee GJ, Park HK, Lee JD (2018) Anti-cancer effect of bee venom on human MDA-MB-231 breast cancer cells using Raman spectroscopy. Biomed Opt Express 9(11):5703–5718. https://doi.org/10.1364/BOE.9.005703

Kähkönen E, Jambor I, Kemppainen J, Lehtiö K, Grönroos TJ, Kuisma A, Luoto P, Sipilä HJ, Tolvanen T, Alanen K, Silén J, Kallajoki M, Roivainen A, Schäfer N, Schibli R, Dragic M, Johayem A, Valencia R, Borkowski S, Minn H (2013) *In vivo* imaging of prostate cancer using [68Ga]-labeled bombesin analog BAY86-7548. Clin Cancer Res 19(19):5434–5443

Kerkkamp H, Bagowski C, Kool J, van Soolingen B, Vonk FJ, Vlecken D (2018) Whole snake venoms: cytotoxic, anti-metastatic and antiangiogenic properties. Toxicon 150:39–49

Kessentini-Zouari R, Jebali J, Taboubi S, Srairi-Abid N, Morjen M, Kallech-Ziri O, Bezzine S, Marvaldi J, El Ayeb M, Marrakchi N, Luis J (2010) CC-PLA2-1 and CC-PLA2-2, two Cerastescerastesvenom-derived phospholipases A_2, inhibit angiogenesis both in vitro and in vivo. Lab Invest 90(4):510–519

Kim DS, Jang YJ, Jeon OH, Kim DS (2007) Saxatilin, a snake venom disintegrin, suppresses TNF-α induced ovarian, cancer cell invasion. J Biochem Mol Biol 40(2):290–294

Klupczynska A, Pawlak M, Kokot ZJ, Matysiak J (2018) Application of metabolomic tools for studying low molecular-weight fraction of animal venoms and poisons. Toxins (Basel) 10(8):306

Koua D, Kuhn-Nentwig L (2017) Spider neurotoxins, short linear cationic peptides and venom protein classification improved by an automated competition between exhaustive profile HMM classifiers. Toxins (Basel) 9(8):245

Kumar N, Tomar R, Pandey A, Tomar V, Singh VK, Chandra R (2017a) Preclinical evaluation and molecular docking of 1,3-benzodioxole propargyl ether derivatives as novel inhibitor for combating the histone deacetylase enzyme in cancer. Artif Cells Nanomed Biotechnol 46(6):1288–1299

Kumar N, Chungh H, Tomar R, Tomar V, Singh VK, Chandra R (2017b) Exploring the interplay between autoimmunity and cancer to find the target therapeutic hotspots. Artif Cells Nanomed Biotechnol 46(4):658–668

LeBeau AM, Brennen WN, Aggarwal S, Denmeade SR (2009) Targeting the cancer stroma with a fibroblast activation protein-activated promelittin protoxin. Mol Cancer Ther 8(5):1378–1386

Lee DH, Pan Y, Kanner S, Sung P, Borowiec JA, Chowdhury D (2010) A PP4 phosphatase complex dephosphorylates RPA2 to facilitate DNA repair via homologous recombination. Nat Struct Mol Biol 17(3):365–372

Li B, Lyu P, Xi X, Ge L, Mahadevappa R, Shaw C, Kwok HF (2018a) Triggering of cancer cell cycle arrest by a novel scorpionn venom-derived peptide-Gonearrestide. J Cell Mol Med 22(9):4460–4473. https://doi.org/10.1111/jcmm.13745

Li L, Huang J, Lin Y (2018b) Snake venoms in cancer therapy: past, present and future. Toxins 10(9):346

Li W, Li Y, Zhao Y, Yuan J, Mao W (2014) Inhibition effects of scorpion venom extracts (*Buthus matensii* Karsch) on the growth of human breast cancer MCF-7 cells. Afr J Tradit Complement Altern Med 11:105–110

Lian W, Lian H, Li Q, Hu A, Liu S (2018) The venom of spider Haplopelma hainanum suppresses proliferation and induces apoptosis in hepatic cancer cells by caspase activation in vitro. J Ethnopharmacol 28(225):169–177

Liu Z, Zhao Y, Li J, Xu S, Liu C, Zhu Y, Liang S (2012) The venom of the spider *Macrothele raveni* induces apoptosis in the myelogenous leukemia K562 cell line. Leuk Res 36(8):1063–1066

Lucena S, Sanchez EE, Perez JC (2011) Anti-metastatic activity of the recombinant disintegrin r-mojastin 1, from the mohave rattlesnake. Toxicon 57(5):794–802

Lucena S, Castro R, Lundin C, Hofstetter A, Alaniz A, Suntravat M, Sánchez EE (2015) Inhibition of pancreatic tumoral cells by snake venom disintegrins. Toxicon 93:136–143

Malekara E, Pazhouhi M, Rashidi I, Jalili C (2020) Anti-proliferative and cytotoxic effect of Iranian snake (*Vipera raddei kurdistanica*) venom on human breast cancer cells via reactive oxygen species-mediated apoptosis. Res Pharm Sci 15(1):76

Mamelak AN, Jacoby DB (2007) Targeted delivery of antitumoral therapy to glioma and other malignancies with synthetic chlorotoxin (TM-601). Expert Opin Drug Deliv 4:175–186

Marcinkiewicz C, Weinreb PH, Calvete JJ, Kisiel DG, Mousa SA, Tuszynski GP, Lobb RR (2003a) Obtustatin: a potent Selective Inhibitor of alpha1 beta1 Integrin *in vitro* and angiogenesis *in vivo*. Cancer Res 63:2020–2023

Markland FS, Swenson S (2013) Snake venom metalloproteinases. Toxicon 62:3–18

Marcinkiewicz C, Lobb RR, Marcinkiewicz MM, Daniel JL, Smith JB, Dangelmaier C, Weinreb PH, Beacham DA, Niewiarowski S (2000) Isolation and characterization of EMS16, a C-lectin type protein from Echismultisquamatus venom, a potent and selective inhibitor of the alpha 2 beta 1 integrin. Biochemistry 39(32):9859–9867

Marcinkiewicz C, Weinreb PH, Calvete JJ, Kisiel DG, Mousa SA, Tuszynski GP, Lobb RR (2003b) Obtustatin: a potent selective inhibitor of α1β1 integrin in vitro and angiogenesis in vivo. Cancer Res 63:2020–2023

Meng Z, Yang P, Shen Y, Bei W, Zhang Y, Ge Y, Newman RA, Cohen L, Liu L, Thornton B, Chang DZ, Liao Z, Kurzrock R (2009) Pilot study of huachansu in patients with hepatocellular carcinoma, non-small cell lung cancer or pancreatic cancer. Cancer 115(22):5309–5318

Miller D, Patil C, Walker D, Kittle D, Nufer K, Yamada M, Butte P, Prow T, Novak J, Black K, Mamelak A (2016) Phase 1 safety study of blz-100 for fluorescence-guided resection of glioma in adult subjects. Neuro- Oncol 18:12–13

Minea R, Helchowski C, Rubino B, Brodmann K, Swenson S, Markland F Jr (2012) Development of a chimeric recombinant disintegrin as a cost-effective anti-cancer agent with promising translational potential. Toxicon 59(4):472–486

Moon DO, Park SY, Heo MS, Kim KC, Park C, Ko WS (2006) Key regulators in bee venom induced apoptosis are Bcl-2 and

caspase-3 in human leukemic U937 cells through down regulation of ERK and Akt. Int Immunopharmacol 6(12):1796–1807
Moreno M, Zurita E, Giralt E (2014) Delivering wasp venom for cancer therapy. J Control Release 182:13
Munawar A, Ali SA, Akrem A, Betzel C (2018) Snake venom peptides: tools of biodiscovery. Toxins 10(11):474
Moore AJ, Devine DA, Bibby MC (1994) Preliminary experimental anticancer activity of cecropins. Pept Res 7:265–269
Naumann GB, Silva LF, Silva L, Faria G, Richardson M, Evangelista K, Kohlhoff M, Gontijo CM, Navdaev A, de Rezende FF, Eble JA, Sanchez EF (2011) Cytotoxicity and inhibition of platelet aggregation caused by an L-amino acid oxidase from *Bothrops leucurus* venom. Biochem Biophys Acta 1810(7):683–694
Ojeda PG, Wang CK, Craik DJ (2016) Chlorotoxin: structure, activity, and potential uses in cancer therapy. Biopolymers 106(1):25–36
Okada M, Ortiz E, Corzo G, Possani LD (2019) Pore-forming spider venom peptides show cytotoxicity to hyperpolarized cancer cells expressing K^+ channels: a lentiviral vector approach. PLoS ONE 14(4):e0215391
Olfa KZ, José L, Salma D, Amine B, Najet SA, Nicolas A, Maxime L, Raoudha Z, Kamel M, Jacques M, Jean-Marc S, el Mohamed A, Naziha M (2005) Lebestatin, a disintegrin from macrovipera venom, inhibits integrin-mediated cell adhesion, migration and angiogenesis. Lab Invest 85:1507–1516
Oršolić N (2012) Bee venom in cancer therapy. Cancer Metastasis Rev 31(1–2):1
Pal P, Roy S, Chattopadhyay S, Pal TK (2015) Medicinal value of animal venom for treatment of cancer in humans—a review. World Sci News 22:128–144
Possani LD, Merino E, Corona M, Bolivar F, Becerril B (2000) Peptides and genes coding for scorpionn toxins that affect ion-channels. Biochimie 82(9–10):861–868
Premzl A, Kovacic L, Halassy B, Krizaj I (2008) Generation of ammodytoxin-anti-cathepsin B immune-conjugate as a model for delivery of secretory phospholipase A2 into cancer cells. Toxicon 51(5):754–764
Rauh R, Kahl S, Boechzelt H, Bauer R, Kaina B, Efferth T (2007) Molecular biology of cantharidin in cancer cells. Chin Med 2:8
Rady I, Siddiqui IA, Rady M, Mukhtar H (2017) Melittin, a major peptide component of bee venom, and its conjugates in cancer therapy. Cancer Lett 402:16–31
Reardon DA, Neyns B, Weller M, Tonn JC, Nabors LB, Stupp R (2011) Cilengitide: an RGD pentapeptide αvβ3 and αvβ5 integrin inhibitor in development for glioblastoma and other malignancies. Future Oncol 7:339–354
Roy A, Jauhari N, Bharadvaja N (2018) Medicinal plants as a potential source of chemopreventive agents natural products and biotechnological implements anticancer plants. Springer, Singapore, pp 109–139
Roy A, Ahuja S, Bharadvaja N (2017) A review on medicinal plants against cancer. J Plant Sci Agric Res 2:008
Roy A, Bharadvaja N (2017) Medicinal plants in the management of cancer: a review. Int J Complement Alt Med 9(2):00291
Sachdeva V, Roy A, Bharadvaja N (2020) Current prospects of nutraceuticals: a review. Curr Pharm Biotechnol. https://doi.org/10.2174/1389201021666200130113441
Sahni A, Simpson-Haidaris PJ, Sahni SK, Vaday GG, Francis CW (2007) Fibrinogen synthesized by cancer cells augments the proliferative effects of fibroblast growth factor-2 (FGF-2). J Thromb Haemost 6:176–183
Sánchez EE, Rodríguez-Acosta A, Palomar R, Lucena SE, Bashir S, Soto JG, Pérez JC (2009) Colombistatin: a disintegrin isolated from the venom of the South American snake (*Bothropscolombiensis*) that effectively inhibits platelet aggregation and SK-Mel-28 cell adhesion. Arch Toxicol 83(3):271–279
Sarray S, Delamarre E, Marvaldi J, Ayeb M, Marrakchi N, Luis J (2007) Lebectin and lebecetin, two C-type lectins from snake venom, inhibit α5β1 and αv-containing integrins. Matrix Biol 26(4):306–313
Saviola AJ, Burns PD, Mukherjee AK, Mackessy SP (2016) The disintegrintzabcanin inhibits adhesion and migration in melanoma and lung cancer cells. Int J Biol Macromol 88:457–464
Shen S, Khazaeli MB, Gillespie GY, Alvarez VL (2005) Radiation dosimetry of 131I-chlorotoxin for targeted radiotherapy in glioma-bearing mice. J Neurooncol 71(2):113–119
Shin VY, Jin HC, Ng EK, Yu J, Leung WK, Cho CH, Sung JJ (2008) Nicotine and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone induce cyclooxygenase-2 activity in human gastric cancer cells: Involvement of nicotinic acetylcholine receptor (nAChR) and β-adrenergic receptor signaling pathways. Toxicol Appl Pharmacol 233:254–261
Shou LM, Zhang QY, Li W, Xie X, Chen K, Lian L, Li ZY, Gong FR, Dai KS, Mao YX, Tao M (2013) Cantharidin and norcantharidin inhibit the ability of MCF-7cells to adhere to platelets via protein kinase C pathway-dependent downregulation of α2 integrin. Oncol Rep 30:1059–1066
Siegel RL, Miller KD, Ahmedin Jemal DVM (2019) Cancer statistics, 2019. CA Cancer J Clin 69:7–34
Siegel RL, Miller KD, Jemal A (2018) Cancer statistics, 2018. CA Cancer J Clin 68(1):7–30
Son DJ, Park MH, Chae SJ, Moon SO, Lee JW, Song HS, Moon DC, Kang SS, Kwon YE, Hong JT (2007) Inhibitory effect of snake venom toxin from *Viperalebetinaturanica*on hormone-refractory human prostate cancer cell growth: induction of apoptosis through inactivation of nuclear factor kappaB. Mol Cancer Ther 6(2):675–683
Song JK, Jo MR, Park MH, Song HS, An BJ, Song MJ, Han SB, Hong JT (2012) Cell growth inhibition and induction of apoptosis by snake venom toxin in ovarian cancer cell via inactivation of nuclear factor κB and signal transducer and activator of transcription 3. Arch Pharm Res 35:867–876
Staniszewska I, Walsh EM, Rothman VL, Gaathon A, Tuszynski GP, Calvete JJ, Lazarovici P, Marcinkiewicz C (2009) Effect of VP12 and viperistatin on inhibition of collagen receptors-dependent melanoma metastasis. Cancer Biol Ther 8(15):1507–1516
Suárez Y, González L, Cuadrado A, Berciano M, Lafarga M, Muñoz A (2003) Kahalalide F, a new marine-derived compound, induces oncosis in human prostate and breast cancer cells. Mol Cancer Ther 2(9):863–872
Suttmann H, Retz M, Paulsen F, Harder J, Zwergel U, Kamradt J, Wullich B, Unteregger G, Stockle M, Lehmann J (2008) Antimicrobial peptides of cecropin family show potent antitumor activity against bladder cancer cells. BMC Urol 3:8
Swenson S, Minea R, Moore J, Mikhail M, Kirakosyan Z, Markland Jr FS (2015) Intraperitoneal delivery of the disintegrin vicrostatin (VCN) effectively limits ovarian cancer growth and progression. Clin Cancer Res 21(16 Suppl):1432. https://doi.org/10.1158/1557-3265.OVCASYMP14-POSTER-THER-1432
Tian J, Paquette-Straub C, Sage EH, Funk SE, Patel V, Galileo D, McLane MA (2007) Inhibition of melanoma cell motility by the snake venom disintegrin eristostatin. Toxicon 49(7):899–908
Tomar V, Kumar N, Tomar R, Sood D, Dhiman N, Dass SK, Prakash S, Madan J, Chandra R (2019) Biological evaluation of noscapine analogues as potent and microtubule-targeted anticancer agents. Sci Rep 9:19542
Verma AK, Prasad SB (2013) Antitumor effect of blister beetles: an ethno-medicinal practice in Karbi community and its experimental evaluation against a murine malignant tumor model. J Ethnopharmacol 148(3):869
Wang K, Yan J, Liu X, Zhang J, Chen R, Zhang B, Dang W, Zhang W, Kai M, Song J, Wang R (2011) Novel cytotoxity exhibition mode

of polybia-CP, a novel antimicrobial peptide from the venom of the social wasp *Polybiapaulista*. Toxicology 288:27–33

Wang L, Dong C, Li X, Han W, Su X (2017) Anticancer potential of bioactive peptides from animal sources. Oncol Rep 38(2):637–651

Yang RS, Tang CH, Chuang WJ, Huang TH, Peng HC, Huang TF, Fu WM (2005) Inhibition of tumor formation by sanke venom disintegrin. Toxicon 45:661–669

Yalcın HT, Ozen MO, Gocmen B, Nalbantsoy A (2014) Effect of ottoman viper (*Montiviperaxanthina* (Gray, 1849)) venom on various cancer cells and on microorganisms. Cytotechnology 66(1):87–94

Zhang YY, Wu LC, Wang ZP, Wang ZX, Jia Q, Jiang GS, Zhang WD (2009) Anti-proliferation effect of polypeptide extracted from scorpionn venom on human prostate cancer cells *in vitro*. J Clin Med Res 1(1):24–31

Zheng J, Lee HL, Ham YW, Song HS, Song MJ, Hong JT (2015) Anti-cancer effect of bee venom on colon cancer cell growth by activation of death receptors and inhibition of nuclear factor kappa B. Oncotarget 6(42):44437–44451

Zouari-Kessentini R, Srairi-Abid N, Bazaa A, El Ayeb M, Luis J, Marrakchi N (2013) Antitumoral potential of Tunisian snake venoms secreted phospholipases A2. Biomed Res Int 2013:391389

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