IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 1, March 2023, pp. 79~86
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i1.pp79-86

Human emotion detection and classification using modified Viola-Jones and convolution neural network

Komala Karilingappa[1], Devappa Jayadevappa[2], Shivaprakash Ganganna[3]
[1]Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology, Tumakuru, India
[2]Department of Electronics and Instrumentation Engineering, Jagadguru Sri Shivarathreeshwara Academy of Technical Education, Bengaluru, India
[3]Department of Electronics and Instrumentation Engineering, M.S. Ramaiah Institute of Technology, Bengaluru, India

Article Info

Article history:

Received Oct 20, 2020
Revised May 26, 2022
Accepted Jun 24, 2022

Keywords:

Convolution neural network
Facial emotion recognition
Gray-level co-occurrence matrix
Linear binary pattern
Robust principal components analysis
Viola-Jones

ABSTRACT

Facial expression is a kind of nonverbal communication that conveys information about a person's emotional state. Human emotion detection and recognition remains a major task in computer vision (CV) and artificial intelligence (AI). To recognize and identify the many sorts of emotions, several algorithms are proposed in the literature. In this paper, the modified Viola-Jones method is introduced to provide a robust approach capable of detecting and identifying human feelings such as angerness,sadness, desire, surprise, anxiety, disgust, and neutrality in real-time. This technique captures real-time pictures and then extracts the characteristics of the facial image to identify emotions very accurately. In this method, many feature extraction techniques like gray-level co-occurrence matrix (GLCM), linear binary pattern (LBP) and robust principal components analysis (RPCA) are applied to identify the distinct mood states and they are categorized using a convolution neural network (CNN) classifier. The obtained outcome demonstrates that the proposed method outperforms in terms of determining the rate of emotion recognition as compared to the current human emotion recognition techniques.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Komala Karilingappa
Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology
Tumakuru, Karnataka, India
Email: komalak@ssit.edu.in

1. INTRODUCTION

The human facial expression is one of the most essential and effective component of Inter-personnel communique. Facial expressions are quite costly. There is merely 7% of the total significant in a verbalized part of the message, 38% of the total signal in tone and 55% in portrayed [1]–[3]. Extracted features is very extensively utilized in surveillance, biometric, psychiatric, military and human-computer interaction (HCI) [4]. Facial images are exploited to recognize the type of emotion in humans. Anger, sadness, happiness, surprise, fear, disgust, and neutral are the seven primary emotions. Human facial expressions [5]–[8], can be used to identify the aforementioned states of emotion. Recognizing the human feelings is the important task. Several researchers have worked on the detection of age, sex and feelings from facial features [9]. Detection of different human emotions using facial expressions is a difficult task. The capacity of the system to differentiate between several faces is a frequent requirement in human-computer interaction. Until recently, computer vision issues were extremely difficult. With the advent of technology, the challenges in computer vision (CV) due to changes in lighting, ageing, hair, and other accessories [10], have become uncomplicated. Face recognition software, on the other hand, is used to enhance ease of access by identifying and verifying individuals based on their facial

attributes.Thus understanding facial attributes is vital for CV-based applications. These attributes and expressions helps for classifying the facial emotions. Artificial intelligence (AI) systems are employed on the basis of current technology innovations since these systems are capable of identification of emotion through facial characteristics [11]. Human emotion detection is still an active research area because of the current technology innovations for HCI in deep learning or convolution neural network (CNN) prototypes [12]–[14]. Various techniques are necessary to detect and categorize human faces, but deep learning methodology is better than other methods because of its huge capacities of varied datasets and quick computing capabilities [15]. Typically, the face recognition and classification involves several phases such as pre-processing, detection, feature extraction and classification. A voila-jones (V-J) technique is used for extracting the features by classifying images with emotion. This is usually followed by emotion classification using Haar and CNN [16]–[18]. The representation of extracted facial images with databases is the main shortcoming for the analyzying the features of lips and eye and the 2-D image. To overwhelme this shortcoming, the extracted images can be investigated with region of interest (ROI) [19]. Facial expression recognition (FER) can be done using statistical-unsupervised techniques like Independent component analysis (ICA) and genetic algorithm. Genetic algorithm is a feature enhancing technique that carry out for foreseeing Facial emotions [20]. Around 55% of total facial emotions is verified to contribute for social connections. Some of the limitations of the V-J algorithm includes a lack of accurate face and facial part recognition owing to lighting and variation issues. It also suffers from an inability to recognize a face and facial parts due to a fast shift in scene illumination and being too sensitive to stiff features in pictures. With low-resolution pictures and uneven lighting variations of the images, the updated algorithm V-J recognizes the face and facet part closely [21]. With an extremely low fictional rate and a high real-time video detection rate, it is quite resilient. It was suggested that the eye and mouth features are very important facial features which the algorithm extracts very effectively. When it comes to detecting different human emotions, it's quite accurate.

2. PROPOSED METHODOLOGY

In the proposed work, a distinctive technique is used for FER system using CNN. It consists of 3 important phases; face recognition, feature extraction followed by emotion classification. A video is taken as an input where the images can be extracted from the input video and then pre-process each of the images. The Gabor filter is used for removing the unwanted noise, blur and shadow from the original images. After pre-processing, the face detection is carried out using the modified V-J algorithm. There are four stages present in the modified V-J algorithm namely, Haar feature selection, Integral Image creation, AdaBoost training and cascading classifier. The Haar-feature is useful to apply on input face images to check whether the faces are present or not in an image. It can be computed as result of addition of all image pixels, and then subtracted to obtain a unique value. If the unique value is greater than the range, then it implies the human face is recognized. Creation of Integral Image is used for evaluating the sum-up of pixels in a particular area of interest of an image. Adaboost is used for generating the robust classifiers from feasible classifiers. It is not only used to reduce the detection of false positive rate but also decreases the difficulty due to the presence of redundant features. Cascade structure is not only utilized for removal of the false positive images as well as utilized to inspect the occurrence of a face in a specific part of an image. This is followed by extraction of features from the image by gray-level co-occurrence matrix (GLCM) and linear binary pattern (LBP). Afterwards, the required feature is selected using principle component analysis (PCA). The particular features are fed to the CNN classifier for classification. The output from the CNN classifier is the type of emotion in the image in question.

The most important phase in FER is face detection to identify all emotions efficiently using the modified V-J algorithm. The face and emotion can be detected using the proposed algorithm. Extracting the features plays a importance in the FER system as a result of enhancing the accuracy of Feeling detection techniques. There are many extraction techniques such as LBP, GLCM, gray level weight matrix (GLWM), traditional gabor filter (TGF) and daubechies wavelet packet features (DBWP). In the proposed methodology, extraction of feature techniques such as GLCM and LBP are used for classifying the texture. Using GLCM dissimilarity, correlation, mean, entropy, variance, average angular second moment, homogeneity, contrast, energy, standard deviation and maximum probability features are extracted. LBP is used as texture operator which symbols the image pixels through adopting the process of thresholding the neighborhood of each pixel. The output of LBP is obtained in the form of binary. Due to discernment of power and computational simplicity [22], LBP is a widely used method in real time applications. The popularity of LBP is due to its robustness toward monotonic varies in gray-scale due to light illumination change. In LBP every pixel value p is compared with the radial distance r of its N neighbors. There are N comparisons for each pixel p and the output for each can be expressed as:

$$LBP\ (m,n)\ =\ \sum_{p=0}^{7} s\ x(\text{gc-gp})2^{l} \quad (1)$$

where, ‘gc’ corresponds to the value of grey scale in the centrally located pixel (xc, yc) and ‘gp’ to the grey scale values of the eight

$$s(z) = \begin{cases} 1, z\ \geq\ 0 \\ 0, z\ <\ 0 \end{cases} \quad (2)$$

neighboring pixels. p is the number of neighboring pixels, s(z) is a threshold function. After feature extraction, feature selection is used to upgrade the performance of the classifier. robust principal components analysis (RPCA) technique is employed for extracting the features from face images and also used to reduce the dimensionality of face images. It is a numerical method that tranforms a set of correlated N face images to a set of eigen face images.

The RPCA was formulated as a non-convex optimization problem defined as,

$$arg\ \ \min_{L,S} rank(L) + \lambda||S||_0 \text{ s.t D=L+S} \quad (3)$$

A set of face images are in training, then it is denoted with large eigenvalues through the greatest eigenfaces for accurate estimation of the face. After this step the result of eigenfaces, each face image can be indicated by permutation of eigenfaces, followed by symbolization in the form of vectors. The input features are compared with standard features of dataset for FER. The features are classified using CNN classifier. CNN comprises sequences of convolutional layers, the output which is correlated only to native areas in the input. This is carried-out through sliding filter, or weighted-matrix with respect to the input. For every point, CNN computing the product of convolution between the input and filter [23], [24].

Figure 1 show the block diagram of the proposed FER system. Initially from real time video, facial image will be captured than fed to pre-processing. In the next stage face detection is done using modified V-J method. Facial feature extraction is done using GLCM and LBP. These methods were also used to distinguish the texture information of images and hence it improves the classification performance. The feature selection using RPCA method is done. The RPCA is a feature selection technique which is used for ease the dimensionality of face data. This step is followed by feeding the image to CNN classifier, where the real time image will be compared with database to detect the facial expression more effectively. Figure 2 shows the flowchart of the proposed methodology, which is self explanatory.

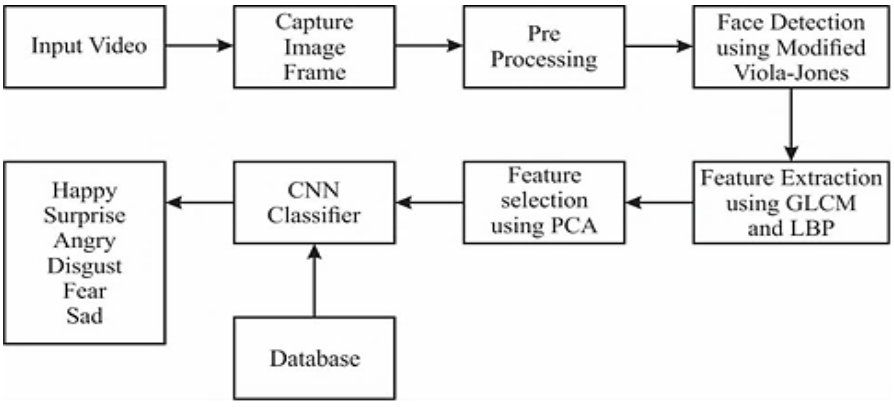


Figure 1. Block diagram of the proposed FER system

3. RESULTS AND DISCUSSION

The proposed work is implemented using the technical computing environment MATLAB. The datasets were collected from Kaggle and karolinska directed emotional faces (KDEF) databases [25]. This dataset comprises 215 images with 7 facial emotions such as happy, sad, surprise, disgust, angry, fear and neutral. The real-time images are used as input images. In the beginning, pre-processing step is used for removal of unwanted images as well it smoothen the images from input datasets using the Gabor filter. For FER, the modified V-J algorithm is used to vary the image intensity and window size. The AdaBoost is not only used to reduce the detection of false positive rate but also decreases the difficulty due to the presence of

redundant features. CNN classifiers are used to classify effectively the different emotion statuses of the input images effectively. The proposed technique yielded an accuracy validation of 95.6%.

The Kaggle and KDEF databases are used as shown in Figure 3, the training and testing sets can be divided through cross-validation. In this validation, the whole database is segregated into three identical sets of images. The segregation is random in nature. Then two sets are combined to use as a training data set. The remaining section of the dataset is used for the testing phase. Figure 4 shows the accuracy and log loss plot of CNN during Training. From figure one can notice the quality of performance of a model as the number of iterations of optimization progresses. Accuracy metric is used to measure the performance in an interpretable way. It is a degree of how accurate the model's likelihood is compared to the correct data. Figure 5 shows Adaboost plot, which gives the relationship between false positive rate and true positive rate. Adaboost is used for adjusting the weights of classifiers during training. The process is repeated as training process iterates. This step ensures that the accuracy of predictions of unusual observation. It is also used to boost the performance of any machine learning algorithm. Figure 6 shows the bargraph of performance of different classifiers for the selected dataset. We can see that CNN has higher performance value among the compared classifiers for the chosen comparision matrix.

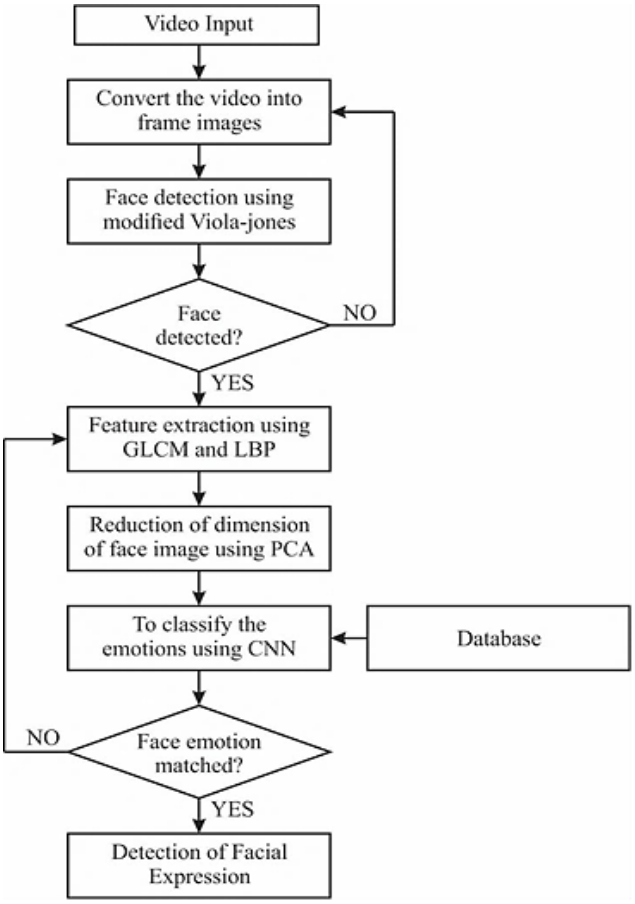


Figure 2. Flowchart of proposed methodology

3.1. Performance analysis

The performance of the proposed work is quantitatively evaluated using the parameters like precision, sensitivity, specificity, accuracy and recall. The confusion matrix of the facial emotion detection is constructed as shown in Table 1 for the merged image. The experimental results show that the proposed technique efficiently detects the facial expressions with high accuracy as compared to the current techniques. Table 2 shows the region of interest and its corresponding real time image. The classification of emotion is displayed above the input image. Table 3 shows the accuracy outcome of CNN classifiers to be more effective in detecting emotions compared to k-nearest neighbor (KNN) and artificial neural network (ANN) classifiers.

Figure 3. Sample dataset used for the proposed work

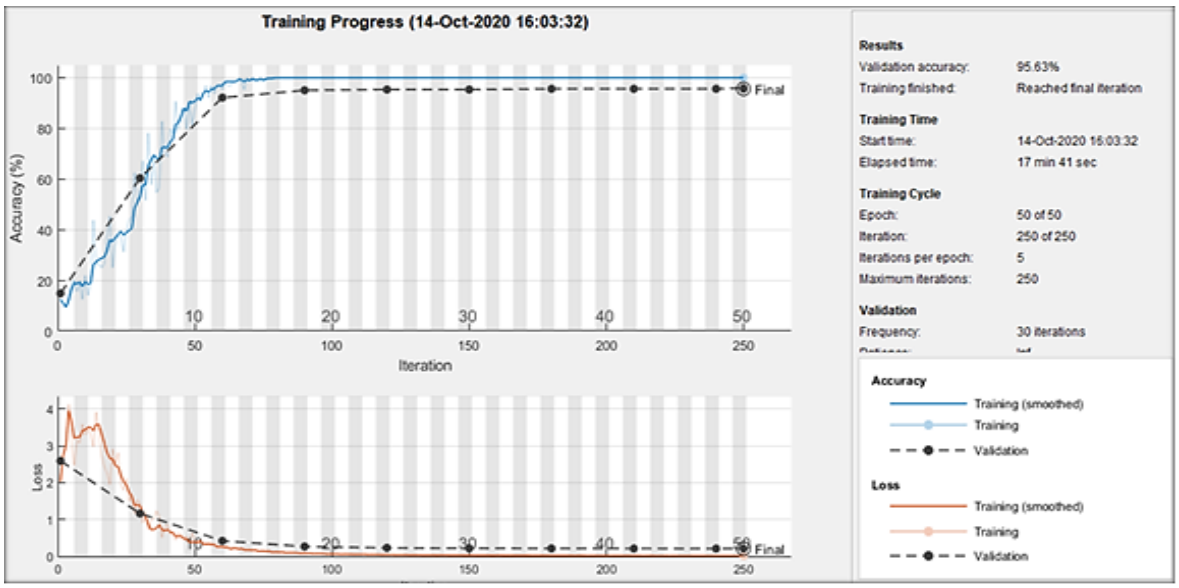


Figure 4. Accuracy and log-loss plot of CNN during training

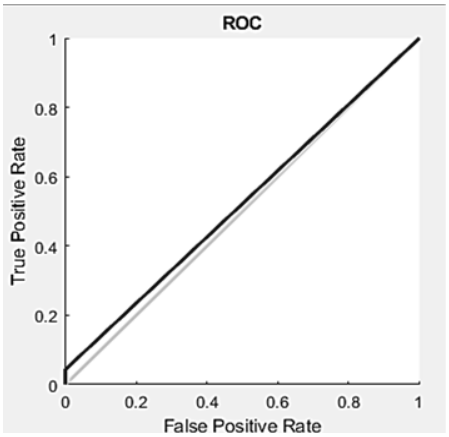


Figure 5. Adaboost plot

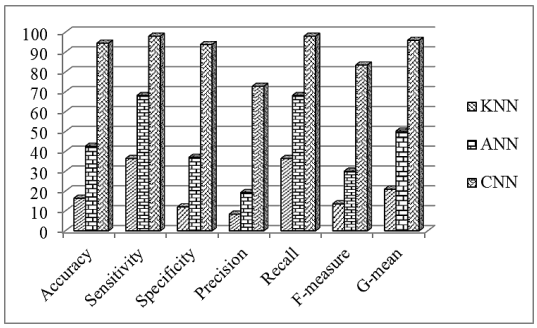


Figure 6. Comparative result of different classifiers

Table 1. Confusion matrix of CNN

		Target class							
		1	2	3	4	5	6	7	
Output class	1	45	0	1	0	0	0	0	97.8%
		13.1%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	2.2%
	2	1	47	0	0	0	0	1	95.9%
		0.3%	13.7%	0.0%	0.0%	0.0%	0.0%	0.3%	4.1%
	3	1	1	47	0	1	0	0	94.0%
		0.3%	0.3%	13.7%	0.0%	0.3%	0.0%	0.0%	6.0%
	4	0	0	0	46	0	1	0	97.9%
		0.0%	0.0%	0.0%	13.4%	0.0%	0.3%	0.0%	2.1%
	5	0	1	0	0	48	0	1	96.0%
		0.0%	0.3%	0.0%	0.0%	14.0%	0.0%	0.3%	4.0%
	6	0	0	1	3	0	48	0	92.3%
		0.0%	0.0%	0.3%	0.9%	0.0%	14.0%	0.0%	7.7%
	7	2	0	0	0	0	0	47	95.9%
		0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	13.7%	4.1%
		91.8%	95.9%	95.9%	93.9%	98.0%	98.0%	95.9%	95.6%
		8.2%	4.1%	4.1%	6.1%	2.0%	2.0%	4.1%	4.4%

Table 2. Real time output results of different classifiers

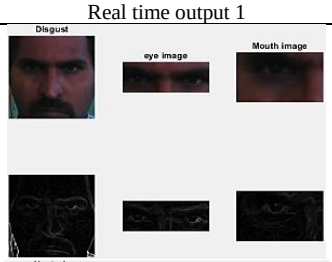
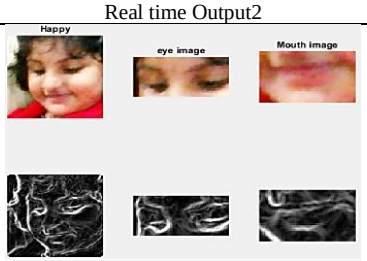
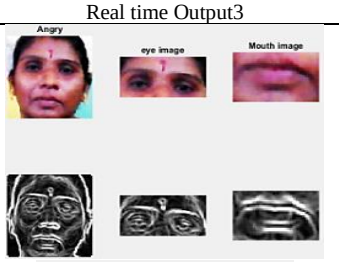
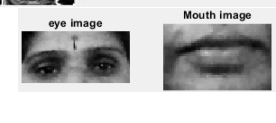
Classifier	Real time output 1	Real time Output2	Real time Output3
ANN			
KNN			
CNN			

Table 3. Performance analysis of different classifiers

Performance metrics (%)	KNN	ANN	CNN
Accuracy	16.39	42.6	94.46
Sensitivity	36.36	68.18	97.96
Specificity	12	37	93.8
Precision	8.33	19.23	72.73
Recall	36.36	68.18	97.96
F-measure	13.56	30	83.48
G-mean	20.89	50.21	95.9

4. CONCLUSION

The performance evaluation of the proposed modified V-J algorithm is carried out with suitable datasets to find the facial emotions from the realtime data-image and also to categorize different emotions. For FER, LPB, GLCM, and RPCA based feature extraction techniques are applied to extract details from face images for reognizing each facial emotion. The entire system is trained and classified using CNN classifiers for FER. The performance of the proposed approach is estimated through the parameters like specificity, sensitivity, precision, recall, and accuracy. The results obtained show that the proposed method efficiently detects emotions in the face images using CNN with an accuracy of 95.33% for different input images.

REFERENCES

[1] T. Chernigovskaya, P. Eismont, and T. Petrova, *Language, music and gesture: informational crossroads*. Springer Singapore, 2021.

[2] S. Mekruksavanich and A. Jitpattanakul, "Biometric user identification based on human activity recognition using wearable sensors: an experiment using deep learning models," *Electronics*, vol. 10, no. 3, p. 308, Jan. 2021, doi: 10.3390/electronics10030308.

[3] A. Swaminathan, A. Vadivel, and M. Arock, "FERCE: facial expression recognition for combined emotions using FERCE algorith," *IETE Journal of Research*, pp. 1–16, May 2020, doi: 10.1080/03772063.2020.1756471.

[4] K. S. Yadav and J. Singha, "Facial expression recognition using modified viola-john's algorithm and KNN classifier," *Multimedia Tools and Applications*, vol. 79, no. 19–20, pp. 13089–13107, May 2020, doi: 10.1007/s11042-019-08443-x.

[5] A. Jaiswal, A. K. Raju, and S. Deb, "Facial emotion detection using deep learning," in *2020 International Conference for Emerging Technology (INCET)*, Jun. 2020, pp. 1–5, doi: 10.1109/incet49848.2020.9154121.

[6] S. K. Mondal, I. Mukhopadhyay, and S. Dutta, "Review and comparison of face detection techniques," in *Proceedings of International Ethical Hacking Conference 2019*, Springer Singapore, 2019, pp. 3–14.

[7] R. Goel, I. Mehmood, and H. Ugail, "A study of deep learning-based face recognition models for sibling identification," *Sensors*, vol. 21, no. 15, p. 5068, Jul. 2021, doi: 10.3390/s21155068.

[8] V. Sreenivas, V. Namdeo, and E. V. Kumar, "Group based emotion recognition from video sequence with hybrid optimization based recurrent fuzzy neural network," *Journal of Big Data*, vol. 7, no. 56, Aug. 2020, doi: 10.1186/s40537-020-00326-5.

[9] L. B. Krithika and G. G. L. Priya, "Graph based feature extraction and hybrid classification approach for facial expression recognition," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2131–2147, Jul. 2020, doi: 10.1007/s12652-020-02311-5.

[10] K. D. Ismael and S. Irina, "Face recognition using viola-jones depending on python," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1513–1521, Dec. 2020, doi: 10.11591/ijeecs.v20.i3.pp1513-1521.

[11] B. Taha and D. Hatzinakos, "Emotion recognition from 2D facial expressions," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, May 2019, pp. 1–4, doi: 10.1109/ccece.2019.8861751.

[12] M. Li, X. Yu, K. H. Ryu, S. Lee, and N. Theera-Umpon, "Face recognition technology development with Gabor, PCA and SVM methodology under illumination normalization condition," *Cluster Computing*, vol. 21, no. 1, pp. 1117–1126, Mar. 2017, doi: 10.1007/s10586-017-0806-7.

[13] M. Nehru and S. Padmavathi, "Illumination invariant face detection using viola jones algorithm," in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Jan. 2017, pp. 1–4, doi: 10.1109/icaccs.2017.8014571.

[14] K. Dang and S. Sharma, "Review and comparison of face detection algorithms," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Jan. 2017, pp. 629–633, doi: 10.1109/confluence.2017.7943228.

[15] L. Shen, H. Wang, L. Da Xu, X. Ma, S. Chaudhry, and W. He, "Identity management based on PCA and SVM," *Information Systems Frontiers*, vol. 18, no. 4, pp. 711–716, Apr. 2015, doi: 10.1007/s10796-015-9551-8.

[16] A. Borovykh, S. Bohte, and C. W. Oosterlee, "Conditional time series forecasting with convolutional neural networks," *arXiv preprint*, 2017, doi: 10.48550/arXiv.1703.04691.

[17] V. K. Gudipati, O. R. Barman, M. Gaffoor, Harshagandha, and A. Abuzneid, "Efficient facial expression recognition using adaboost and haar cascade classifiers," in *2016 Annual Connecticut Conference on Industrial Electronics, Technology & Automation (CT-IETA)*, Oct. 2016, pp. 1–4, doi: 10.1109/ct-ieta.2016.7868250.

[18] D. Dagar, A. Hudait, H. K. Tripathy, and M. N. Das, "Automatic emotion detection model from facial expression," in *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, May 2016, pp. 77–85, doi: 10.1109/icaccct.2016.7831605.

[19] A. Garg and R. Bajaj, "Facial expression recognition & classification using hybridization of ICA, GA, and neural network for human-computer interaction," *Journal of Network Communications and Emerging Technologies (JNCET)*, vol. 2, no. 1, pp. 49–57, 2015.

[20] N. Mahajan and H. Mahajan, "Emotion detection algorithm," *International Journal of Electrical and Electronics Research*, vol. 2, no. 2, pp. 56–60, 2014.

[21] P. Yaffe, "The 7% rule: fact, fiction, or misunderstanding," *Ubiquity*, vol. 2011, no. October, pp. 1–5, Oct. 2011, doi: 10.1145/2043155.2043156.

[22] K. Nozaki, H. Ishibuchi, and H. Tanaka, "Adaptive fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 238–250, 1996, doi: 10.1109/91.531768.

[23] I. Paliy, "Face detection using Haar-like features cascade and convolutional neural network," in *international conference on" modern problems of radio engineering, telecommunications and computer science"(TCSET)*, 2008, pp. 375–377.

[24] G. UKharat and S. V. D. Ul, "Emotion recognition from facial expression using neural networks," in *2008 Conference on Human System Interactions*, May 2008, pp. 422–427, doi: 10.1109/hsi.2008.4581476.

[25] T. Abidin and W. Perrizo, "SMART-TV: a fast and scalable nearest neighbor based classifier for data mining," in *Proceedings of the 2006 ACM symposium on Applied computing - SAC '06*, 2006, pp. 536–540, doi: 10.1145/1141277.1141403.

BIOGRAPHIES OF AUTHORS

Mrs. Komala Karilingappa obtained B.E in Electronics and Communication Engineering and M.Tech. in Digital Electronics and Communication from Visvesvaraya Technological University, Belagavi in the year 2000 and 2010 respectively. Currently she is working as Assistant Professor, Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India. She has more than 17 years of teaching experience and she published 8 papers in National and Intrernational Journals and conferences, her area of research interest is image processing. She can be contacted at email: komalak@ssit.edu.in.

Devappa Jayadevappa received BE Degree in Instrumentation Technology from Siddaganga Institute of Technology, Tumkur, M.Tech. Degree from SJCE, Mysore specialization in Biomedical Instrumentation and Ph.D. from Jawaharlal Nehru Technological University, Andrapradesh. He is currently working as a Professor, Department of Electronics and Instrumentation Engineering, Jagadguru Sri Shivarathreeshwara Academy of Technical Education, Bengaluru. He has more than 22 years of teaching and industrial experience. He has published more than 100 papers in International and National Journals and Conferences. He is the reviewer for various National and International Journals published across the world. His areas of interests are Digital image processing, Medical imaging, Biomedical Signal Processing and Industrial Automation. He can be contacted at email: djayadevappa@jssateb.ac.in.

Shivaprakash Ganganna received B.E. degree in Instrumentation Technology from SIT, Tumakuru, Karnataka, India, in 1991. He has received a M.Tech. in Bio-medical instrumentation from SJCE, Mysuru, India in 1995. He has received a Ph.D. from Visveswaraya Technological University, Belagaum. He is currently working as an Associate Professor with the Department of Electronics and Instrumentation Engineering, Ramaiah Institute of Technology, Bangalore, Karnataka, India. His research area includes VLSI design, image processing, signal processing, and industrial automation. He can be contacted at email: shivaprakash@msrit.edu.