Facial Expression Recognition using SVM With CNN and Handcrafted Features

G. Priyanka, S. Pavithra

Abstract*: The facial expression recognition system is playing vital role in many organizations, institutes, shopping malls to know about their stakeholders' need and mind set. It comes under the broad category of computer vision. Facial expression can easily explain the true intention of a person without any kind of conversation. The main objective of this work is to improve the performance of facial expression recognition in the benchmark datasets like CK+, JAFFE. In order to achieve the needed accuracy metrics, the convolution neural network was constructed to extract the facial expression features automatically and combined with the handcrafted features extracted using Histogram of Gradients (HoG) and Local Binary Pattern (LBP) methods. Linear Support Vector Machine (SVM) is built to predict the emotions using the combined features. The proposed method produces promising results as compared to the recent work in [1].This is mainly needed in the working environment, shopping malls and other public places to effectively understand the likeliness of the stakeholders at that moment.

***Keywords*: SVM, CNN, Handcrafted Features, Combined Features, HoG, LBP**

I. INTRODUCTION

Facial expression can make effective communication than words. The lots of information are conveyed through expressions before we perform certain actions. Cultures and languages may differ for people around the world but the emotions and expressions remain the same. Social learning and culture is the main part of facial expression. Some of the facial expressions are anger, disgust, fear, happiness, sadness, surprise, and contempt, embarrassment, interest, pain, and shame. We can easily understand other persons mind through their facial expression.The eyes are typically viewed as vital options of facial expressions. Aspects like blinking rate will probably let others know whether or not someone is anxious or whether or not he or she is lying. Also, eye contact is taken into account a very important side of social communication. However, there are cultural variations relating to the social behavior of maintaining eye contact. Emotions can easily convey the person's psychology state. Some psychologist can even understand the hidden meaning of each facial expression. For emotion classification task, the facial expression has high inequity. In earlier works, CNN is used for extracting temporal and spatial features and provides the effective model for classification [3], [4], [10]. But still it leads to overfitting problem due to the lack of availability of large number of training samples in each emotion class. In order to avoid that, CNN with less number of hidden layers is proposed for feature extraction. But it may drop some high level features in the input image. To avoid that, Appearance based high level features are extracted from the images. For better recognition result the CNN and appearance based approach features are combined. In previous works, random forest, k nearest neighbour classifier, are used for building the classification model. The performance of the proposed system is evaluated on three publicly available datasets. The performance analysis shows that our proposed facial expression recognition systems achieve high accuracy compared to the other existing systems. The rest of this paper organized as follows: Related works is discussed in Section 2, Section 3 detail description about proposed system. Section 4 illustrates the performance of the proposed system and Section 5 is the Conclusion and future work.

II. RELATED WORKS

In criminal interrogating process, medical field and computer aided training we can easily find the emotional state of the personnel like experts using the exemplars. In earlier works, automatic facial expression recognition method is addressed by manually extracting and selecting predominant features. Then these features are applied to classifiers like SVM, Random Forest, etc., to predict the emotions. Feature extraction is the foundation for building the model in more efficient way. Geometric features, Appearance based features and auto learned features are the three major category of facial feature extraction technique. Geometric based approaches discussed in [20], [23], [25] mainly focused on fiduciary points and the distance between each point to describe the different facial emotions. On the other hand appearance based approaches extract entropy information of pixel's intensity from the cropped region or whole recognized face using optical flow [26], HoG [19] and LBP [22]. Auto learned feature extraction methods can extract the features without the human interaction like deep neural networks as mentioned in [7]. After feature extraction, next step is to build the classifier model using the existing or newly introduced learning algorithm. The most important part of the vision system is the representation and not the classification so less intention is put on classification part.

Manuscript published on November 30, 2019.
* Correspondence Author

G. Priyanka*, Assistant Professor (Senior Grade), Department of CSE, Mepco Schlenk Engineering College, Sivakasi – 626005, India.

S. Pavithra, PG Student, Department of CSE, Mepco Schlenk Engineering College, Sivakasi – 626005, India.

Retrieval Number: D7802118419/2019©BEIESP
DOI:10.35940/ijrte.D7802.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

But the performance of the system is fully depends on this classification part. Some of the general learning algorithms used to build the classifier models are k-Nearest-Neighbour (k-NN) [2],[14], Support Vector Machine (SVM) [8],[15],[17], decision tree [9],[12],

Artificial Neural Network(ANN) [27],[29] and exemplar based SVM [1]. The exemplar based SVM model described in [1] provide the answer to the question “Which previously perceived emotion expression is the most similar expression to the given sample?” But it takes lot of time to build N different exemplar and still it lacks in achieving better accuracy for some of the benchmark databases like JAFFE.

The manual feature extraction and selection process is difficult and it is surpassed by the introduction of deep neural networks. Features are extracted and model is built in an unsupervised or semi-supervised manner [3],[4],[10],[11]. A multi-layered neural network system is employed to join low-level features (pixels) with noticeable highlights, and construct the model. These strategies have accomplished genuinely engaging and promising outcomes in the machine vision applications. In the recent work of [1], exemplar based SVM was constructed and it provides answer to the question like ‘which emotion this image associate to?’ but it has low accuracy for JAFFE dataset. And also it takes lot of time to construct N number of Support Vector Machine models with one positive and N-Si number of negative samples, where Si is the same set of positive samples. To avoid it, in the proposed work, multiple number of CNN was constructed to extract the features automatically. CNN is intended to naturally and adaptively learn spatial hierarchies of features through back-propagation by utilizing various structure blocks like convolution layers, pooling layers, and fully connected layers. In this proposed system, to increase the performance of the JAFFE dataset, automatically extracted CNN features are combined with the extracted hand crafted features like LBP and HOG. Finally the SVM model is constructed to classify the features. SVM well suited for high dimensional data and it also classifies unstructured data like image and texts. SVM mainly used in most of the applications because it gives directly classification labels not the probability based on the estimation process.

III. DATASET DESCRIPTION

Our proposed facial expression recognition system’s performance is evaluated on the three publicly available datasets namely Cohen–Kanade+ (CK+) [21], JAFFE [28].

CK+ [21]: CK+ dataset is the extended version of the original CK dataset which is introduced in 2010. It contains facial expression of 593 image sequences of 123 individuals.

JAFFE [28]: This JAFFE dataset contains 10 Japanese females individual of facial expression image sequences. Totally 213 images are available in JAFFE dataset. Each emotion has 30 images on an average. Fig. 1. shows the seven facial expression of an individual person in the JAFEE dataset.

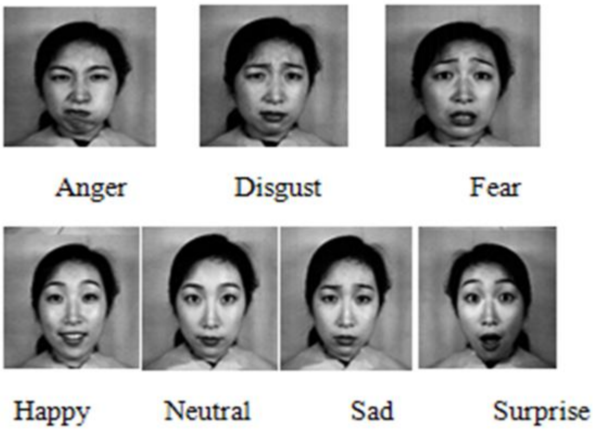

Fig. 1. Seven Facial expression of one individual person (JAFFE Dataset)

IV. PROPOSED SYSTEM

The block diagram for the proposed system is illustrated in Fig. 2. First we collected the facial expression image sequences and then noises are removed in the pre-processing step (see section A). Then features are extracted using CNN and handcrafted methods (see section B), after that CNN and handcrafted features are combined (see section C). Finally SVM is built for classification (see section D)

A. Pre-Processing

The center of the eyes and nose tip of facial expression images are extracted using the algorithm proposed in [46]. Finally the images are resized to 110 x 150 resolutions. No other pre-processing steps are needed.

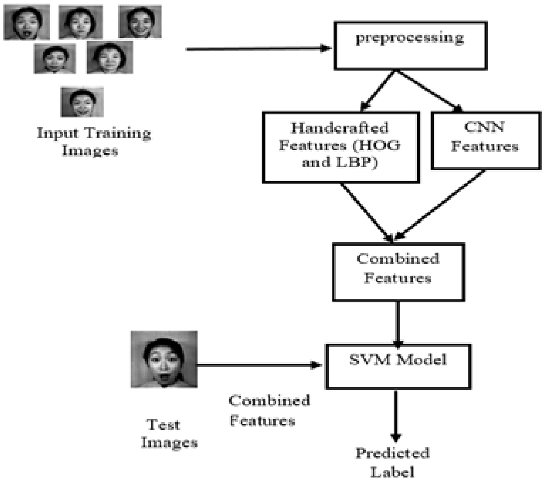

Fig. 2. Block diagram of proposed system

B. Feature Extraction

Features are extracted using two popular techniques. Appearance based feature extraction and automatic feature extraction techniques.

- **Appearance Based Feature Extraction**
  - **Local Binary Pattern**

    The LBP extraction strategy is as smooth as non-parametric technique that describes spatial data of the pixels with respect to their neighbor pixels. This process is terminated by marking a label (decimal value) to every pixel using the equation (1). The histograms are formed using the labels to represent the images.

$$\text{LBP}=\sum_{1}^{n-1} b(c_i - c_g)2^i \qquad (1)$$

where c_i and c_g are the gray-level of the center and neighborhood pixels respectively, and n is the number of neighbors.

- **Histogram of Gradients**

The shape of an object is described using the local gradient method in HOG feature extraction. Then the direction (θ) and the magnitude (g) of each pixels gradients are calculated using the equations (2) and (3) where C_x and C_y defines the horizontal and vertical gradients respectively.

$$g=\sqrt{c_x^2+c_y^2} \qquad (2)$$

$$\theta = \arctan\frac{c_x}{c_y} \qquad (3)$$

The images are divided into small cells (regions) to capture the spatial information of the oriented gradients.

- **Automatic Feature Extraction**

In automatic feature extraction, CNN is mainly used to extract the features without the human interactions. When compared with other feature extraction technique CNN has more efficient feature extraction process. Here three different CNNs are used for extracting the features from three resolution images like 110*110, 90*90, 64*64 resolutions and it is depicted in the Fig. 3. Three different CNN architectures has 6 convolutions and one fully connected and one softmax classification layer at the end. As there is no dependency between one CNN architecture with another, it can be executed in parallel.

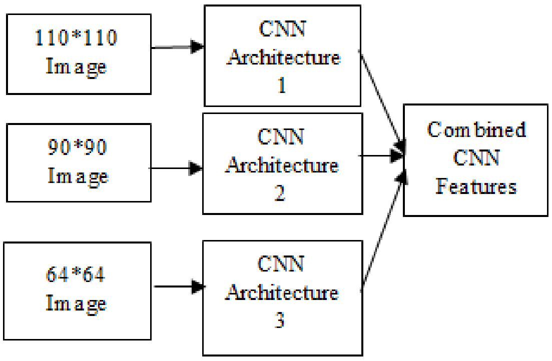

Fig.3. Feature extraction using three different CNN architectures

The three different CNNs along with the layer description and hyper parameter tuning metrics is illustrated in the tables I, II and III. For an image, output from the three CNNs are three row feature vectors. Next step is to convert the three rows into single row feature vector as it belongs to the same input image.

Table -I: CNN Architecture I

Layer	1	2	3	4	5	6
Stage	**Conv**	**Conv**	**Conv**	**Conv**	**conv**	**FC**
Input Size	110x11	54x54	25x25	11x11	9x9	3x3
Filter Size	7x7	7x7	7x7	7x7	3x3	-
Conv Stride	2x2	2x2	2x2	2x2	2x2	-
Pooling	Max	Max	Max	Max	max	-
Pooling Size	2x2	2x2	2x2	2x2	2x2	-
Pooling	2x2	2x2	2x2	2x2	2x2	-
Neurons	256	128	64	128	128	7

Table II: CNN Architecture II

Layer	1	2	3	4	5	6
Stage	**Conv**	**Conv**	**Conv**	**Conv**	**conv**	**FC**
Input Size	90x90	46x46	23x23	12x12	6x6	3x3
Filter Size	2x2	3x3	2x2	3x3	3x3	-
Conv Stride	2x2	2x2	2x2	2x2	2x2	-
Pooling	Max	Max	Max	Max	max	-
Pooling Size	2x2	2x2	2x2	2x2	2x2	-
Pooling	2x2	2x2	2x2	2x2	2x2	-
Neurons	256	128	64	128	128	7

Table III: CNN Architecture III

Layer	1	2	3	4	5	6
Stage	**Conv**	**Conv**	**Conv**	**Conv**	**conv**	**FC**
Input Size	64x64	50x50	25x25	13x13	7x7	4x4
Filter Size	7x7	3x3	3x3	3x3	2x2	-
Conv Stride	2x2	2x2	2x2	2x2	2x2	-
Pooling	Max	Max	Max	Max	-	-
Pooling Size	2x2	2x2	2x2	2x2	-	-
Pooling Stride	2x2	2x2	2x2	2x2	-	-
Neurons	256	128	64	128	128	7

The model trained under three different epochs as 30, 50 and 100 epochs. Batch size is constant for all iteration. Learning rate is fixed as 0.001 and Learn rate drop factor is set as 0.1.

C. Combine Features

The appearance features extracted using HoG and LBP and automatically extracted features using CNN techniques are combined together to yield the better result.

D. Support Vector Machine

After feature extraction the model is built to classify the emotions. Liner SVM Model is used to classify the features extracted by both appearance and automatic techniques. In order to avoid the problem of over-fitting Linear SVM is mainly used to build the classification model. Equation (4) clearly states the inner product between weight vector v_e of sample s_e and bias b_e,

$$\Omega_e(v,b) = v^T s_e + b_e \qquad (4)$$

$v^T s_e + b_e$ = 0 is the distance from hyper plane which is a similar measure of the output Ω in equation (4).

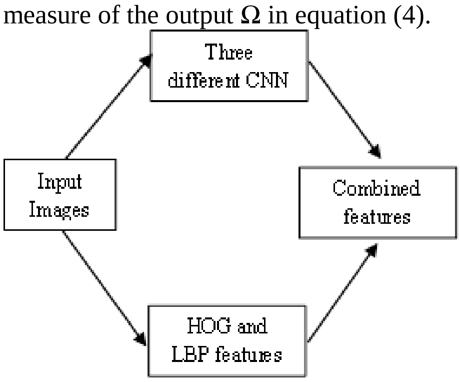

Fig.4. Combined features of appearance and automatic techniques

V. FACIAL EXPRESSION RECOGNITION RESULT

The performance of the facial expression recognition system is evaluated on three dataset CK+, JAFFE. In order to improve the performance of the system, two different types of features are extracted and combined. For each image five feature vectors are extracted. Out of that three feature vectors are extracted using three different CNN architectures and two are extracted using appearance based models. These five feature vectors are combined as a single feature vector for each image. And then it is fed as input to the classifiers. Table IV shows the emotion category-wise accuracy for three different datasets CK+, JAFFE. As CK+ dataset contains only 25 images for fear emotion, misclassification rate is high for fear. Similarly sad emotion has the highest misclassification rate in JAFFE dataset.

Table IV: Emotion wise accuracy for CK+, JAFFE

Emotions	**CK+**	**JAFFE**
Anger	100	100
Disgust	100	100
Fear	92	100
Happiness	100	100
Neutral	100	100
Sadness	100	85
Surprise	100	100
Average	**98.80**	**95.63**

A. Comparison with engineered and learnt based representations

Engineered methods are supervised methods. Most of the traditional classification techniques follow supervised model. In [18] for given facial expression images the Gabor feature are extracted and model was made using SVM. Our model outperforms it with an increase in an average prediction rate of 1% on JAFFE dataset.

Learnt methods are unsupervised method. Mostly Deep neural network follows an unsupervised method. In [11] eight layer neural network was built to find out the face parts and then features are extracted. Finally the classification model was built. Comparing the result of [11] with our proposed system, the proposed system has an average of 6.40% higher prediction rate for CK+ database. In [3] author used a two deep neural network one is used to find out the temporal appearance features and another one is used for temporal geometric feature. Our proposed system approximately provides better accuracy at the rate of 2.90% than the existing system. The below table V and VI clearly explains that our proposed system work well for benchmark datasets.

Table-V: Emotion Recognition rate (%) comparison with engineered approaches

Paper	**Method**	**JAFFE**
[18] (2011)	3D Gabor features + SVM	91.00
[16] (2012)	Contours + Simnet	-
[13] (2014)	Meta Probability Codes + SVM	86.38
[5] (2015)	Curvlet + online sequential learning machine	94.65
[6] (2016)	Local Fisher discriminant analysis	77.02
[3] (2017)	Pyramid + SVM + single-branch decision tree	91.43
[1] (2018)	Exemplar-based + SVM	92.53
ours	CNN + SVM	**95.63**

Table-VI: Emotion Recognition rate (%) comparison with learnt approaches

Paper	**Method**	**CK+**
[10] (2012)	Temporal appearance features + temporal geometry features	98.50
[3] (2014)	Temporal features + spatial features	96.64
[4] (2015)	Linear combination of localized basis function	96.02
[1] (2018)	Exemplar-based + SVM	97
ours	CNN + SVM	**98.80**

VI. CONCLUSION AND FUTURE ENHANCEMENT

The main objective of the proposed system is to improve the emotion recognition accuracy for benchmark datasets like CK+, JAFFE using engineered and learnt approaches. Accuracy mainly depends on the feature extraction techniques used. Here two different types of methods are used for extracting the features. Automatic (CNN based) and appearance based approaches (LBP and HoG) are used to extract the features. Totally five feature vectors are extracted from each image and it is combined together for improving the system's overall performance.In order to improve the classification accuracy the five facial feature vectors are combined into single feature vector. Then support vector machine is used for classification to predict the seven emotion categories: happy, sad, angry, fear, neutral, disgust and anger. Our proposed system works well on both JAFFE and CK+ datasets. But it has more misclassification rate for fear emotion in CK+ dataset because CK+ has 25 original images only. In future, we can use auto-encoders type or Generative Adversarial Neural Network (GAN) to increase the number of images to train the models efficiently thereby reducing the objective loss function for classification.

REFERENCES

1. Nacer Farajzadeh, Mahdi Hashemzadeh "Exemplar-based facial expression recognition", Information Sciences, 460–461, (2018) 318–330.
2. S.K.A. Kamarol, M.H. Jaward, H. Klviinen, J. Parkkinen, R. Parthiban "Joint facial expression recognition and intensity estimation based on weighted votes of image sequences", Pattern Recognition Letters, 92 (2017), pp. 25-32.
3. K. Zhang, Y. Huang, Y. Du "Facial expression recognition based on deep evolutional spatial-temporal networks", IEEE Transactions on Image Processing (2017).
4. E. Sariyanidi, H. Gunes, A. Cavallaro "Learning bases of activity for facial expression recognition", IEEE Trans. Image Processing, 26 (4) (2017), pp. 1965-1978.
5. A. Uar, Y. Demir, C. Gzeli "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering", Neural Computing and Applications, 27 (1) (2016), pp. 131-142.
6. Z. Wang, Q. Ruan, G. An "Facial expression recognition using sparse local fisher discriminant analysis", Neuro computing, 174 (2016), pp. 756-766.
7. J. Schmidhuber "Deep learning in neural networks: an overview", Neural Networks and applications. 61 (2015), pp. 85-117.

8. M.F. Valstar, T. Almaev, J.M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, J.F. Cohn “Fera 2015-second facial expression recognition and analysis challenge”, Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, 6, IEEE (2015), pp. 1-8.
9. A. Dapogny, K. Bailly, S. Dubuisson “Pairwise conditional random forests for facial expression recognition”, Proceedings of the IEEE International Conference on Computer Vision (2015), pp. 3783-3791.
10. H. Jung, S. Lee, J. Yim, S. Park, J. Kim “Joint fine-tuning in deep neural networks for facial expression recognition”, Proceedings of the IEEE International Conference on Computer Vision (2015), pp. 2983-2991.
11. M. Liu, S. Li, S. Shan, R. Wang, X. Chen “Deeply learning deformable facial action parts model for dynamic expression analysis”, Asian Conference on Computer Vision, Springer (2014), pp. 143-157.
12. M.K.A. El Meguid, M.D. Levine “Fully automated recognition of spontaneous facial expressions in videos using random forest classifiers”, IEEE Transactions on Affective Computing, 5 (2) (2014), pp. 141-154.
13. N. Farajzadeh, G. Pan, Z. Wu “Facial expression recognition based on meta probability codes”, Pattern Analysis and Applications., 17 (4) (2014), pp. 763-781.
14. S. Wan, J. Aggarwal “Spontaneous facial expression recognition: a robust metric learning approach”, Pattern Recognition Letters., 47 (5) (2014), pp. 1859-1868.
15. E. Owusu, Y. Zhan, Q.R. Mao” An svm-adaboost facial expression recognition system”, Applied Intelligence., 40 (3) (2014), pp. 536-545.
16. H.-C. Lee, C.-Y. Wu, T.-M. Lin “Facial expression recognition using image processing techniques and neural networks”, Advances in Intelligent Systems and Applications-Volume 2: Proceedings of the International Computer Symposium ICS 2012 Held at Hualien, Taiwan, December 1214, 2012, Springer Science & Business Media (2012), p. 259.
17. M.F. Valstar, M. Mehu, B. Jiang, M. Pantic, K. Scherer “Meta-analysis of the first facial expression recognition challenge”, IEEE Transactions System, Man and Cybernetic. Part B (Cybernetics), 42 (4) (2012), pp. 966-979.
18. L. Zhang, D. Tjondronegoro “Facial expression recognition using facial movement features”, IEEE Transaction on. Affective Computing., 2 (4) (2011), pp. 219-229.
19. O. Dniz, G. Bueno, J. Salido, F. De la Torre “Face recognition using histograms of oriented gradients”, Pattern Recognition Letters., 32 (12) (2011), pp. 1598-1603.
20. M.-Y. Chen, C.-C. Chen “The contribution of the upper and lower face in happy and sad facial expression classification”, Vision Research, 50 (18) (2010), pp. 1814-1823.
21. P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews “The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression”, Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, IEEE (2010), pp. 94-101.
22. C. Shan, S. Gong, P.W. McOwan “Facial expression recognition based on local binary patterns: a comprehensive study”, Image Vision Computing, 27 (6) (2009), pp. 803-816.
23. S.-C. Cheng, M.-Y. Chen, H.-Y. Chang, T.-C. Chou “Semantic-based facial expression recognition using analytical hierarchy process”, Expert System with. Applications. 33 (1) (2007), pp. 86-95.
24. L.-F. Chen, Y.-S. Yen “Taiwanese Facial Expression Image Database”, Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan (2007).
25. A.S.M. Sohail, P. Bhattacharya “Classification of facial expressions using k-nearest neighbor classifier”, International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications, Springer (2007), pp. 555-566.
26. M. Yeasin, B. Bullot, R. Sharma “From facial expression to level of interest: a spatio-temporal approach”, Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2, IEEE (2004).
27. L. Ma, K. Khorasani “Facial expression recognition using constructive feedforward neural networks”, IEEE Transaction System, Man and Cybernation. Part B (Cybernetics), 34 (3) (2004), pp. 1588-1595.
28. M.J. Lyons, J. Budynek, S. Akamatsu “Automatic classification of single facial images”, IEEE Transaction on Pattern Analysis and Machine Intelligence., 21 (12) (1999), pp. 1357-1362.
29. Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu “Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron”, Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, IEEE (1998), pp. 454-459.

AUTHORS PROFILE

G. Priyanka received the Bachelor of Engineering degree in Computer Science and Engineering from Anna University, Tirunelveli, India in 2011 and the Master of Engineering in Computer Science and Engineering from Anna University, Chennai, Tamilnadu, India, in 2013. She is in teaching and research for the past 7 years and 3 years respectively and currently, she is working as Assistant Professor (Senior Grade) in Computer Science and Engineering Department at Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India. Her area of research interests includes Video Analytics, Machine Learning & Deep Learning techniques, and Data Analytics. She has authored or co-authored about 10 publications in International Journal/Conference level. She is a life member of Computer Society of India (CSI) and Indian Society for Technical Education (ISTE).

S. Pavithra completed her B.E Computer Science and Engineering in the year 2017 and she is pursuing M.E Computer Science and Engineering in Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India. In her under graduation course she did project in data mining techniques. In her post-graduation studies she decided to carry out her research project using the popular deep learning techniques for computer vision kind of tasks. She is very keen in programming and her research interest includes Image and Video Processing, data mining, Deep Learning techniques. Her subject of interest includes data mining, machine learning, data base management systems, cloud computing, digital image processing. She is good in programming C, C++, Python, Matlab.