International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 4230-4235

Recognizing Audience Feedback through Facial Expression using Convolutional Neural Networks

M.Kanipriya[1]*, R. Krishnaveni[2], M. Krishnamurthy[3]

[1]Research Scholar, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India.
ORCID: 0000-0002-8033-7898

[2]Professor, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India.

[3]Professor, KCG College of Technology, Chennai, Tamilnadu, India.

Abstract

Viewers opinion is a main element in assessing the usefulness of a forum, discussions, or assembly. On the other hand, challenges exist in acquiring reliable, eminent comments from spectators. Traditional tools such as opinion poll was considered to be unsuccessful while estimating the real sensation of the spectators. Most of the complications related to opinion poll arise due to insufficient number of replies, deferred reactions, and predisposition of the contestant. Aspect detection and emotion recognition algorithms are used to get rid of these misconceptions and then the overall opinion of the spectators can be evaluated. The introduced method can mine features from the realistic audiovisual shots, logged motion pictures, or photographs. Then the system identifies the aspects from the extracted information and investigates sentiments which are being disclosed over non verbal communications. The opinions are then categorized into various sentiment polarities and the system represents the detected opinion information pictorially.

Keywords: Sentiment analysis, Feedback, Face detection, Emotion detection, dataset, Haar features, CNN

I. INTRODUCTION

Audience opinion are greatly required and are significant because it is used to evaluate the overall effectiveness of the research work. The feedback methods which are used to collect spectators view such as surveys do not have high precision. This is because some of the viewers does not attend the survey, late to submit the survey or they are biased. This type of feedback method can be used in seminars, lectures, summits, conferences, meetings or a public reveal of any products. That is, the feedback system works if the video of the audience does exist. Face detection algorithms are used to discover the faces from the video, then with the help of emotion recognition algorithms and dataset, the sentiments of the audience can be known, then classify it and visually signify the data.

Face detection is a sub-field of Object detection. In object-class Detection, the task is to discover the positions and dimensions of all entities in an image that belong to a given class. Face-detection algorithms focus on the discovery of anterior part of human faces. Face detection is used in several places. They are applied in camera for auto-focus, security systems, classifying persons into different groups.

Emotion recognition is the method of recognizing human sentiment from an input text, audio, picture or video. Emotion detection examines the given input, perceives the emotions and categorizes it. The main application of emotion recognition is sentiment analysis. Sentimental analysis is used in recommendation systems, reviewing a product or an item, twitter and to analyze market growth for future prediction.

The sentiment detection analyzes the expressions: angry, fear, Disgust, happy, neutral, sad, surprise. These identified emotions are then classified into three groups:

- Positive(Happy, Surprise)
- Negative(Angry, Fear, Sad, Disgust)
- Neutral

These groups are then plotted on a graph over a time period of the input video. The output graph represents the overall positive, negative opinion of the viewers.

II. RELATED WORKS

The prior works that focused on the aspects are studied that are related to the proposed work such as similar ideas, related techniques used and comparison of those techniques.

T. Shiva et al. [6] suggested a methodology to compute the effect of actions using Emotion detection. The approach is to identify the facial attribute from the audiovisual feed with OpenCV and use emotion recognition technique to investigate the emotion of the person. Tensorflow is used to train the neural network model. The model is used to recognize the emotion and the emotions are visualized into charts.

Kriti dang et al. [1] studied and compared different face detection algorithms. They assessed the algorithms and decided that Viola Jones algorithm is superior than other face detection algorithms on the basis of precision and recall. Viola Jones algorithm has the highest value for precision and recall.

Precision and recall are the measures used to assess the relevance of the algorithm. Precision also known as positive predictive value which states the part of related occurrences retrieved and provides facts about the false alarms. Recall contributes the data about the number of entities spotted inappropriate occurrences retrieved. For the perfect system, the more it is close to 1 more precise the outcomes of the algorithm.

Neha A. Chinchanikar [2] revised various Facial Expression Recognition methodologies proposed in other papers. She also contributed a brief on several widely accessible Facial Expression Recognition Databases. She concluded that Facial Expression Recognition using deep learning yields enhanced outcomes compared to traditional methods like HoG and LBP. CNN with dense connections provides better results. Likewise, Moe Moe Htay et al. [7] surveyed emotion recognition using facial expression and offered understandings on numerous techniques, Datasets and state of art complications.

Heechul Jung et al. [8] developed two deep network models DNN and CNN for facial expression recognition. These models were compared and the CNN had improved performance than DNN in terms of recognition rates. They also performed quantitative and qualitative assessment.

Hongli Zhang et al. [3] recommended a system for Facial Expression Recognition based on convolutional neural network and image edge detection. Haar classifiers are used to detect human faces and Adaboost is used to increase the detection rate. Then CNN is used for facial expression detection. The CNN is a feed-forward neural network, which can extract features from a two-dimensional image and boost up network factors by using a back propagation algorithm. CNN is composed of convolution layer, pooling layer, full connection layer and softmax layer. The novel facial expression recognition method using a CNN model extracts facial features efficiently. Compared to existing methods, the proposed method can automatically acquire pattern features and moderate the incompleteness incurred by simulated design features.

III. METHODOLOGY

This section deliberates about the proposed system, algorithms used and the brief of overall working of the system. The schematic diagram of the proposed system is shown below:

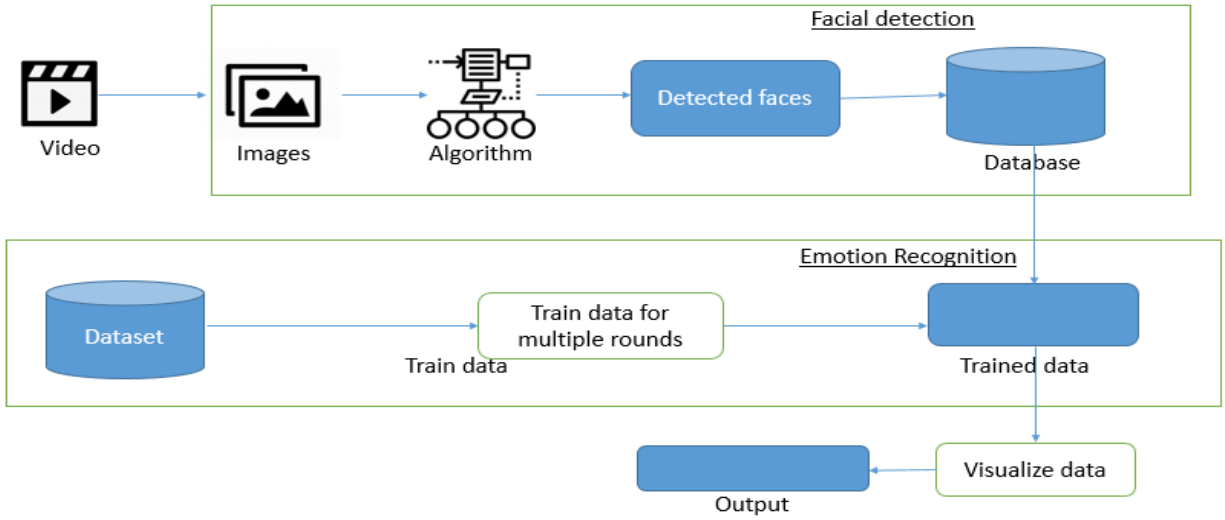

Fig 1: Schematic Diagram of CNN – FER

The complete system is separated into three main fragments:

1. Face detection
2. Emotion detection
3. Data Visualization

A. FACE DETECTION

Facial detection [10] is an artificial intelligence (AI) based computer technology which can be employed to various fields that detects human faces in pictures or videos. Face detection systems can be splitted into two modules based on the techniques they use.

1. Feature based techniques
2. Image based techniques

Here Viola Jones algorithm is used for face detection which uses Haar Feature Selection to identify the aspects.

It is elucidated in the review paper [3], there are four Haar features used in the Viola-Jones algorithm: edge features, linear features, center features and diagonal features and how they are applied in detecting the faces.

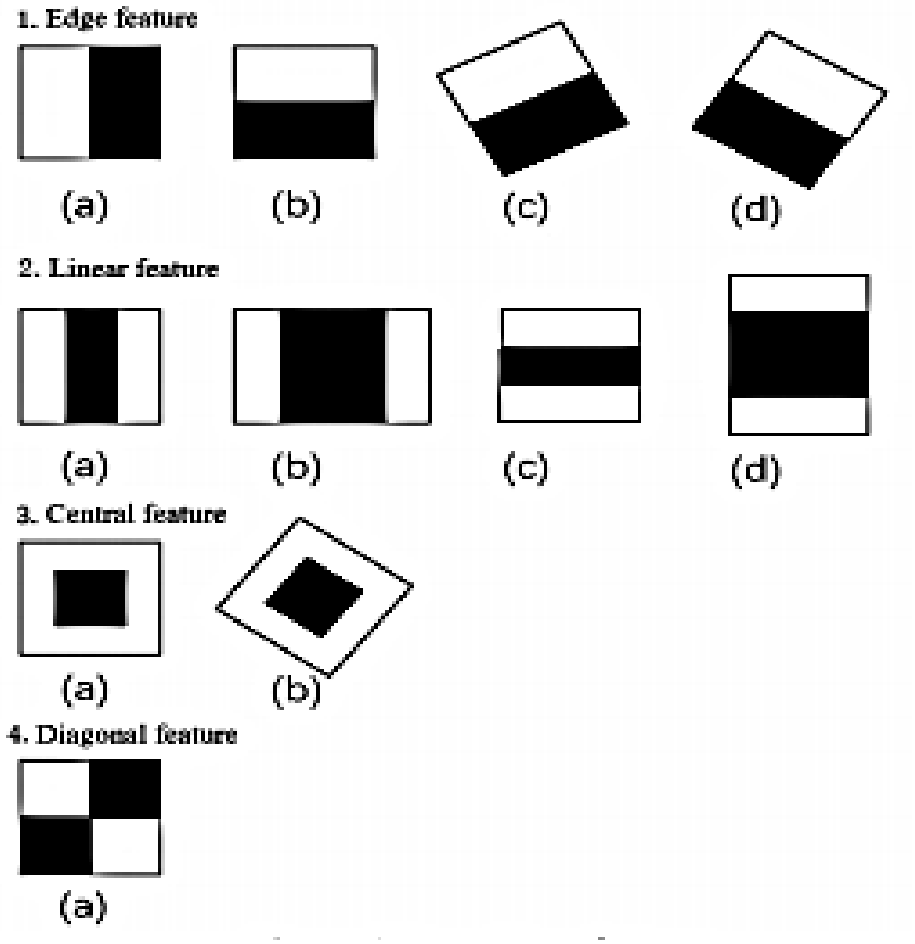

Fig 2: Haar features

The Viola Jones Object Detection framework was considered to be the first framework to provide reasonable entity recognition rates. It can be utilized for discovering entities in practical applications but it is predominantly useful in face detection application. The detection rate of this framework is fairly high (true positive rate) and very low false positive rate which makes the algorithm robust and it also processes the images rapidly. Its main goal is face detection, not recognition, that is to distinguish faces from non-faces, which is the first step of recognition.

Four main steps followed in this algorithm are:

- **Haar Feature Selection:** A Haar-like feature move rectangular window through the image and checks for feature like eyes, nose and mouth to determine the faces in the image
- **Creating an Integral Image:** A fundamental picture is framed by figuring the square shapes present at (x,y) into a solitary picture portrayal which helps in speeding the method.
- **Adaboost Training:** AdaBoost (Adaptive Boosting) is a machine learning meta-algorithm used with other algorithms to boost performance.
- **Cascading Classifiers:** Cascading Classifier is a machine learning based approach where a cascade function is trained from many positive and negative images which is then used to detect objects in other images.

B. EMOTION RECOGNITION

Facial Expression Recognition is a technology that detects the sentiment [4] in the human faces present in a picture or a video. Many applications like human-computer interaction (computer responding / interacting with humans after analyzing what human feels), computer forensics (in the case of lie detection), pain detection, the field of education (i.e. distance learning where teachers determine whether the student understood the course), games and entertainment (for asserting user experience) and so on.

But before finding the emotions, datasets are required. Datasets are a set of labeled images with different emotions. First, the model is trained and then the trained model is used to compute the emotions in the input image or video. There are different datasets available. FER-2013 [9] data set contains 28,000 images that are labeled. The dataset was created in 2013 for learning focused on three challenges: the

black box learning, the facial expression recognition challenges and the multimodal learning challenges. The images are 48 * 48 pixels gray scale of faces in seven expressions: six basic expressions and neutral.

Here FER-2013 is used as a dataset and train it with Convolution Neural Network(CNN) [5].

Fig 3: Sample images in FER-2013 dataset

The principle of deep learning technique is to create a deep neural network like human brain structure, which acquires more advanced feature expression of data level by level through multi-hidden non-linear structure. This mechanism of spontaneously acquiring the core instructions of huge facts create the mined features more crucial depiction of the data and thus the classification results can be significantly improved. For a two-dimensional image input, the neural network model can infer it layer-by-layer from the pixels initially acquired by the computer to edges, parts, contours of objects, objects understood by the human brain and then categorize it directly within the model to achieve better recognition results.

The CNN is a feed-forward neural network, which can extract features from a two-dimensional image and elevate network factors by using a back propagation algorithm. Common CNNs usually consist of three basic layers: a convolution layer, a pooling layer and a connective layer. Each layer is composed of several two-dimensional planes namely feature maps and each feature map has many neurons.

The proposed CNN model includes three convolutional layers namely three pooling layers and two fully connected layers. Input image is a gray scale face image. The weights of CNN are locally linked and pooled spatially. 3×3 filters are used with stride one for each convolutional layer. The model is divided into five Blocks. There are 64 filters exist for the convolution layers in the first block, 128 filters for the convolution layers in the second block, 256 filters for the convolution layers in the third and fourth block and 512 filters for the convolution layers in the fifth block. The stimulation function for each convolutional layer is ReLu.

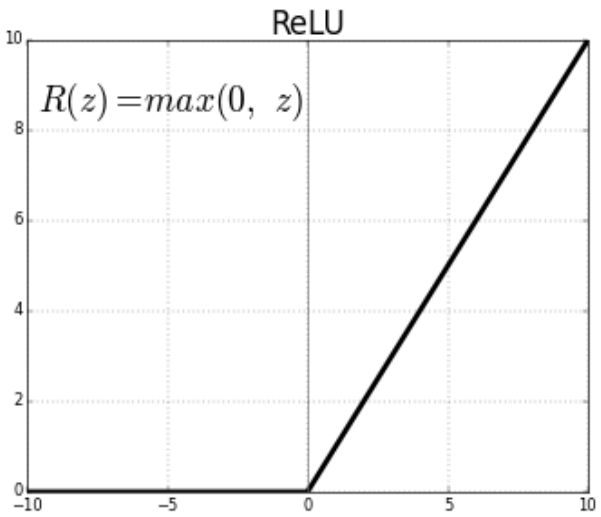

Fig 4: ReLu Activation

Max pooling layers are used for sharing the layers and size of the filter is set to 2×2. As a result, these layers reduce the size of the input image partially. The Softargmax formula is given below:

The CNN has two completely attached layers which have 512 numbers of nodes each. The total count of nodes are determined empirically. The stimulation function of these completely attached layers is also ReLu. Moreover, the model is normalized by including a dropout method at the end of each block. Lastly, the softargmax layer is tracked by two completely attached layers. The total count of nodes in this layer is set to 7, because there are 7 types of sentiments used in our work.

C. DATA VISUALISATION

Data visualization is the graphical illustration of information and facts. By means of pictorial components like charts, graphs and maps, data visualization tools deliver a manageable way to perceive and comprehend trends, outliers and patterns in data.

The recognized emotions are then served into the data visualization module. Here, Matplotlib is used as a visualization tool. By using Matplotlib, the overall opinion of the video can be acquired. The emotions (positive, negative or neutral) are marked on the y-axis resultant to the time marked on the x-axis.

IV. RESULTS

A simple website is hosted on a local server from which the user videos are retrieved. After the input video has been uploaded, the user can view the overall emotion of the audience or review the change in positive, negative or neutral emotions over the time.

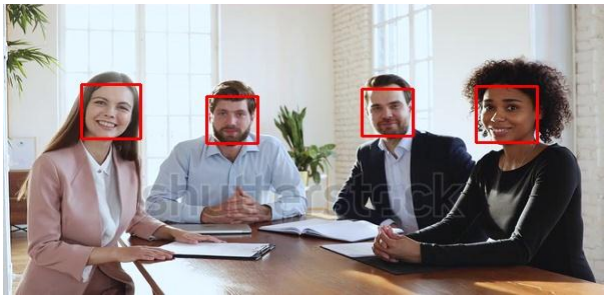

Fig 5: Faces detected by the proposed system

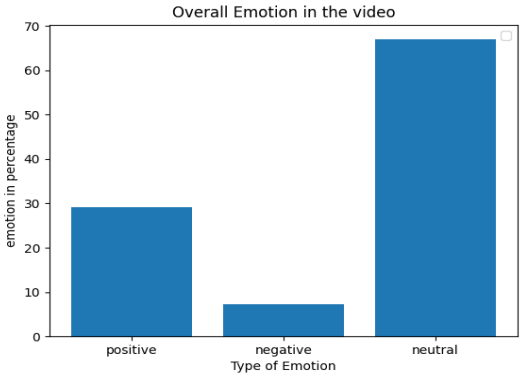

Fig 6: Output showing overall emotion of the audience

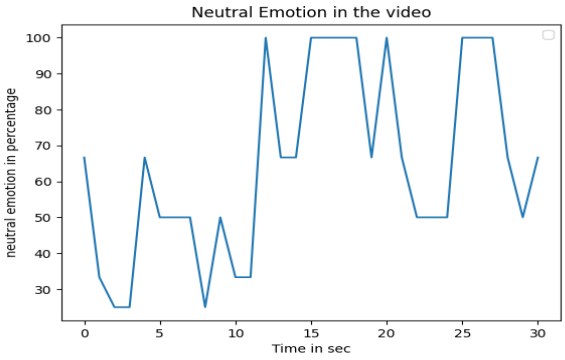

Fig 7: Output showing the neutral emotion over a period of time

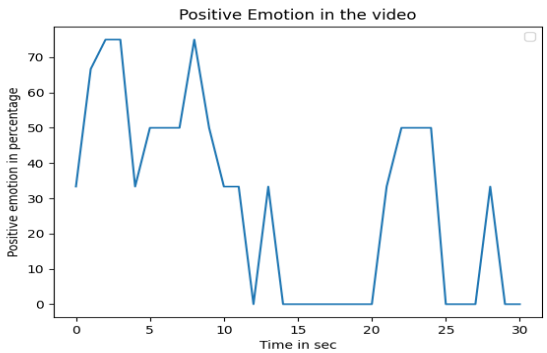

Fig 8: Output showing the positive emotion over a period of time

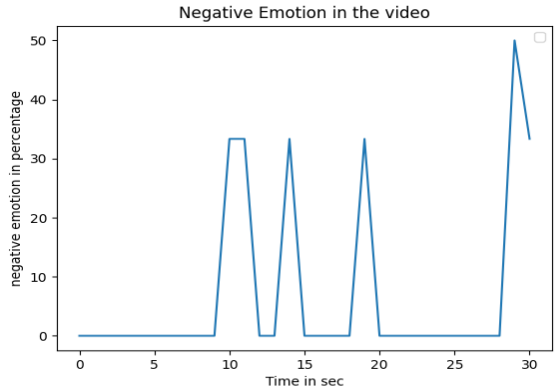

Fig 9: Output showing the negative emotion over a period of time

The Proposed model which has been used for identifying the emotions, is trained for nearly 40 iterations and achieves accuracy of approximately 65%. The accuracy and loss graph of the trained model vs dataset are shown below:

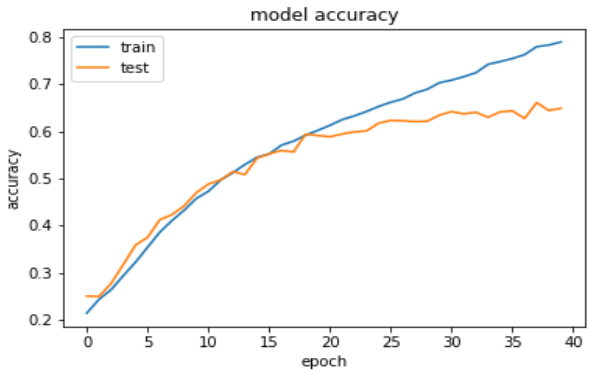

Fig 10: Accuracy graph

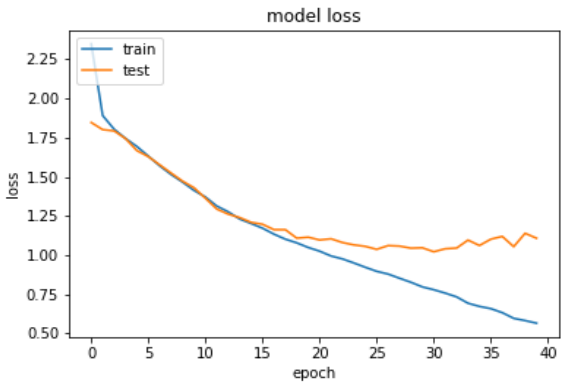

Fig 11: Loss graph

V. CONCLUSION

The proposed system can be an alternative / added method to regular tools like surveys and forms. But the pre-condition is there must be an input video of the audience. So the proposed

system can be used in places like seminars, conferences and meetings. The main challenge of the system is the algorithm detects only full frontal faces. So, the video must contain the audience's clear picture to get the exact emotions.

REFERENCES

[1] Kirti Dang, Shanu Sharma, "Review and Comparison of Face Detection Algorithms". 7th International Conference on Cloud Computing, Data Science & Engineering, 2017

[2] Neha A. Chinchanikar, "Facial Expression Recognition Using Deep Learning: A Review". International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue: 06 2019

[3] Hongli Zhang, Alireza Jolfaei, Mamoun Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing". IEEE Access Volume 7, 2019

[4] Kah Phooi Seng, Li-Minn Ang, "Video Analytics for Customer Emotion and Satisfaction at Contact Centers". IEEE Transactions on Human-Machine Systems, 2017

[5] Yong Li, Jiabei Zeng, Shiguang Shan, Xilin Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism". Journal of Latex Class Files, Volume 14, NO. 8, August 2018

[6] T. Shiva, T. Kavya, N. Abhinash Reddy, Shahana Bano, "Calculating The Impact Of Event Using EmotionDetection". International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 8 Issue-7 May, 2019

[7] Moe Moe Htay, Zin Mar Win, "Survey on Emotion Recognition Using Facial Expression". International Journal of Computer (IJC) Volume 33, No 1, pp 1-10, 2019

[8] Heechul Jung, Sihaeng Lee, Sunjeong Park, Byungju Kim, Junmo Kim, Injae Lee, Chunghyun Ahn, "Development of deep learning-based facial expression recognition system".Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), 2015

[9] Si Miao, Haoyu Xu, Zhenqi Han, Yongxin Zhu, "Recognizing Facial Expressions Using a Shallow Convolutional Neural Network" IEEE Access Volume 7, 2019

[10] Nabil Tahmidul Karim, Sanjana Jain, Jednipat Moonrinta, Matthew N. Dailey, Mongkol Ekpanyapong, "Customer and Target Individual Face Analysis for Retail Analytics" International Workshop on Advanced Image Technology (IWAIT), 2018

AUTHORS PROFILE

Ms. Kanipriya M, received B.E degree in Computer Science and Engineering from Anna University Chennai and M.E. degree in Computer Science and Engineering from Anna University Chennai, India, in 2010 and 2012, respectively. She is now pursuing Ph.D. degree in Information and Communication Engineering at Hindustan Institute of Technology and Science Chennai. She has published research papers in International journals and conferences including reputed journals. Her research interests includes large database mining techniques, incremental data mining techniques, Big Data and Security, Image Processing and Machine Learning and social network analysis and mining.

Dr.Krishnaveni R. is currently working as professor at Hindustan Institute of Technology and Science, Chennai. She is a Ph.D holder from Anna University. She completed her B.E degree from Bharadhiyar University and M.E degree from Madras University. She has published around 30 research papers in International journals and conferences including Springer, IFIP, JCTN, IJNS and many other reputed journals. Her areas of research interests include Biometrics Security, Web Security, Information Security, Networking and Security, Cloud Computing and Security, Big Data and Security, Cyber Security, Image Processing and Machine Learning.

Dr. M Krishnamurthy is currently working as a Professor and Head in the Department of Computer Science and Engineering, KCG College of Technology. He leads a research group of Computer Science and Engineering faculty in KCG. He has published more than 46 papers in refereed international conferences and journals. His primary work and research interests include large database mining techniques, incremental data mining techniques, and social network analysis and mining. He also received prestigious award of "Best Java Architect" and "Distinguish Faculty in Data Mining" from Cognizant Technology Solutions and Venous International Foundation, Chennai in the year 2000 and 2016 respectively. He served as reviewer on several international journals including Knowledge and Information Systems (KAIS) Journal, International Journal of Physical Sciences and China Communications. International Journal of Knowledge and Information Systems. International Journal of Physical Science. Journal of King Saud University (Elsevier Editorial System). International Journal of Advanced Intelligence Paradigms (IJAIP). International Journal of Advanced Intelligence Paradigms (Inderscience) and ACM Computing Science.