APPLICATION OF SOFT COMPUTING

Emotion recognition system for E-learning environment based on facial expressions

Farzana Begum[1] · Arambam Neelima[1] · J. Arul Valan[1]

Abstract

The popularity of E-learning has grown significantly due to the continuous growth of Internet usage and related technologies. However, the lack of face-to-face interaction in E-Learning poses a challenge in detecting the emotions of the learners. Although existing emotion recognition systems can identify the six universal emotions, they are unable to recognize the emotions specific to the E-learning environment, such as Confusion, Boredom, Concentration, and Self-Confidence. To tackle this challenge, a new emotion recognition system is proposed that considers multiple portions of the face. The proposed method utilizes the Viola–Jones algorithm for face detection, local binary patterns (LBP) for extracting the local facial features, and fuzzy neural network for classification. The experimental results demonstrate that the proposed system performs better than existing methods, achieving a higher prediction accuracy of 91.25%.

Keywords E-learning · Emotion recognition · Facial expressions · Local binary patterns (LBP)

1 Introduction

In today's world due to the enormous increase in the personal computer usage, interaction between human and machine is also widely spreading. The traditional way of learning in classrooms involves face-to-face interaction where the teachers were able to detect the emotions of the students according to their facial expressions. E-learning plays a very vital role in today's education system where recognizing emotions are found to be difficult than the traditional learning. Thus, recognition of emotions in the E-learning environment has been in focus of the researchers since last decades. To encourage the modern education system and to improve the quality of teaching in E-learning environment, recognizing of emotions from facial expressions comes into picture.

✉ Farzana Begum
  farzanab247@gmail.com

  Arambam Neelima
  neelimaarambam@yahoo.co.in

  J. Arul Valan
  valanmspt@yahoo.co.in

[1] Department of Computer Science and Engineering, NIT Nagaland, Chumukedima, Dimapur 797103, India

Several methods have been developed in E-learning environment for interpreting the emotional state and concentration degree of the students (Krithika and Lakshmi Priya 2016). The existing Emotion recognition systems mainly concentrate on identifying the six universal emotions, i.e., Happiness, Sadness, Anger, Disgust, Fear and Surprise from small portions of the face by cropping only the eyes and mouth region (Yang et al. 2018). Detection of accurate emotions plays an important role in E-learning environment as emotions have a very crucial effect on learning where factors like illumination and pose of the image are not much involved.

Most of the existing E-learning emotion recognition system can identify the emotions only into classes of six universal emotions which are Happiness, Anger, Sadness, Disgust, Surprise and Fear. However, classification of these universal emotions doesn't help much in E-learning environment as we are more interested in knowing whether the target audience is interested or not. For example, if a participant's emotion is classified as 'happy.' It doesn't mean that he/she has understood what is being taught. The instructors are more interested in knowing the emotions which are desirable for E-learning environment. Thus, a method has been proposed for evaluating the emotional

state of the participants which detects the desirable emotions of the E-learning environment.

This paper classifies the desirable emotions from the dataset having various facial expressions of different subjects where the issue of change in illumination is addressed. Viola–Jones face detection algorithm is used for detecting various parts of the face, and local binary pattern (LBP) is used for extracting the features from the face. Then, classification of desirable emotion is done with the help of fuzzy neural network.

The main contribution of this paper is to identify the emotions which are desirable for E-learning environment other than the six universal emotions. Different portions of the face are considered in this study by using Viola–Jones face detection algorithm and local binary pattern (LBP) is used for extracting the features from the face. To achieve better prediction accuracy the proposed emotion recognition system has used fuzzy neural network for classification.

The paper is organized as follows: Sect. 2 describes the literature review. Details of the dataset are discussed in Sect. 3. Section 4 outlines the methodology of the proposed method with necessary steps. In Sect. 5, result analysis along with a comparative analysis is discussed. Finally, Sect. 6 includes the conclusion.

2 Literature review

Krithika and Lakshimi Priya (2016) detect the eyes and head movement for measuring the concentration level of the student. Three levels of concentration are detected—High, Medium and Low by monitoring the head rotation and eyes movement. Then, the concentration level of the student is detected by analyzing the entire three components data together. The major limitation is that it considers only eyes and head movement for identifying the concentration level of the student and also fails to recognize other possible required emotions.

Yang et al. (2018) proposed an emotional recognition model by detecting eyes and mouth region and classifying the emotions into six universal categories. Here, only mouth region is added along with the eyes for identifying the six basic emotions from the standard JAFFE database.

Ayvaz et al. (2017) developed a method for facial emotion recognition using various machine learning algorithms for classification. The limitation of this method is that it only classifies the six universal emotions into positive (happy and surprise) and negative emotions (sadness, disgust, fear and anger).

Ashwin et al. (2015) detect facial emotion recognition for multiple faces in a single frame for a group of E-learning students using SVM (support vector machine)-based supervised machine learning technique. They have classified the emotions into seven major categories: Happy, Sadness, Disgust, Anger, Fear, Surprise and Neutral, from the standard database in which the performance of the teaching strategy is not focused much.

Deshmukh et al. (2018) developed a facial expression recognition system which identifies facial expressions of students in real time and serves as a feedback to the instructor. It classifies the emotions into categories like—Neutral, Yawning, Sleeping and Smiling. Here, misclassifications among emotions are detected which is further solved by applying an ensemble approach which take more computation time during classification.

Various methods have been proposed for emotion recognition in E-learning such as fuzzy superior Mandelbrot sets, complex interval-valued Pythagorean fuzzy sets (Kahraman et al. 2020), and Schweizer–Sklar prioritized aggregation operators. However, these methods fail to classify the learner's emotion into emotions which are desirable for E-learning environment. Most of the methods are designed for classification of universal emotions.

3 Dataset collection

Due to the absence of standard dataset for emotion classification in E-learning, a dataset has been created. Several parameters have been considered for classifying the desired emotions. Table 1 gives the details of the parameters, and some sample images are shown in Fig. 1. The dataset contains face images of 40 different people showing various facial gestures in real-life environment with pixel size ranging between 200×200–400×400 in order to identify the desirable emotions of E-learning system. A dataset of 56,994 different faces is attained showing different movements of each portions of the face, i.e., eyes, mouth, eye brows, eyelids, lips, forehead, chin and head movement. This dataset is obtained to identify the desirable emotions from which 80% of the dataset is used for training and 20% for testing. Examples from the dataset are defined below:

For instance, in order to recognize "*Confused*" emotion, characteristics/patterns like brow lowerer, chin raiser, lip corner depressor, lid and lip tightener are considered.

For "*Boredom*" dimpler, lip pressor, yawning, lip tightener, sleeping or eyes closed are considered.

For identifying "*Concentration*" interested like outer brow raiser, upper lid raiser, inner brow raiser, chin raiser and for less interested eyes and head movement not facing the screen are examined.

For recognizing "*Self-Confidence,*" normal eye facing the screen is considered.

Table 1 Required emotions and the parameters defined to recognize those emotions

Emotions	Parameters
Confused	Brow Lowerer, Chin raised, Lid Tightener, Lip Corner Depressor and Lip Tightener
Boredom	Dimpler, Lip Pressor, Yawning, Lip Tightener and Sleeping
Concentration	*Interested*: Outer brow raiser, Upper lid raiser, Inner brow raiser
	Less Interested: Eyes and Head movement
Self-Confidence	Normal eye facing the screen

4 Proposed methodology

The proposed method develops an emotion recognition system for E-learning environment by considering multiple portions of the face (eyes, eyebrows, eyelids, forehead, mouth and chin) into emotion categories like Confused, Boredom, Concentration and Self-Confidence.

The method consists of two phase: (A) the training phase and (B) the testing phase. Figure 2 depicts the flowchart of the training phase.

4.1 Input image

Face images of various facial gestures of different people in real-life environment are captured and used as an input to the model. The captured images consist of characteristics of multiple portions of the face like eyes, mouth, eyebrows, eyelids, forehead, chin and lips of different people expressing four emotions: Confused, Boredom, Concentration and Self-Confidence.

4.2 Pre-processing

Image Resize- The input image is resized to a standard size 200×300 by using bilinear interpolation. Resizing of image is required as the size of the image captured from real life may vary. The resize image is then passed through a Gaussian low pass filter using Eq. (1) where the image quality is preserved. Gaussian low pass filter is used for removing noise from the image as it is an effective measure to enhance the robustness against rotations, blurring of an image etc.

As the results of the noise removal have a strong influence on the quality of the image, Gaussian low pass filter is applied in the proposed method where image quality is preserved before further processing.

$$F(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where $G(x, y, \sigma)$ Gaussian kernel, $I(x, y)$ image intensity at location (x, y), $F(x, y, \sigma)$ filtering result, σ standard deviation

The degree of smoothing is controlled by 'σ,' i.e., larger the value of 'σ,' the more intensive is the smoothing and this filter is also found to be computationally efficient.

4.3 Face detection

Face detection is carried out using Viola–Jones algorithm (Viola and Jones 2001) where the face is detected from the filtered image obtained in previous step.

4.4 Image segmentation

Image segmentation is the process of subdividing an image into multiple segments. So, the detected face is further segmented into various regions (eyes, eyebrows, forehead, mouth, chin and lips) of the face image.

4.5 Feature extraction

Feature extraction plays a major role for the whole recognition process. Even the best classifier will fail to achieve an accurate recognition rate if inadequate features are extracted. Here, the features are extracted based on LBP.

Local binary patterns (LBP) feature extraction process is performed for extracting the features from various detected face regions. LBP feature descriptor is based on the local texture representation providing specific and detailed face information which can be used not only to select faces, but also to provide face information for recognition. One of the advantages of using LBP feature descriptor is that it provides a compact representation to describe a patch of an image and is found to be more robust to noise. LBP feature descriptor (Huang et al. 2011) is a powerful approach to describe local structures which labels every pixel of an image by thresholding a 3×3 neighborhood of each pixel with the centered value and considering the results as a binary number and 256-bin histogram of the LBP labels computed over region is used as a texture descriptor. LBP can be computed using Eq. (2).

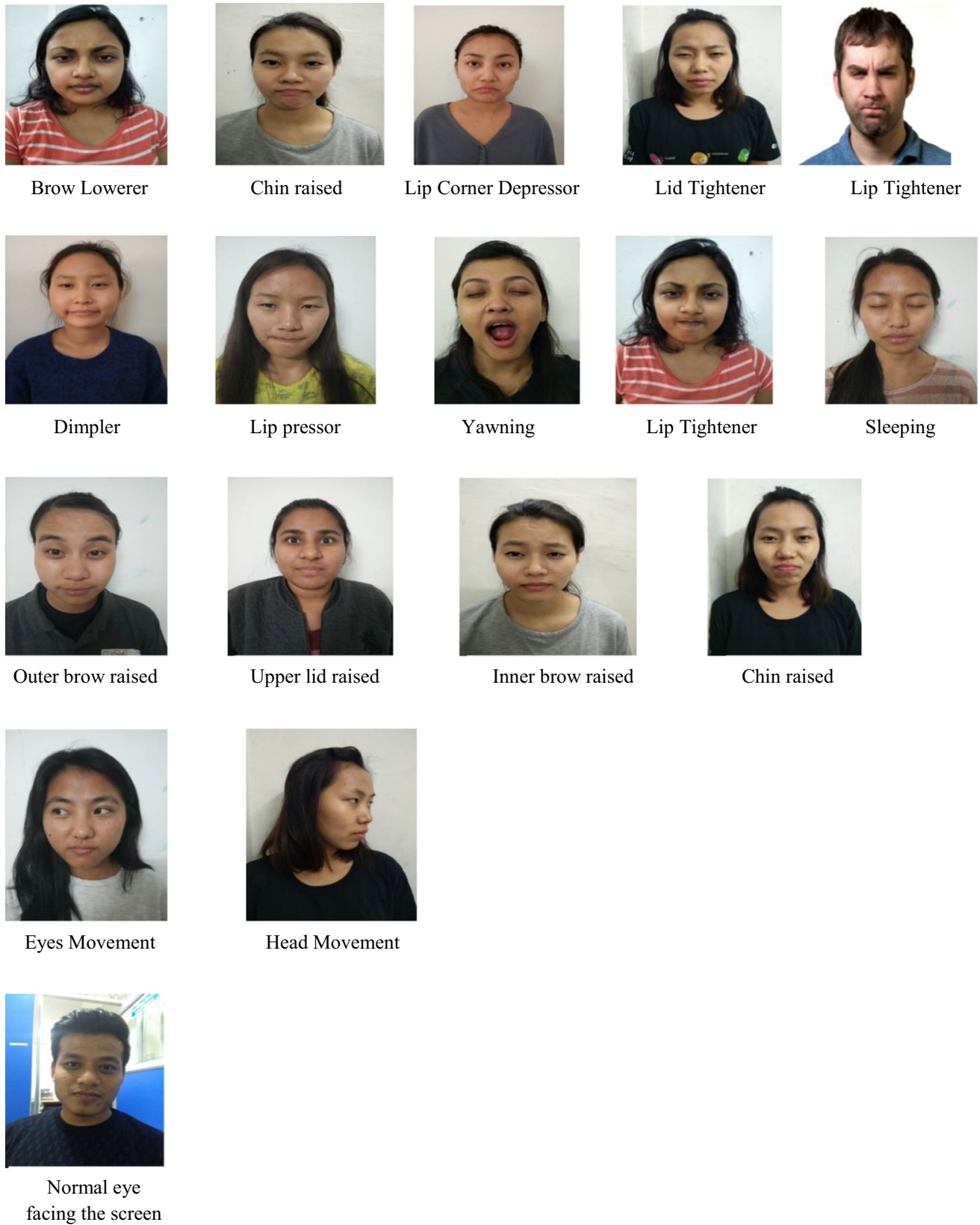$$\text{LBP}(x_c, y_c) = \sum_{n=0}^{7} 2^n S(i_n - i_c) \tag{2}$$

Fig. 1 Face sample images from the dataset

Fig. 2 Training model of the proposed system
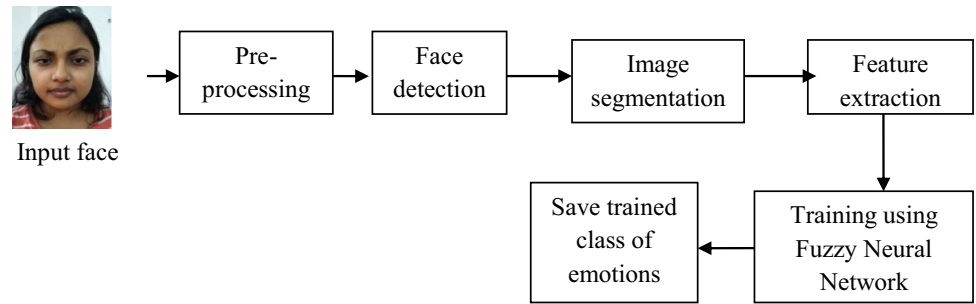

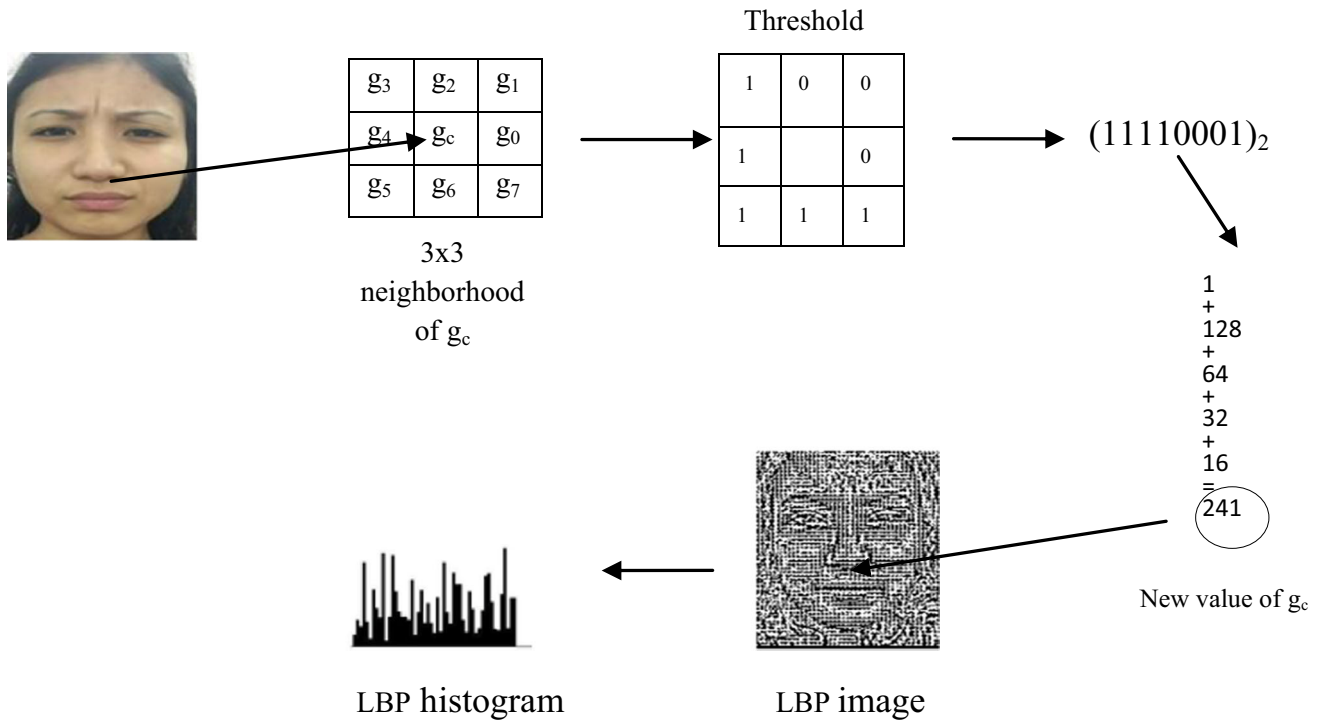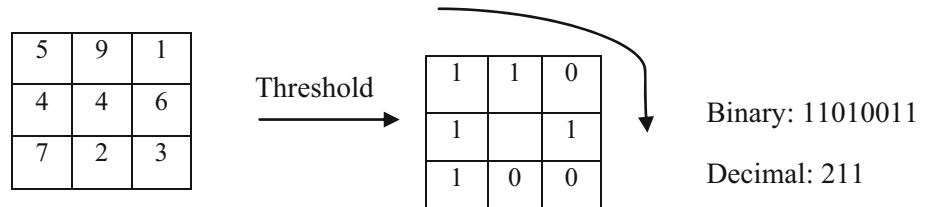Fig. 3 Shows an example of how a binary pattern is generated

Fig. 4 Depicts the various steps of LBP histogram generation

where in this case n runs over the 8 neighbors of the central pixel c, i_n and i_c are gray level values at c and $s(u) = 1$ if $u > = 0$ and 0 otherwise

After labeling an image with the LBP operator, a histogram of the labeled image is formed which contains information about the distribution of the local micro-patterns (Fig. 3).

LBP histogram is generated from all the face segments of an image and concatenated to form the final feature vector (FV) (Fig. 4).

4.6 Fuzzy neural network for training

The feature vector generated in the previous step is then fed as an input to fuzzy neural network (Kavitha et al. 2017)
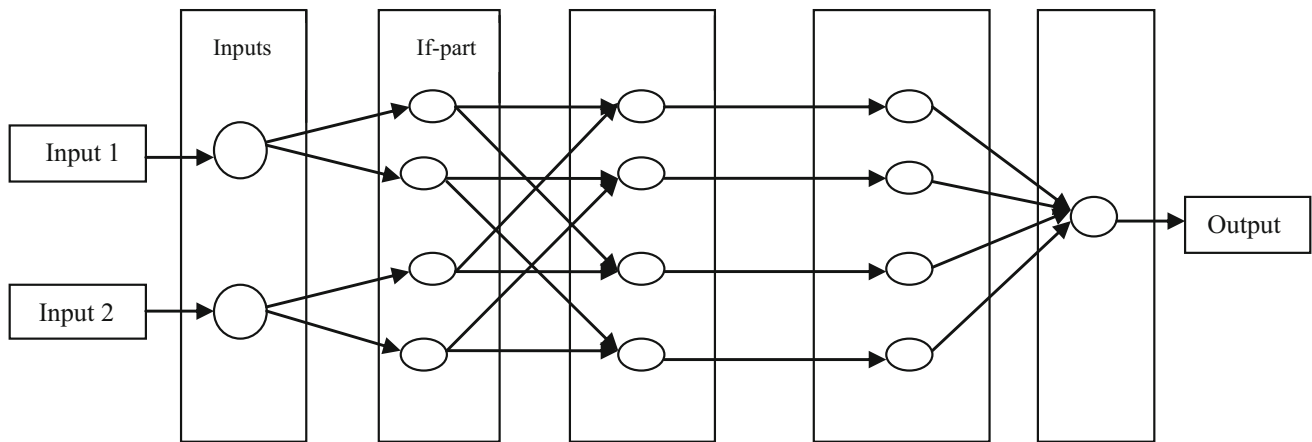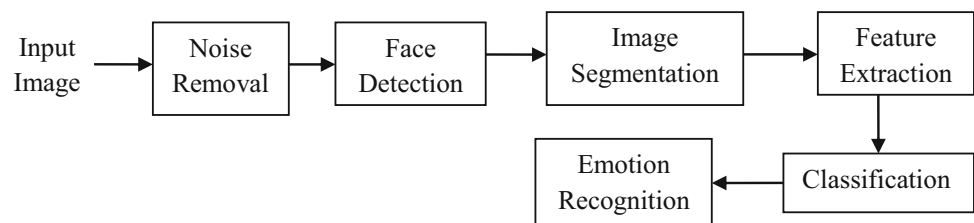
Fig. 5 The architecture of a fuzzy-neuro system

Fig. 6 Depicts the block diagram of the testing phase

which captures the benefits of both neural networks and fuzzy logic principles in a single framework. Fuzzy neural network is based on the concepts of fuzzy set theory, fuzzy if–then rules thus transparent to the user where the learning algorithms can learn both fuzzy sets and fuzzy rules and operate on local information, whereas other networks are black boxes to the user which cannot be interpreted.

A fuzzy neural network can be viewed as a multi-layer feedforward neural networks as shown in Fig. 5.

The process involves: membership functions, fuzzy logic operators and if–then rules.

A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value or degree of membership between 0 and 1. Fuzzy logic starts with the concept of fuzzy set. A fuzzy set is a set without a crisp, clearly defined boundary. It can contain elements with only a partial degree of membership. The fuzzy rule base is characterized in the form of if–then rules in which preconditions and consequents involve linguistic variables. A single fuzzy if–then rule assumes the form

If x is A then y is B

where A and B are linguistic values defined by fuzzy sets in the ranges x and y, respectively.

Here, the 'if' part of the rule 'x is A' is called the antecedent (precondition) or premise, while the 'then' part of the rule 'y is B' is called the consequent or conclusion.

Interpreting an if–then rule involves evaluating the antecedent and then applying that result to the consequent. Similarly, all the consequents are affected equally by the result of the antecedent. The consequent specifies a fuzzy set be assigned to the output. The implication function then modifies the fuzzy set to the degree specified by the antecedent. For multiple rules, the output of each rule is a fuzzy set. The output fuzzy sets for each rule are then aggregated into a single output fuzzy set. Finally, the resulting set is resolved to a single number.
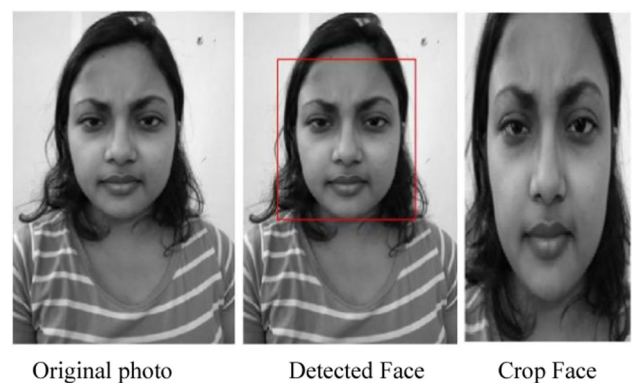

Original photo   Detected Face   Crop Face

Fig. 7 Detecting face from the original image using Viola–Jones

Fig. 8 Segmented parts of the detected face

Eyes   Eyebrows   Forehead   Mouth
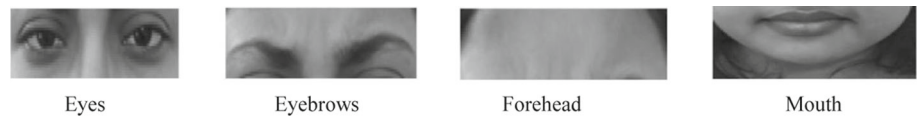
Fig. 9 LBP images of the segmented face parts (eyes, eyebrows, forehead, and mouth)

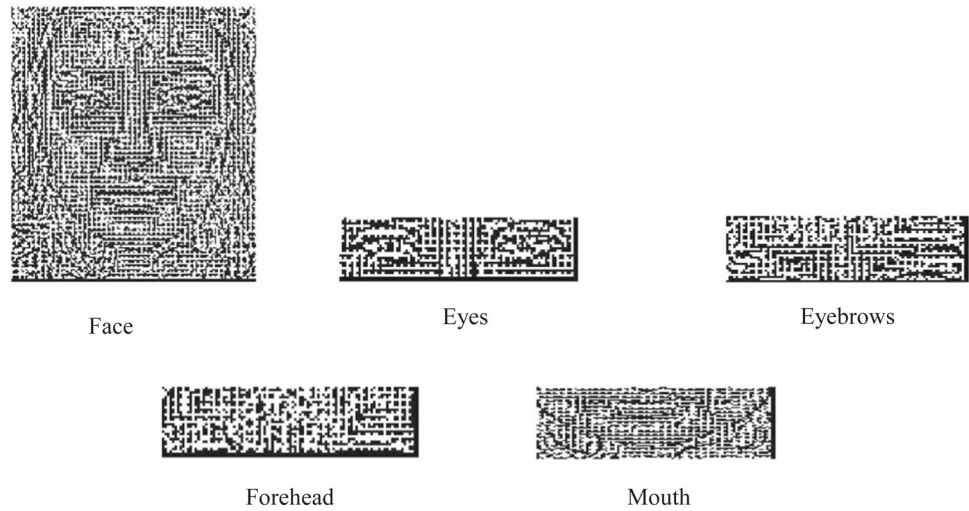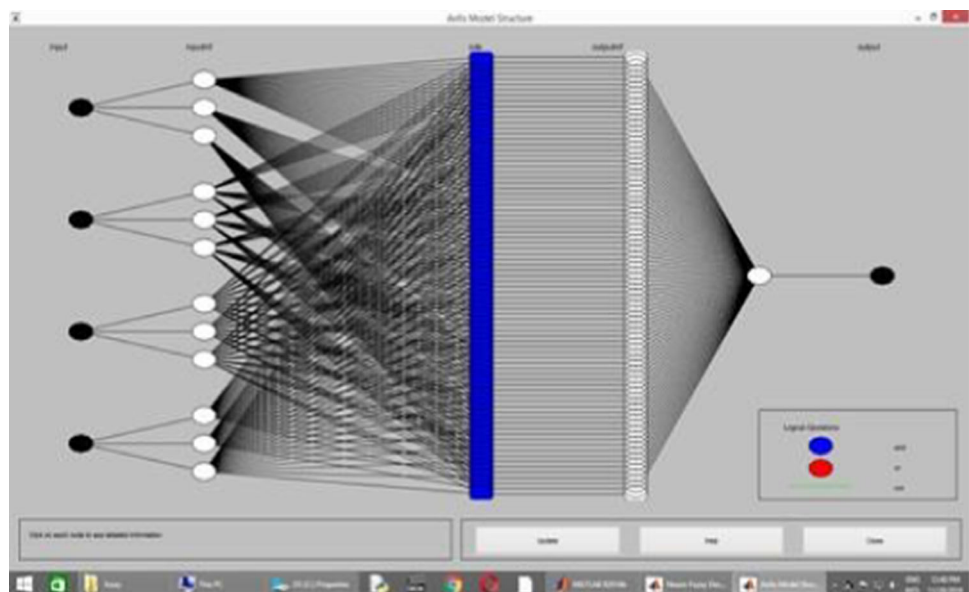Face   Eyes   Eyebrows

Forehead   Mouth

Fig. 10 Structure of a neuro-fuzzy system with four input node

Thus, the concept of fuzzy logic used in fuzzy neural network makes the development and implementation much simpler resulting in higher accuracy. As the learning procedure of fuzzy neural network operates on local information, it is found to be beneficial for the proposed method. By allowing users to insert prior knowledge into a network as rules, it learns features in the dataset and adjusts the system parameters accordingly.

In the testing phase, the same algorithm is used but on the test image instead of the training image for emotion recognition (Fig. 6).

5 Experimental results

The proposed method is simulated using MATLAB platform. At first, images from dataset are taken and noise is removed by using Gaussian low pass filter. Viola–Jones algorithm is applied for detecting the faces from the images as shown in Fig. 7.

The detected face is then further segmented into various parts (eyes, eyebrows, forehead, and mouth) as shown in Fig. 8.

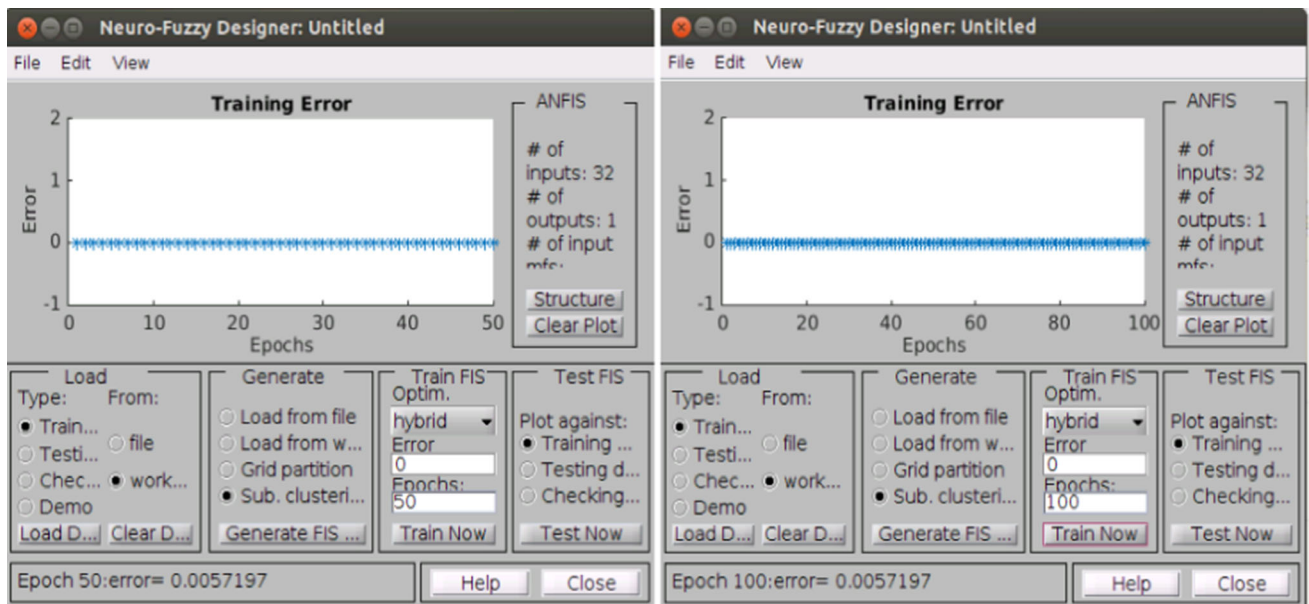After segmenting face into different parts, LBP method is applied for extracting the features.

Fig. 11 Training error rate at 50 and 100 epochs

Table 2 Error rate at different epochs

Number of epochs	Error rate
10	0.0058
30	0.0057307
50	0.0057197
100	0.0057197

Table 3 Accuracy at different bin numbers having different bin sizes

Number of bins	Bin size	Accuracy	Feature size
4	64	0.8575	16
8	32	0.9125	32
16	16	0.8600	64
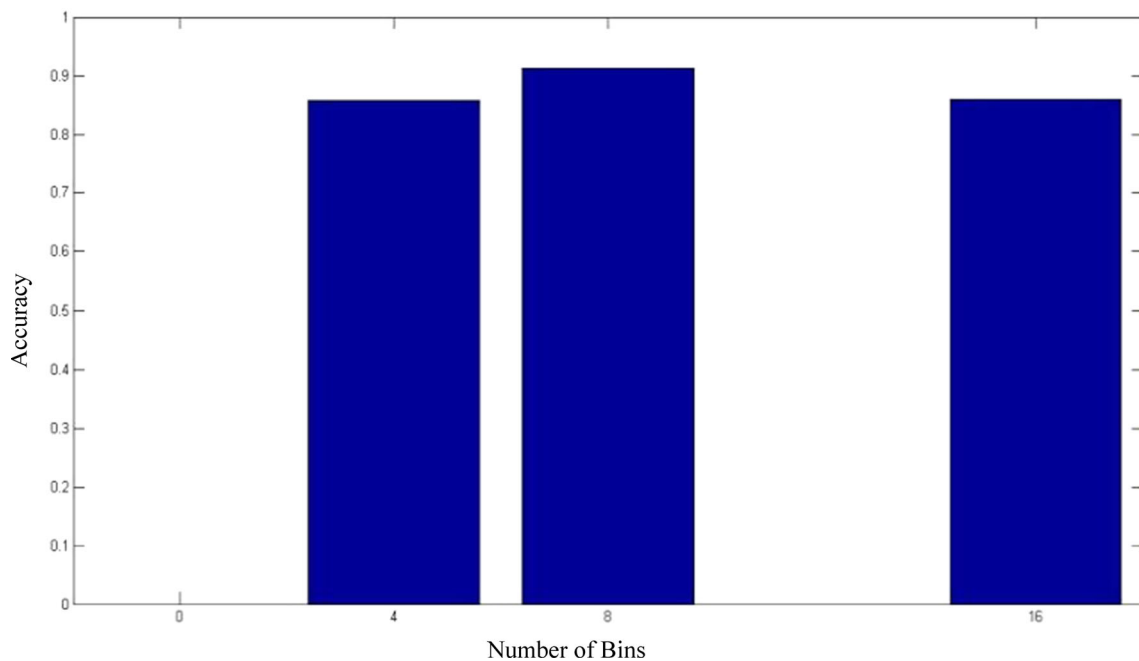
Fig. 12 Accuracy levels at different numbers of bins

Table 4 Performance analysis of existing methods and proposed method

Methods	Emotions classified	Accuracy
Krithika and Lakshmi Priya (2016)	High, medium and low concentration level	#
Yang et al. (2018)	Sad, surprise, happy, anger, disgust and fear	87.53%
Ashwin et al. (2015)	Sad, disgust, fear, anger, happy, surprise and neutral	89%
Deshmukh et al. (2018)	Smiling, neutral, yawning and sleeping	74.15%
Proposed method	Confused, boredom, concentration and self-confidence	91.25%

Figure 9 shows the LBP images of different segmented parts of the face.

Figure 10 depicts the fuzzy neural system which was generated after training the network with data consisting of four inputs of the dataset.

Finally, the dataset was trained with different number of epochs (10, 30, 50, 100) and an optimized result was obtained with 50 epochs having error rate = 0.0057197 as shown in Fig. 11 (Table 2).

The dataset was also trained on different number of bins (4 bins, 8 and 16 bins) having different bin size (64, 32, 16) as shown in Fig. 12 and an optimized result was obtained with 8 number of bins (32 bin size) acquiring higher prediction accuracy of about 91.25% with 50 epochs (Tables 3 and 4).

6 Conclusion

An emotion recognition system for E-Learning environment has been proposed to identify the emotions, namely Confusion, Boredom, Concentration, and Self-Confidence, by analyzing the facial expressions. The system utilizes the Viola–Jones algorithm to detect faces, which has a high detection rate, and segments the faces into various parts. Local binary patterns (LBP) feature extraction is applied to extract the features from each segment. The training dataset consisted of 80% of the data, while the remaining 20% was used for testing. Ultimately, the proposed method achieved an optimized result with 8 bins, yielding a prediction accuracy of approximately 91.25%.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Farzana Begum, Dr. Arambam Neelima and Dr. J. Arul Valan. The first draft of the manuscript was written by Farzana Begum, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript

Declarations

References

Ashwin TS, Jose J, Raghu G, Reddy GRM (2015) An E-learning system with multifacial emotion recognition using supervised machine learning. In: IEEE seventh international conference on technology for education

Ayvaz U, Gürüler H, Devrim MO (2017) Use of facial emotion recognition in E-learning systems. Inf Technol Learn Tools 60(4)

Deshmukh SP, Patwardhan MS, Mahajan AR (2018) Feedback based real time facial and head gesture recognition for E-learning system. In: International conference on data science and management of data

Huang D, Shan C, Ardebilian M, Wang Y, Chen L (2011) Local binary patterns and its application to facial image analysis: a survey. IEEE Trans Syst Man Cybern Part C (Appl Rev) 41(6)

Kahraman C, Bolturk E, Onar SC, Oztaysi B (2020) Modeling humanoid robots facial expressions using pythagorean fuzzy sets. J Intell Fuzzy Syst 39(5):6507–6515

Kavitha S, Paul V, Jothi Swaroopan NM (2017) Biometric emotion recognition using adaptive neuro fuzzy inference system. Middle-East J Sci Res 25(8):1644–1649

Krithika LB, Lakshmi Priya GG (2016) Student emotion recognition system (SERS) for E-learning improvement based on learner concentration metric. Proc Comput Sci 85:767–776

Viola PA, Jones MJ (2001) Rapid object detection using a boosted cascade. In: IEEE computer society conference on computer vision and pattern recognition CVPR, pp 511–518

Yang D, Alsadoon A, Prasad PWC, Singh AK, Elchouemi A (2018) An emotion recognition model based on facial recognition in virtual learning environment. Proc Comput Sci 125:2–10