Research Article

CLBPGNN Convoluted Local Binary Pattern based Grouping Neural Network for Face Emotion Recognition

Lakshmi Priya G. G.* and Krithika L. B.

Scholl of Information Technology and Engineering, VIR University, Vellore, India

___

Abstract

Due to the variations of facial expressions, the process of emotion recognition is a highly demanding and challenging task in many image processing applications. So, the traditional works introduced different techniques for facial image recognitions, which includes the drawbacks of reduced accuracy and recognition rate. Thus, this work aims to develop a new emotion recognition system by implementing different image processing techniques. Initially, the image obtained from the video sequence is preprocessed by employing the Neighborhood Pixel Intensity based Filtering (NPIF) technique. Then, the Viola Jones algorithm is implemented to extract the face by suppressing the background and foreground regions, which increases the accuracy of recognition. After that, the Convoluted Local Binary Pattern (CLBP) technique is employed to extract the patterns of the background suppressed image. Based on the extracted key points, the emotion of the given image is recognized by using the Grouping based Neural Network (GNN) classification technique.

Keywords: Emotion Recognition, Neighborhood Pixel Intensity based Filtering (NPIF), Viola Jones Algorithm, Convoluted Local Binary Pattern (CLBP), Grouping Neural Network based Classification.

___

1. Introduction

Emotion recognition is one of the interest research areas in current days, because the facial emotion of human is used to understand the person's mood in a particular time. The automatic facial emotion recognition system is highly useful in day-to-day applications [1,2]. For instance, detecting the user's mood in a Human Computer Interaction (HCI), distrustful person identification in railway stations, airports and other public places [3] is a highly critical task. Processing the images that obtained from the videos has been used in different applications such as surveillance, action recognition, tracking, face detection and emotion recognition [4,5]. Due to the emergence of advantaged technologies, these applications are integrated with some electronic devices for sharing the mood and perception of people [6]. So, the analysis of people's mood detection is a main role in video processing. Typically, the expression of the facial image is identified by using the key parts of eyes, nose and mouth. The facial expression recognition contains three stages: face detection, feature extraction and expression classification.

a) Problem Identification

Typically, the emotional reactions of human beings is complicated at some times, so it is very difficult to predict the emotions due to the mixture of reactions [7, 8]. It includes anger, fear, happy, sadness, contempt, disgust, and surprise. Moreover, detecting emotions from the videos is also a critical task, because the video processing considers many factors that includes quality, and resolution [9, 10]. In the traditional works, various emotion prediction techniques are developed, but it has the key issues of blurred region separation inaccurate foreground and background suppression [11]. Also, it lacks with the facial key point orientation based problems. The extraction of face region is hard for the persons to understand the extracted feature space and user views [12]. To solve these problems, this paper aims to develop an efficient facial emotion recognition system by employing different image processing techniques.

b) Objectives

The major objectives of this paper are as follows: To remove the noise and to increase its quality, a Neighborhood Pixel Intensity based Filtering (NPIF) is used. To suppress the foreground and background regions for detecting the face from the filtered image, a Viola Jones algorithm is used. To extract the features for further processing, the Convoluted Local Binary Pattern (CLBP) technique is utilized. To accurately classify the emotion of the given image, a Grouping Neural Network (GNN) based classification technique is introduced.

c) Organization

The rest of the sections organized in the paper are as follows: various image processing techniques such as pre-processing face detection, feature extraction, and classification used for emotion recognition are surveyed in Section II. Then, the description about the proposed methodology is presented with the clear flow diagram in Section III. The experimental results of the proposed emotion recognition system are analyzed, and its effectiveness is proved by comparing it with the existing techniques in Section IV. Finally, the paper

___

is concluded and the enhancements that will be implemented in future are stated in Section V.

2. Related Works

In this section, the traditional works related to facial emotion recognition are investigated with its benefits and limitations.

Happy, et al [3] introduced a new framework for recognizing the facial expression of the human based on the selected facial patches. This work includes the stages of preprocessing, face detection, facial patches extraction, and classification. Here, the low pass filtering was utilized to eliminate the noise from the image for face localization. Then, the Region of Interest (ROI) is extracted for the points of eyes and nose based on the geometrical positions of the face. Also, the binary image was obtained by performing the morphological dilation operation. However, this paper required to reduce the computational cost and execution time of the recognition process. *Teijeiro-Mosquera, et al* [13] implemented a Computer Expression Recognition Toolbox (CERT) for recognizing the facial expressions of the human. Here, the personality impression was predicted from the face by extracting the four types of cues. Then, the relationship between each facial expression was analyzed by using big five traits of the image. Based on the duration and relative location, the prediction capability was calculated with better annotation impressions. But, this work failed to segment the speech and non-speech of a verbal content. *Zheng* [14] employed a Group Sparse Reduced Rank Regression (GSRRR) mechanism for describing the facial feature vectors. The Linear Discriminant Analysis (LDA) was utilized to reduce the dimensionality of the feature vector. The processes that involved in this work were as follows: facial region division, feature extraction, dimensionality reduction, and classification.

The advantage that observed from this work was, it has an increased recognition performance. *Kamarol, et al* [15] suggested the Spatio Temporal Texture Map (STTM) technique for extracting the better features for facial expression recognition. Here, the viola jones algorithm was used to eliminate the unwanted boundary from the given image for efficiently recognizing the face image. The emotions that recognized in this work were as follows: anger, disgust, happiness, sadness, surprise, and fear. The merit of this paper was, it reduced the computational cost with better accuracy. But, it required reducing the computational complexity by implementing the suitable machine learning classification mechanism. *Uddin, et al* [16] developed a Deep Belief Network (DBF) for recognizing the facial expression by extracting the patterns of the image. For this purpose, a Local Directional Position Pattern (LDPP) technique was implemented, which considered the strength positions of each pixel. Also, the Principle Component Analysis (PCA) and Generalized Discriminant Analysis (GDA) were utilized to clearly illustrate the expressions of the image. The disadvantage of this work was, it has an increased time and cost consumption. *Kabir, et al* [17] designed a new feature description, namely, Patterns of Oriented Motion Flow (POMF) for recognizing the facial expression of the image from the video. The Hidden Markov Model (HMM) was also used to produce objective sequences of the codebook. Here, the directional motion information was obtained by using an enhanced local texture descriptor. However, this paper failed to prove the effectiveness of the suggested approach.

Wang, et al [18] used a Nature Visible and Infrared Facial Expression (NVIE) construction procedure for analyzing the significant differences between the posed and spontaneous facial expressions. The authors stated that the facial image sequences were highly informative compared than the apex images for emotion recognition. *Mostafa, et al* [19] designed a Random Forest Classifier (RFC) for detecting the facial expression from the given video sequences. In this work, the relevancy of various facial patches was extracted for recognizing different emotions. Here, six universal expressions were recognized based on the position and size of the active regions. Also, some facial points were accurately detected with reduced computational cost by the use of facial landmark detection model. *Zhang, et al* [20] suggested a Deep Neural Network (DNN) based classification technique for analyzing the facial expressions. In this recognition system, a Scale Invariant Feature Transform (SIFT) technique was utilized to extract the meaningful features of the face image. Based on the structure of low level input feature, the projection layer and convolutional layer were included in this system for reducing the space complexity. The advantage of this work was, it has better recognition accuracy.

Yin, et al [21] introduced a Multiple fusion layer based Ensemble Classifier of Stacked Auto Encoder (MESAE) for recognizing the emotions of the facial image. Here, the over fitting was avoided by estimating the artificial feature vector. Also, the uncertainty of the empirical evaluation was avoided by using the structural learning framework. Moreover, the visualization and quantitative evaluation were performed to analyze the feasibility of the recognition system. However, it required to improve the generalization capability of the classifier, which was the major limitation of this work. *Kaya, et al* [22] employed deep CNN for recognizing the facial expressions from the video sequence. Here, the importance of flexible registration and fine tuning were analyzed for obtaining the features. Moreover, the false detections were automatically eliminated by employing a Principle Component Analysis (PCA) method. Also, different visual descriptors were utilized to split the video into equal length volumes of the image. The drawback of this paper was, it failed to prove the efficacy of the suggested recognition system by evaluating different performance measures. *Katsimerou, et al* [23] developed an automatic mood prediction model for recognizing the emotions of the facial image. This work includes two stages such as mood prediction and emotion recognition. Here, the authors stated that more complex models were required to estimate the mood prediction.

Elaiwat, et al [24] suggested a Restricted Boltzmann Machine (RBM) for recognizing the facial expressions based on its features. Here, the relationship between the image pairs was estimated by using different transformations. Moreover, a quadripartite Contrastive Divergence (CD) algorithm is employed to describe the joint distribution between the input sets. In this model, the relationships between the image pairs were estimated based on the independent hidden sets. Also, the most suitable features were extracted by the use of deep complex network structure. The drawback that observed from this work was, it required to improve the computational efficiency of the recognition system. *Guo, et al* [25] implemented a Dynamic Facial Expression Recognition (DFER) model for analyzing the facial expressions of the given image. The motive of this work is as follows: a Diffeomorphic growth model was utilized to extract the feature movements of the facial image.

A sparse group wise registration model was implemented to determine the facial features of the entire image.

Here, the corresponding query sequence was compared with the atlas sequence for determining the expression stage of recognition. Also, the Longitude Facial Expression Atlases (LFEA) technique was considered in this work for comparing the results of the suggested technique. However, it failed to increase the robustness of the suggested method. *Li, et al* [26] developed a multi-kernel learning model for recognizing the emotions from the given video sequence. Here, four different types of features were included such as: hybrid features, multimodal features, two-level deep features, and SVM based decision optimization. Here, the multiclass decision rules were optimized by the distribution of predicted class. The advantage of this work was, it has an increased accuracy by extracting the features of the image. But, it lacks with the drawbacks of increased computational and time complexity. *Patwardhan and Knapp* [27] suggested a rule based decision model for recognizing the emotions with the use of multimodal system. This paper focused the following things: A human annotator system was developed to extract the knowledge based on the rules. The rule base decision was considered for recognizing the emotions. However, it required to increase the level of accuracy during emotion recognition. *Sanchez-Mendoza, et al* [28] designed a Facial Action Coding System (FACS) for detecting the emotions from the facial image based on the geometric and appearance features. Here, the dimensionality of the data was reduced by performing the feature extraction. Also, the histograms of action units were utilized to classify the set of positive and negative attitudes. However, this paper failed to prove the efficiency of the recognition system by employing various performance measures. In this survey, the benefits and demerits of different image processing techniques used for emotion recognition were surveyed. From this, it is analyzed that the existing techniques have the following drawbacks: reduced accuracy, computationally inexpensive, minimized robustness and misclassification results.

To solve these issues, the proposed work aims to introduce an efficient face recognition system by using an enhanced preprocessed, feature extraction and classification techniques.

3. Proposed Method

In this section, the clear description about the proposed facial emotion recognition system is presented. The motive of this paper is to accurately detect the emotions of the human by implementing an improved preprocessing, feature extraction and classification techniques. Initially, the given video sequence is split into frames, then it is preprocessed by eliminating the noise, increasing the smoothness, and visual quality. For this purpose, a Neighborhood Pixel Intensity based Filtering (NPIF) is developed, which efficiently removes the noise. Then, the background separation and foreground extraction processes are performed by using the viola jones algorithm. After that, different key points such as left eye, right eye, nose and mouth are extracted, and its patterns are also extracted by using the Convoluted Local Binary Pattern (CLBP) technique. Consequently, the Grouping Neural Network (GNN) is utilized to recognize the emotion of the face based on the extracted patterns. The flow of the proposed work is depicted in Fig 1 (a), which includes the following stages: preprocessing, background separation

and foreground extraction, differential key point extraction, pattern extraction, and classification.
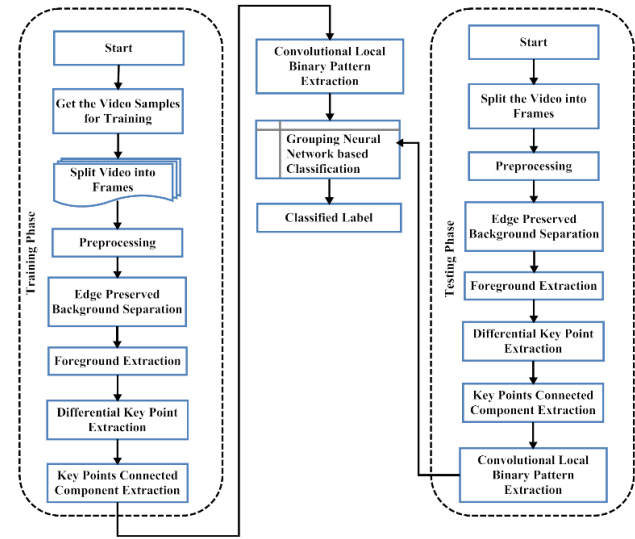

Fig 1 (a). Flow of the proposed face emotion recognition system

a) Preprocessing

Preprocessing is the beginning stage of many image processing applications, which efficiently removes the noise from the given image. Generally, a face occupies a small part of image, in order to extract this part, it must be preprocessed. Here, the frame that obtained from the video sequence is given as the input for filtering, in which the size of image is extracted by estimating the row and column. After getting the input, the row and column size of the image is extracted at first, and the *3 × 3* window is projected on the image for variable temp, which is shown in Fig 1 (b).
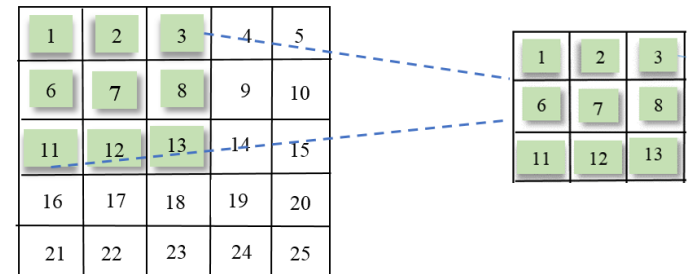

Fig 1 (b). *3 × 3* window matrix

The mean value is estimated and applied on temp, then the resultant is stored in M_{temp}. After that, the difference between the window pixels and the mean value of the window pixels is calculated, which is represented in *3 × 3* in matrix. During this process, the pixels that are all greater than the value of two is validated, if the condition is satisfied, the median of temp is assigned to M_{P_w}. Then, the original pixel values are replaced by the M_{P_w} pixel values. The original and preprocessed images by using NPIF are depicted in Fig 2.

Algorithm I – Neighborhood Pixel Intensity based Filtering

Input: Raw Image;

Output: Filtered Image;

Step 1: Get the image as input;

Step 2: Estimate the row and column size of the image;

Step 3: for i = 2 to r

Step 4:   for j = 2 to c //Where, r –> row size and c –>
*   column size of the image;*

Step 5:   *Project 3×3 window on the image for the variable P_w;*

Step 6:   *Apply mean value to temp and store the result in M_{temp};*

Step 7:   *$P_d = temp - M_{temp}$; //Where, P_d indicates the difference between the window pixels and the mean value of window pixels, It is represented in 3×3 matrix;*

Step 8:   *If the difference is greater than the value of two, the flag value can be incremented;*

Step 9:   *if $N_{P_d} > 2$, then*

$M_{P_w} = median (P_w)$; //Where, N_{P_d} defines the total number of pixels that are greater than two in P_d;

Step 10:   *Replace the original image pixel values by M_{P_w} pixel values;*

Step 11:   *End if;*

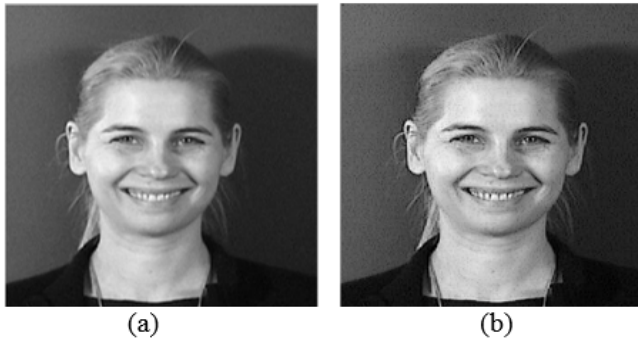*Step 12:*   *End for loop;*

Step 11Get the output;

Fig 2 (a). Original image and (b). Preprocessed image

After preprocessing the image, the background and foreground areas are separated by using the viola jones algorithm. Here, the foreground suppression is mainly performed to detect the facial image by suppressing the background points. The detected face region, and the background suppressed images are depicted in Fig 3 (a) and Fig 3 (b).
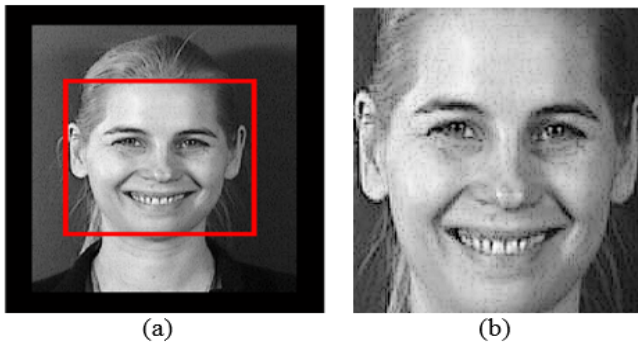

Fig 3 (a). Face detection and (b). Background suppressed image

b) Feature Extraction

After detecting the facial image, the patterns are extracted with the use of Convoluted Local Binary Pattern (CLBP) technique. Extracting the useful features points is important for representing the characteristics of human face. It converts the image pixel into a higher level representation of color, texture, shape, motion and spatial configurations. In this algorithm, the extracted key points such as eyes, nose, and mount are given as the input. Then, the row and column size

of the image are estimated by projecting the 3×3 window on the input image. Further, the double median of $P_{w_{CLBP}}$ is calculated and the value is stored in $M_{P_{wCLBP}}$. If the input image pixel is greater than the value of $M_{P_{wCLBP}}$, the temporary variable T_1 is assigned as 1; otherwise, the T_1 is assigned as 0. Here, T_1 is mainly used for constructing the pattern and size of T, which may vary for each iteration. Consequently, the initial pattern is estimated and the kernel function is calculated for sharpening the image through replacing the pixels of P_i by T_2. In this technique, the key points include nose, eyes, and mouth are separately extracted, which are depicted in Fig 4. The detailed procedure of the CLBP based feature extraction is explained as follows:

Algorithm II – Convoluted Local Binary Pattern Extraction

Input: Isolated eye, nose and mouth from the input image using viola jones;

Output: Convoluted LBP pattern;

Step 1:   Get the input;

Step 2:   Estimate the row and column size of the image;

Step 3:   for i = 2 to r

Step 4:   for j = 2 to c //Where, r is the row size and c is the column size;

Step 5:   Project 3×3 window on the input and assigned as $P_{w_{CLBP}}$;

Step 6: $M_{P_{wCLBP}} = median (median (P_{w_{CLBP}}))$

Step 7:   if input image pixel $> M_{P_{wCLBP}}$

Step 8:   Assign $T_1 = 1$;

Step 9:   else

Step 10: $T_1 = 0$; //T_1 is the temporary variable used for constructing the pattern and size of T may vary for each iterations;

Step 11:   Estimate the pattern P_i

$$P_i = T_1(i-1, j-1) * 2^7 + T_1(i-1,j) * 2^6 + T_1(i-1, j+1) * 2^5 + T_1(i, j+1) * 2^4 + T_1(i+1, j+1), 2^3 + T_1(i+1,j) * 2^2 + T_1(i+1, j-1) * 2^1 + T_1(i, j-1) * 2^0;$$

Step 12:   end for;

Step 13:   end for;

Step 14:   for k = 2 to m − 3

Step 15:   for l = 2 to n- 2 //Where, m and n represents the row and column of P_i;

Step 16:   Project the 3×3 window on P_i;

Step 17: $T_2 = P_i * K$; //Where, K is the kernel function that is used for sharpening the image and T_2 is the temporary variable;

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$ //Kernel function is used to sharpen the image;

Step 18:   After getting the value of T_2, the value of P_i is replaced by T_2, which is assigned in as a label of P_{final}.
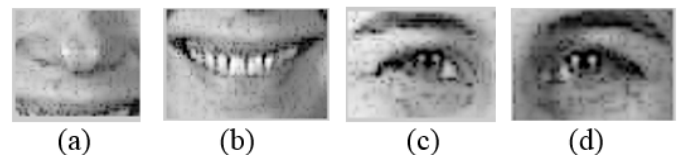
Step 19:   end for;

Step 20:   end for;

Fig 4 (a). Nose detection (b). Mouth detection (c).Left eye detection and (d). Right eye detection

c) Grouping Neural Network Based Classification

After extracting the patterns, the Grouping Neural Network (GNN) based classification technique is employed to classify the emotion of the given image. The proposed classification technique is developed based on the traditional Neural Network (NN). It contains three layers such as input layer, hidden layer, and output layer. In which, the input layer takes the input features, and the output layer can produce the result as a classified label. In the hidden layer, the classification process is performed based on the input features obtained from the input layer. In this technique, the training and testing features obtained the feature extraction stage are given as the input. After that, the unique value of the features are calculated and stored in the variable T_c, then the length of T_c is estimated. Consequently, the row and column size of the training and testing data are validated. The clusters are formed based on the training features, then its mean value is estimated and stored in P_{r1}. The weight value of the edges are also calculated, based on this the mean of the training data is estimated. Finally, the difference between Pr_{F1} and Pr_{test} is calculated and stored in Pr_{final}, then the minimum of Pr_{final} is produced as an output label. The clear procedure of the proposed GNN based classification is illustrated as follows:

Algorithm III – Novel Grouping Neural Network based Classification

---

Input: *Training features, testing features and labels;*
Output: *Classified output;*
Step 1: *Get the type of classes that is denoted as T_c;*
Step 2: *Check whether the row size of the training data, and label it as equal or not;*
Step 4: *Check whether the column size of training data and testing data are equal or not;*
Step 5: *for i = 1 to N // N represents the number of classes (i.e. Smile, Disgust, Confused, etc);*
   for j = 1 to length of label
        if $T_c(i) = label\ (j)$;
                $G_i = training\ feature\ (j,:)$
                //Where, G_i – group
        end if;
     end for;
   end for;
Step 6: *for m = 1 to C_{G_i}//Where, C_G - Column size of the group;*
Step 7: *$P_{r1} = mean(G_i(:,1))$//Where, $P_{r1}, P_{r2}, P_{r3} \dots P_{rN}$*
Step 8: *end for;*
Step 9: *$W = mean(T_c)$ //Estimating the weight value for edges;*
Step 10: *$Pr_{w1} = W * Pr_1$ //$Pr_{w1}, Pr_{w2}, Pr_{w3} \dots Pr_{wN}$;*
Step11: *$Pr_{test} = W * mean\ (testing\ data)$*
Step 12: *$Pr_{F1} = mean\ (Pr_{w1})$; //$Pr_{F1}, Pr_{F2}, Pr_{F3} \dots Pr_{FN}$*
Step 13: *$Pr_{final} = Pr_{F1} \sim Pr_{test}$;*
Step 14: *Classified output = min (Pr_{final})*

---

4. Performance Analysis

In this section, the simulation results of the existing and proposed emotion recognition techniques are experimented by employing different measures.

4.1 Dataset Description

For evaluating the performance, the MMI dataset [29] is used. The MMI dataset is mainly developed for solving the issue of automatic human behavior analysis which contains a large number of visual data with facial expressions of different persons. Also, this dataset contains 2900 videos with high resolution images of 75 subjects.

The proposed work is implemented in Matlab 8.1 (R2013a) with system configuration of Intel core i5-4200U of RAM 8GB with speed 2.30 GHz.

4.2 Performance Measures

The performance of the proposed work is evaluated using sensitivity, specificity, accuracy, precision and recall. The measures of sensitivity and specificity are widely used in many imaging applications for analyzing the betterment of the technique. These measures are estimated as follows:

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

$$Specificity = \frac{TN}{TN+FP} \tag{2}$$

Where, TP represents the true positive, FP represents the false positive, TN indicates the true negative and TP indicates the true positive. Based on these, the accuracy of the recognition system can be calculated, which states that how accurately the proposed technique recognize the emotions. It is calculated as follows:

$$Accuracy = \frac{TN+TP}{(TN+TP+FN+FP)} \tag{3}$$

The precision and recall are essential measures, which are used for evaluating the correctness of the emotion recognition, in which the recall is same as the sensitivity. The precision is estimated as follows:

$$Precision = \frac{TP}{(TP+FP)} \tag{4}$$

4.3 Experimental results of proposed work

Fig 5 shows the sensitivity, specificity and accuracy of the proposed CLBP-GNN technique. In this analysis, the increased sensitivity, specificity, accuracy value obtained is 93.33%, 96.96% and 93.33% respectively. From the results, it is observed that the proposed CLBP-GNN provides better results since GNN performs classification by finding the weight value with its corresponding mean for classifying the labels. Also, it integrates the benefits of the CLBP, which extracts the features for classification. Thus, the sensitivity, specificity and accuracy of the proposed CLBP-GNN are increased efficiently.

4.4 Comparative Analysis

The traditional techniques [25] such as Dynamic Facial Expression Recognition (DFER), Longitude Facial Expression Atlases (LFEA) are considered for proving the superiority of the proposed CLBP-GNN technique.

The average recognition is defined as the accurate recognition rate of the classification technique. Here, the recognition rate is estimated for both existing and proposed facial emotion recognition techniques, which are illustrated from Figure 7 to 11. Moreover, this measure is estimated with respect to varying emotions of anger, disgust, fear, happiness, and sadness, where the number of folds represents the training set. When compared to the traditional DFER and LFEA techniques, the proposed CLBP-GNN technique has the increased recognition rate since it preserves the edge information by efficiently discriminating the expressions of

the face. Moreover, it suppresses the shape variations for an accurate recognition.
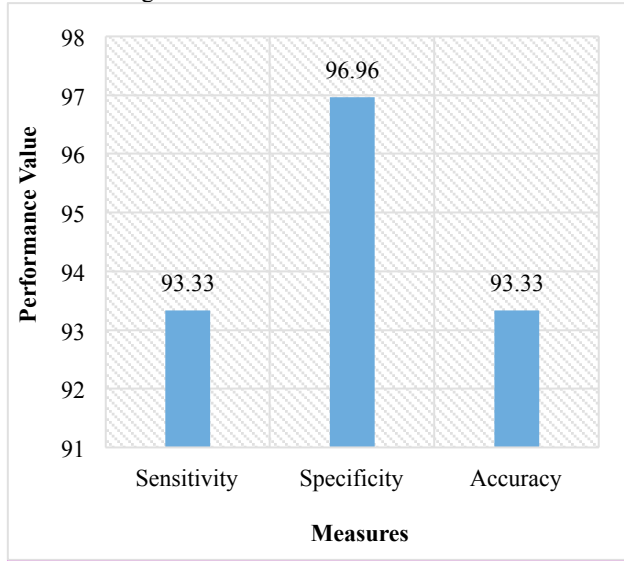

Fig 5. Sensitivity, specificity, and accuracy

Fig 6 shows the precision and recall values of the existing and proposed technique, where CLBP-GNN has the improved precision and recall.
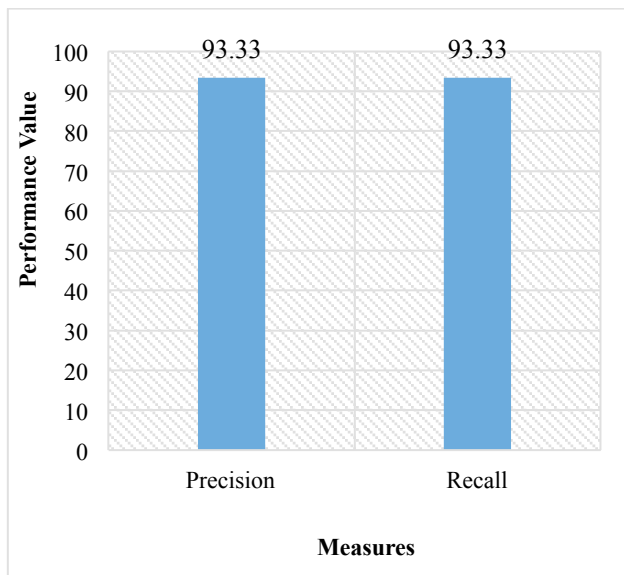

Fig 6. Precision and recall

By using the GNN technique, the recognition rate for an anger emotion is increased from 80% to 96.4%, for disgust the rate is increased from 75% to 98%, for fear the rate is increased from 84% to 97%, for happiness the rate is increased from 84% to 99% and for sadness the rate is increased from 75% to 95%. Because, the proposed GNN technique recognize the emotions based on the patterns of the image. Also, it has the ability to detect the emotions with increased quality and resolution. Thus, the proposed CLBP-GNN provides the increased recognition rate for all kinds of emotion.
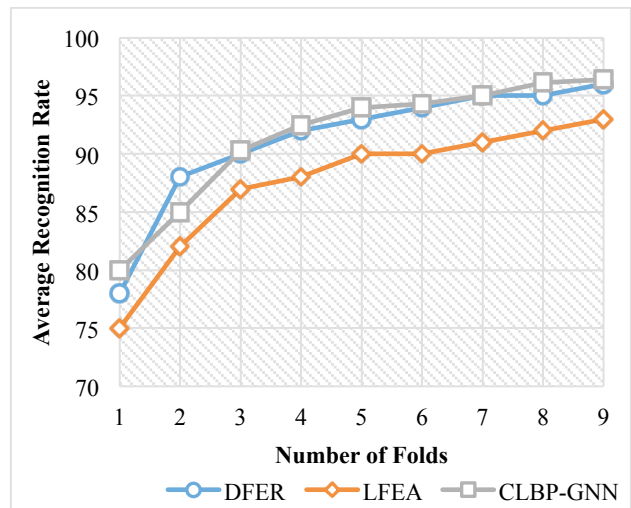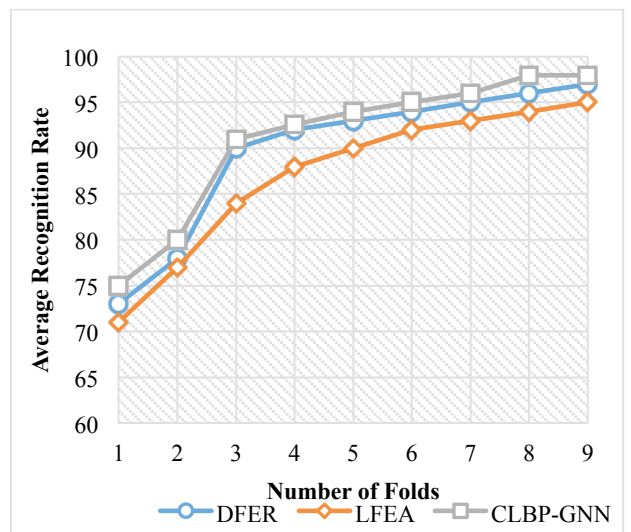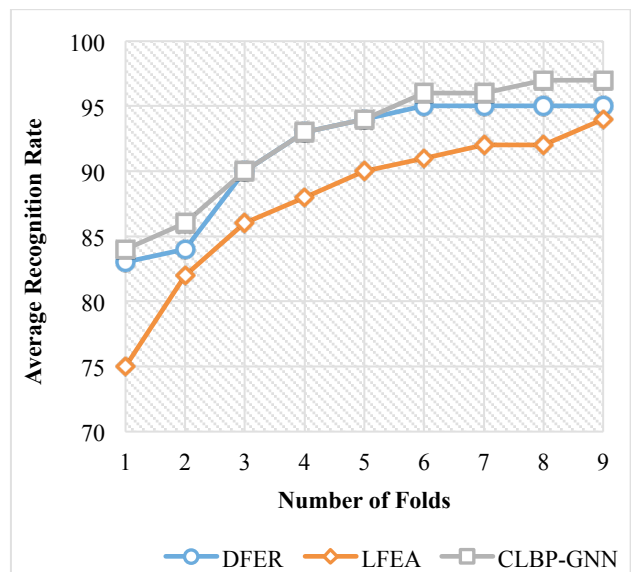

Fig 7. Anger
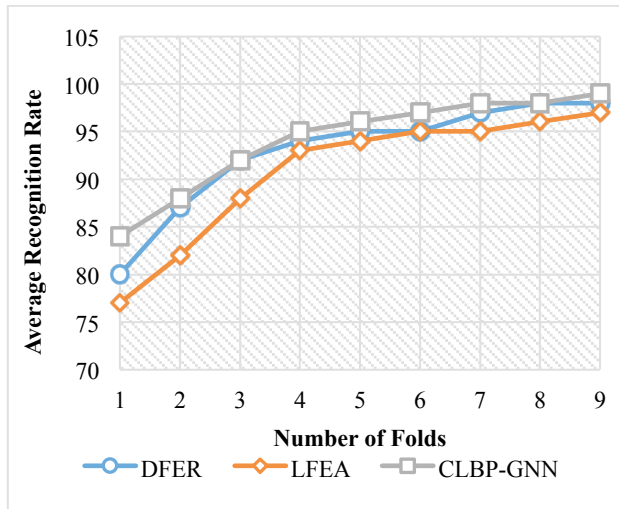

Fig 8. Disgust

Fig 9. Fear
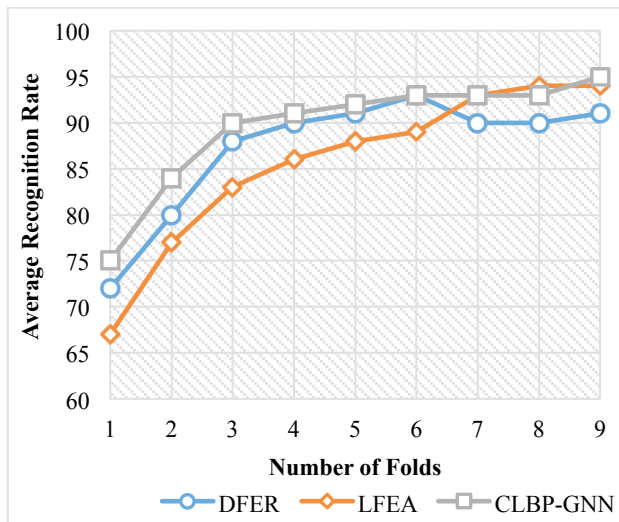
Fig 10. Happiness

Fig 11. Sadness

5. Conclusion and Future Work

This paper proposed a new recognition system, namely, CLBP-GNN for accurately identifying the emotion of the human face. Here, an efficient image processing techniques are implemented in preprocessing, feature extraction, and classification stages. Initially, the NPIF algorithm is employed to eliminate the noise and to increase the quality of the input image frame. To separate the face region from the preprocessed image, the background and foreground regions are extracted by the use of viola jones algorithm. After that, the CLBP technique is utilized to extract the patterns of the face region. Based on the obtained patterns, the key points that include left eye, right eye, nose and mouth are separately identified. Then, the training and testing features of the obtained key points are given as the input of GNN classification, where the emotion of the human face is produced as the classified label output. The major benefits of this work are increased accuracy, reduced misclassification rate, and improved performance.

During experiments, the results of the CLBP-GNN are evaluated by employing different performance measures. Moreover, its efficacy is analyzed by comparing it with the traditional classification techniques. The overall results state that the proposed CLBP-GNN outperforms the other techniques by efficiently extracting the patterns of key points.

In the future work, this work can be improved by reducing the computational complexity of the emotion recognition process. Also, the feature selection can be implemented for improving the accuracy of recognition.

References

1. Yu Z, Zhang C, editors. Image based static facial expression recognition with multiple deep network learning. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (2015).
2. Rahulamathavan Y, Phan RC-W, Chambers JA, Parish DJ. Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. IEEE Transactions on Affective Computing;4(1):83-92,(2013).
3. Happy S, Routray A. Automatic facial expression recognition using features of salient facial patches. IEEE Transactions on Affective Computing,6(1):1-12,(2015).
4. Owusu E, Zhan Y, Mao QR. A neural-AdaBoost based facial expression recognition system. Expert Systems with Applications.;41(7):3383-90,(2014).
5. Mollahosseini A, Chan D, Mahoor MH, editors. Going deeper in facial expression recognition using deep neural networks. Applications of Computer Vision (WACV), 2016 IEEE Winter Conference (2016).
6. Zavaschi TH, Britto AS, Oliveira LE, Koerich AL. Fusion of feature sets and classifiers for facial expression recognition. Expert Systems with Applications;40(2):646-55, (2013).
7. Beaudry O, Roy-Charland A, Perron M, Cormier I, Tapp R. Featural processing in recognition of emotional facial expressions. Cognition & emotion;28(3):416-32, (2014).
8. Xu Y, Li X, Yang J, Lai Z, Zhang D. Integrating conventional and inverse representation for face recognition. IEEE transactions on cybernetics;44(10):1738-46, (2014).
9. Sikka K, Dykstra K, Sathyanarayana S, Littlewort G, Bartlett M, editors. Multiple kernel learning for emotion recognition in the wild. Proceedings of the 15th ACM on International conference on multimodal interaction,(2013).
10. Kahou SE, Bouthillier X, Lamblin P, Gulcehre C, Michalski V, Konda K, et al. Emonets: Multimodal deep learning approaches for emotion recognition in video. Journal on Multimodal User Interfaces,0(2):99-111, (2016).
11. Liu M, Li S, Shan S, Chen X. Au-inspired deep networks for facial expression feature learning. Neurocomputing;159:126-36, (2015).
12. Li Y, Wang S, Zhao Y, Ji Q. Simultaneous facial feature tracking and facial expression recognition. IEEE Transactions on Image Processing;22(7):2559-73, (2013).
13. Teijeiro-Mosquera L, Biel J-I, Alba-Castro JL, Gatica-Perez D. What your face vlogs about: expressions of emotion and big-five traits impressions in YouTube. IEEE Transactions on Affective Computing;6(2):193-205, (2015).
14. Zheng W. Multi-view facial expression recognition based on group sparse reduced-rank regression. IEEE Transactions on Affective Computing;5(1):71-85, (2014).
15. Kamarol SKA, Jaward MH, Parkkinen J, Parthiban R. Spatiotemporal feature extraction for facial expression recognition. IET Image Processing;10(7):534-41,(2016).
16. Uddin MZ, Hassan MM, Almogren A, Alamri A, Alrubaian M, Fortino G. Facial Expression Recognition Utilizing Local Direction-Based Robust Features and Deep Belief Network. IEEE Access;5:4525-36, (2017).

17. Kabir MH, Salekin MS, Uddin MZ, Abdullah-Al-Wadud M. Facial Expression Recognition from Depth Video with Patterns of Oriented Motion Flow. IEEE Access. (2017).

18. Wang S, Liu Z, Wang Z, Wu G, Shen P, He S, et al. Analyses of a multimodal spontaneous facial expression database. IEEE Transactions on Affective Computing;4(1):34-46,(2013).

19. El Meguid MKA, Levine MD. Fully automated recognition of spontaneous facial expressions in videos using random forest classifiers. IEEE Transactions on Affective Computing;5(2):141-54, (2014).

20. Zhang T, Zheng W, Cui Z, Zong Y, Yan J, Yan K. A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition. IEEE Transactions on Multimedia;18(12):2528-36, (2016).

21. Yin Z, Zhao M, Wang Y, Yang J, Zhang J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Computer Methods and Programs in Biomedicine;140:93-110, (2017).

22. Kaya H, Gürpınar F, Salah AA. Video-based emotion recognition in the wild using deep transfer learning and score fusion. Image and Vision Computing,(2017).

23. Katsimerou C, Heynderickx I, Redi JA. Predicting mood from punctual emotion annotations on videos. IEEE Transactions on Affective Computing;6(2):179-92, (2015).

24. Elaiwat S, Bennamoun M, Boussaid F. A spatio-temporal RBM-based model for facial expression recognition. Pattern Recognition;49:152-61,(2016).

25. Guo Y, Zhao G, Pietikäinen M. Dynamic facial expression recognition with atlas construction and sparse representation. IEEE Transactions on Image Processing;25(5):1977-92, (2016).

26. Li W, Abtahi F, Zhu Z, editors. A deep feature based multi-kernel learning approach for video emotion recognition. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction,(2015).

27. Patwardhan A, Knapp G. Augmenting Supervised Emotion Recognition with Rule-Based Decision Model. arXiv preprint arXiv:160702660,(2016).

28. Sanchez-Mendoza D, Masip D, Lapedriza A. Emotion recognition from mid-level features. Pattern Recognition Letters;67:66-74, (2015).

29. MMI Facial Expression Database 2016. Available from: https://mmifacedb.eu/,(2016).