
An Approach for Learner Categorization Based on Emotions in Intelligent Adaptive E-Learning Environment

Myneni Madhu Bala¹, Haritha Akkineni^{2,*}
and Chennupalli Srinivasulu¹

¹*Institute of Aeronautical Engineering, Hyderabad, India*

²*P.V.P Siddhartha Institute of Technology, Vijayawada, India*

*E-mail: baladandamudi@gmail.com; aharithapvpsit@gmail.com;
schennupalli@gmail.com*

**Corresponding Author*

Received 08 January 2022; Accepted 23 March 2022;

Publication 05 July 2022

Abstract

The pandemic across the globe has constrained the change from a conventional face to face to e-learning platforms. The most challenging task during online learning is to be aware and support the emotional side of students. In existing environments, the emotion of the listener consideration is lagging. This can be provided by capturing the emotions of the listener through facial expressions. In general, the most common facial expressions are happy, sad, anger, fear, disgust, neutral and surprise. This knowledge can be used to classify different listeners. Hence in this article, we proposed a novel approach to identify an emotion based learner category in the development of Intelligent Adaptive E-Learning Environment by using Convolution Neural Network. The major work is composed of emotion detection model and learner categorization. The emotion detection model is trained by using a standard FER2013 dataset and it is extended with live streams of learners. The results of emotion detection model are extended to categorize the learners

Journal of Mobile Multimedia, Vol. 18_6, 1709–1732.

doi: 10.13052/jmm1550-4646.18611

© 2022 River Publishers

by fusing emotions and comprehend as Active, Evaluative, Passive and Non-Listener. The proposed model is trained using 100 epochs and achieved an accuracy of 94.44% in the training phase. This knowledge helps to interpret learner's participation in e-learning environment.

Keywords: Convolution neural network, e-learning system, face emotion recognition.

1 Introduction

Online education is seeing its peak and almost the education system is seeing its transition. According to research, there is a close and consistent relationship between a person's facial gestures and emotions. Facial expression is among the most potent, natural, and universal ways for humans to communicate their emotional responses and intentions across national boundaries, racial group, and gender.

Students' emotions play an Emotions rule one's focus and sometimes can empower them to affect their self-regulation, so they play a significant role in their academic achievement. According to [1] self-regulated and motivated learning mediate the effects of emotions on educational excellence. Sometimes positive emotions will have positive impact on student learning whereas negative emotions affect in a negative way.

Research has already been carried identification of expression from faces. There are also some experimental investigations going on in the e-learning system applications. If there is no research done on how to know the expression of student, then it can be absurd because every educator's will be differing and they judge according to their perspective and knowledge they owe. These judgments may also differ in many situations because of several factors acting on them. But there is lack of suitable strategies to talk about the role of emotions in learning.

The emotions are responsible for the impact on intelligence like understanding a conversation, making a decision and also in the behavioural understanding in the aspect of a human being. Likely emotions play pivot role in the process of communication. Human emotions can be detected using voice signals, body language, facial expressions, etc. In communication, 55% of message is conveyed through facial expression of a speaker, 38% is expressed by speaker's voice and the original message to be conveyed will only contribute 7% in the verbal communication. In these cases, the automated and real time expression recognition in faces plays a significant

role in human-machine interaction. The facial expression recognition and classification is very useful from facilitating humans to many clinical practices and e-learning. Real time detection of emotions from faces has a wide variety of applications [2] in the fields of education, medical care, transportation, alert system, pain monitoring for patients and communication, and also has outstanding advantages at some specific scenarios. In education, this technology can help teachers in measuring students learning levels. The categories of the learners are active, passive, non-listener and evaluative listener.

To improve the detection of emotions in online platforms, several techniques have been implemented using machine learning algorithms in extracting facial features and taking smarter decisions. The main drawback is that there is no particular extraction algorithm to solve different problems and the concerns on the accuracy as it becomes constant when data size increases. To handle huge data, to find out the hidden patterns and learn relationships deep learning algorithms have been relied upon. Most of the methods adopted were weak in generality which is most essential to evaluate the practicality of the model [3]. Deep learning algorithms are capable of resolving this issue, and they are also robust in unfamiliar environments. Recent work has shown that convolutional neural networks (CNNs) performed well in tackling computer vision problems, particularly in FER [4, 5], due to its effectiveness in feature extraction and classification tasks, and innumerable models based on CNN structure are continuously suggested and have attained better outcomes than conventional techniques.

The primary goal of this study is to propose an effective method for analyzing student face emotions which can be integrated to the e-learning system and also understanding about the student's environment. This article concentrates on emotion identification in online classes and ensures that teachers receive regular feedback expressed through facial expressions, allowing them to flexibly adjust teaching programmes and, ultimately, improve the efficiency and effectiveness of online education. The paper is organized as follows: a brief background on emotions and related works in field of facial expression recognition. The next section followed by Proposed solution which describes the dataset used then followed about how the computer work and identify faces in images, the proposed CNN model, and the usage of identified emotions from images to integrate to the E-learning system to detect the type of learner. Then the next section describes the experimentation and results followed by their applications. Finally conclusions and references are presented.

2 Related Work

In the last 10 years, research has met exponential growth in the field of detection and recognition of emotions. A Huge count of innovative approaches and robust strategies has been anticipated. Generally, detection is about finding positions of emotions and recognition is about finding the class which given image belongs to.

2.1 Emotion and Learning

Recent studies [6, 7], have demonstrated that emotions have a significant impact on students' educational process. It suggests that the positive emotions influence the students' learning in terms of motivation, attention to studies and self-regulation whereas negative emotions affect the students' learning in performance and achievement. If the students' learning emotion is positive then student may be interested in learning and may end with positive emotion and same in the case of negative emotion. Different types of emotions will be triggering different views which can be related or unrelated to specified task and these also differ with intensity of emotion caused. Many researchers have shown the students experience will be caused by a great variety of emotions which can be task related or sometimes personal. Those social emotions such as hope, shame, boredom etc affect the students learning.

Liu et al. [8] have done a comparative study on different texture operators and feature descriptors like HOG and LBP and obtained a recognition rate of 98.3%. The linear filter local Gabor for texture analysis was used by Xing et al. [9] and obtained an accuracy of 95.1%. To obtain the facial expression combined with deep motion features Gupta et al. [10] developed a multi-speed Auto encoder network. Lopes et al. [11] presented a CNN model for facial expression classification with a main focus on issues related brightness of the image.

2.2 Categorizing Facial Expression & Its Features

For the need to categorize different emotions into different classes, a universal technique [12] which helps the computer to understand the features to be considered and extract from face to identify the emotions. This features position will be changing with respect to type of emotion a face possess. The change of features is given in Figure 1 and Table 1.

Along with this, FER2013 [13], a large-scale and unconstrained database introduced in the ICML 2013 Challenges in Representation Learning,

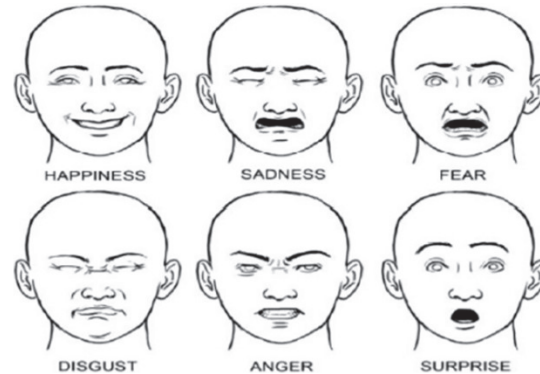


Figure 1 Emotion types.

labelled its facial images into anger, disgust, fear, happiness, sadness, surprise, and neutral, which has been widely used in designing facial expression recognition (FER) systems.

2.3 Facial Expression Recognition Techniques

In 1971, Ekman and Friesen have started working on Facial Emotion recognition. They have given the basic six emotions that can be identified in the face like happiness, sadness, anger, fear, disgust and surprise. This information contained in emotions will help to depict the mental state of an individual. Each and every person have a different way of showing the emotion, also individual person can also vary the expression of an emotion. This becomes one of the hard problem for a machine to deal with facial expressions. Sometimes the data may also be inadequate in terms of the illumination of light on camera or person, the head positions which makes difficult to find the features which determine the emotion of face and also other problems.

Further [14], there are being carried out a variety of researches within the area of e-learning systems to include the emotions so as to understand the emotional side of students. In a paper, it is stating that “Intelligent Tutorial Systems (ITS) are being privileged with feedback capabilities”[15]. They are able to send appropriate or effective and signals to learners and in their response, with an effective detection framework they ensure the emotional safety participation in the learning process. Similar works are also carried by combining emotion framework to ITS introduced by [16], which is telling that the system implemented in two varieties one is proactive system and another is reactive system. These individual systems will be mapping

Table 1 Universal emotion recognition

| Emotion | Description | Motion of Facial Part |
|------------------|---|--|
| Happiness | It is a feeling of pleasure and contentment in the way things are going and having enthusiasm for life. Secondary emotions are cheerfulness, pride, pleasure, thrill and relief. Happiness and health are interconnected. | Open eyes, cheeks raised, lip corner pulled up and mouth edge up. |
| Sadness | The feelings of grief, disappointment, disinterest and hopelessness. Prolonged sadness can affect one's health. Secondary emotions are suffering, hopelessness, hurt, despair and pity. | Lip corner pulled down, inner corner of eyebrows raised, outer eyebrow down, mouth edge down and closed eye. |
| Fear | This is a strong emotion that stems from our self preservation. We feel fear while we're in danger. When we are afraid, our physiological reactions ensure that we are prepared to deal with the risks in our environments. Secondary emotions are dread, horror, worry, panic and dread. | Mouth open, jaw dropped, inner eyebrow up and outer eyebrow down |
| Disgust | It occurs as a reaction to unpleasant or unwanted situations. Like anger, feelings of disgust can help to protect from things you want to avoid. | Lip corner depressor, eyebrows pulled down, nose wrinkle, lower lip depressor. |
| Anger | It is characterized by feelings of hostility, frustration, and agitation. Anger can be expressed in a number of ways, either through the tone from one's voice, yelling, or through physiological reactions such as the flushing of one's face or the use of assertive body language. While anger is generally regarded as a negative emotion, it can also be beneficial at times. Whenever we want to resolve the issue, it can empower us to develop answers or make decisions. Secondary emotions are irritation, annoyance, hate, frustration and dislike | Eyebrows pulled down, open eye, upper and lower lids pulled up, teeth shut and lips tightened. |
| Surprise | It is a physiological response to being startled. It occurs when unexpected things happen. Secondary emotions are surprise, astonishment and amazement. | Eyebrows up, open eye, mouth open, jaws dropped |

different components, their inter relationships and the dependencies in order to capture and to classify into different classes. This is done to support the positive as well as negative emotions in a creative way using a rich structure in the Intelligent Tutorial Systems.

Similarly, in [17], the authors proposed a system that monitors and identify the emotions of learners in e-learning system. It provides a real time feedback. It does by detecting the learner's concentration level through a continuous monitor of rotation of head and movement of eyes. This is done to enhance e-learning methodologies to improve content delivery of any course. The authors in [18] proposed a hybrid information system which is an interactive e-learning system to detect the learner's emotion. The purpose of this system is to give feedback to educator about the learners' present emotional state based on facial expression. This gives the educator information about the emotional state of learner in a virtual learning environment so that teaching may be improved.

Similarly, some authors discussed use of facial expression recognition for developing intelligent system to guide learner so as to enhance learning process. They have provided methods to supervise learners' work and their type of answering question in real time. The interaction in these types of learning environment also ensuring to recognize learner facial expression after they have finished a particular task or statement. The system continuously tracks the learners' knowledge.

Other researchers have introduced a fuzzy approach. These will handle unspecific behaviour in process of learning. For example, the authors in [19] used "fuzzy logic to gain new user's level by combining the emotion variable obtained by emotion recognition, exercise variables and current level in intelligent tutoring system".

Investigations were conducted on the emotional presence of learner using a qualitative methodology. Positive and negative emotions involved in learning were identified as two broad themes in the analysis [20]. Positive emotions included joy, enthusiasm, and excitement for the flexibility of online learning, which were stronger and more frequent in earlier months; pride and contentment for completing course requirements; and surprise and excitement for the emotional world of digital information exchange. Negative emotions included fear and anxiety for the unknown method of online learning and its demands (technology, time management, structure); alienation and the need for connectedness, which arose during first weeks of the semester and then when the students found it difficult to find satisfying forms of communication with their classmates and their instructor; and stress

and guilt for the inability to balance numerous roles and responsibilities, which is common in online learning.

Some recent studies focused on hierarchical features in CNN for facial expression recognition [21]. They proposed a image classifier to extract facial features and identify emotions using single face image. A general CNN which is derived through fusion of features like important mid-level and high-level features at different levels is considered for real time processing. A video discriminator analyzes single face image from a video using image classifier and determines the emotion. The proposed model outperformed single and ensemble models on the FER2013 dataset. In video-based facial expression recognition, the temporal model whose backbone is the proposed ensemble of MLCNNs also achieved comparable performance and outperformed other single video models.

In [22] the MBCC-CNN model is proposed which is a combination of residual connection, Network in Network, and tree structure approaches put together. This fusion of features prevent gradient vanishing and enhance the discrimination ability of the model in the receptive field and increases the recognition performance. The system can recognize facial expressions quickly and accurately, which is helpful to realize the intelligent and real-time application of expression recognition.

Methods including hybrid features like pixel and geometry gained popularity. In [23] a network consisting of Spatial Attention Convolutional Neural Network (SACNN) and a series of Long Short-term Memory networks with Attention mechanism (ALSTMs) is used. SACNN is used to extract the expressions from static face images and ALSTMs is designed to explore the potentials of facial landmarks for expression recognition. A deep geometric feature descriptor is proposed to characterize the relative geometric position correlation of facial landmarks. By jointly combining SACNN and ALSTMs, the hybrid features are obtained for expression recognition.

3 Proposed Solution

The proposed learner categorization system with automatic emotion detection is summarized as:

- Step 1: Extraction of learner images from online classes
- Step 2: Resizing to a standard dimension as 48×48 pixels.
- Step 3: Extracting features from faces using HAAR cascade algorithm.
- Step 4: Design a deep neural network (CNN) based emotion detection model with an optimized hyper parameters.

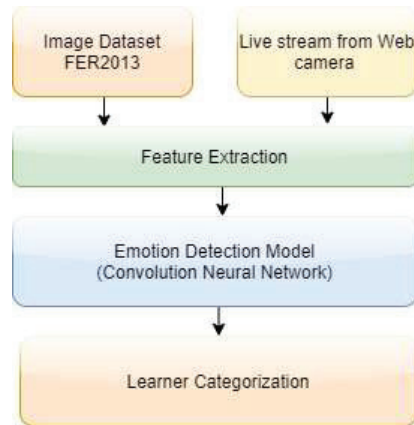


Figure 2 Schematic diagram of emotion based intelligent adaptive e-learning environment.

Step 5: Learner categorization with help of fused emotions

Step 6: Demonstration on an adaptive e-learning environment

Schematic Diagram of emotions based intelligent adaptive e learning environment is given in Figure 2.

3.1 Exploring the Data

The model is trained and tested on FER2013 [24] data set. FER2013 [25] is an open-source dataset which is first created for an ongoing project. It is split into training and testing samples with split ratio 80:20. The FER2013 data set contains total of about 35887 images in which 28709 images are to train the model and 7178 images are to validate the model. The data set consists of 48×48 pixel grey scale images of faces. The faces have been adjusted in a way that the face is centred more or less and also occupies the same quantity of space in each image. Emotion labels in the dataset. The detailed categorization of dataset is given in Table 2.

3.2 Feature Extraction

Face detection is the first step in the FERS. Face detection can be realized on images. As a part of pre-processing the facial images from Google meet are identified and cropped using HAAR Cascade and OpenCV library. As plenty of people are moving to the online platforms especially for classroom teaching the actual scenes are complex and changeable due to various factors like lighting etc. FER in controlled environments or captured in labs is

Table 2 Dataset categorization

| Class Number | Photo Quantity | Class Name |
|--------------|----------------|------------|
| 0 | 4593 | Angry |
| 1 | 547 | Disgust |
| 2 | 5121 | Fear |
| 3 | 8989 | Happy |
| 4 | 6077 | Sad |
| 5 | 4002 | Surprise |
| 6 | 6198 | Neutral |

generally considered to be less practical for feature extraction or classification purpose. Whereas the uncontrolled databases, such as FER2013, are collected from complex environments with vastly different backgrounds, occlusions, and illuminations; these scenes are more similar to the actual situations and can serve as an important source and will be more appropriate for the work which we considered.

Effective pre-processing can reduce the interference of face-like objects in the background when detecting faces in an image and then standardize the face images according to the heuristic knowledge, which will effectively improve the efficiency of the model chosen for analysis.

The colour images are converted into grey scale images. This conversion is important because the colour images mostly don't recognize the important edges and features.

Face detection is done using HAAR Cascades on OpenCV. This is an effective way to achieve face detection as suggested by Paul Viola and Michael Jones [26]. This approach is based on the machine learning process. Here the cascade function is trained in many positive images which means that that contain face images and negative images which means those that contain faceless images. This is used to find faces in photos. HAAR features are basically of three groups. That is the features of a line, edge and four rectangles [27]. Here, each element is made up of a white square and a black square. Then subtracting the sum of all the pixels in the white square from the ratio of all the pixels in the black square we get the same number of elements. The purpose of these features is to distinguish the faceless part from the face part and to indicate the presence of characters in certain images. The detected faces are cropped. Figure 3 shows the one of the pre-processing step extraction of faces from images.

By this pre-processing step the image is turned into a greyscale image then the faces are detected and those faces are cropped using the Haar cascade function. The cropped image in online platforms is of various sizes for



Figure 3 Pre-processing of captured image for extracting face.

example 563×362 pixels. Then the cropped image is resized to 48×48 -pixel image and those resized images are fed to the proposed CNN model.

3.3 Emotion Detection System

Prior to recognition of the facial expression an analysis on the facial landmarks is being carried out. It denotes the changes that occur in sections of the face when the emotional moods changes. The edges of these sections are determined and followed by facial landmarks.

In the proposed work, a convolution neural network model is being used to recognize emotions based on facial expression from face, integrated to e-learning system. For this purpose, a publicly available database named FER2013 [26] to carry out the experiment. Extracting features in face is done using the proposed CNN model. Images from the pre-processing step are fed to the CNN model. The main benefit of CNN is that it accepts 2D images directly as input data. The proposed CNN architecture consists of a total of 13 layers which consists of 4 convolution layers, 3 dropout layers, 3 Max pooling layers, 1 flatten layer and 2 fully connected layers. The first one contains 32 filters having 46×46 size, the second layer contains 64 filters third and fourth layers contain 128 filters each having the size of 20×20 and 8×8 respectively.

3.3.1 Convolution layer

The convolution layer is used to extract features such as shapes, corners and edges to obtain feature maps. Each convolution layer uses a Stride equal to 2 and a valid padding.

The output of convolution layer is calculated as follows given as Equations (1) and (2).

$$Width = (W - F_w + 2P)/S_w + 1 \quad (1)$$

$$Height = (H - F_h + 2P)/S_h + 1 \quad (2)$$

Here, (W, H) are the input size of given image, (F_w; F_h) are the size of filter, P is zero padding and (S_w; S_h) are the stride of the convolution operation.

At the end of each convolution layer, we apply the activation function ReLu. Rectified Linear Units (ReLus) are used as the activation function in the convolutional and max pooling layers to avoid gradient outburst and to ensure faster processing speed during the back-propagation operation, which can be formulated as shown in Equation (3)

$$f(x) = \begin{cases} 0, & x \leq 0; \\ x, & x > 0. \end{cases} \quad (3)$$

Softmax is used as the activation function in the output layer, the input of which is the output from fully connected layer which is given as Equation (4)

$$S_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}} \quad (4)$$

Where K represents the output dimensions of the layer and S_i represents the probability of the result i, i = 1, 2, ..., K. The Figure 4. shows the CNN architecture which shows a detailed view of the process adopted.

3.3.2 Max pooling layer

It is used to reduce the size of the input image by half of its original size if the stride is 2 when the image is too large. The pooling layer comes after each convolution layer. It is a dimension reduction technique that reduces the dimensions of a layer in order to create a new layer, allowing the network to train faster by focusing on the most relevant information.

MAX pooling with 2×2 filter is considered. The size of the flatten vector is 700 which feed a dense layer with 128 nodes. It accepts an input of W1 × H1 × D1 and produces W2 × H2 × D2, where D is the depth. It is calculated as shown in Equations (5), (6) and (7).

$$W2 = (W1 - F_w)/S_w + 1 \quad (5)$$

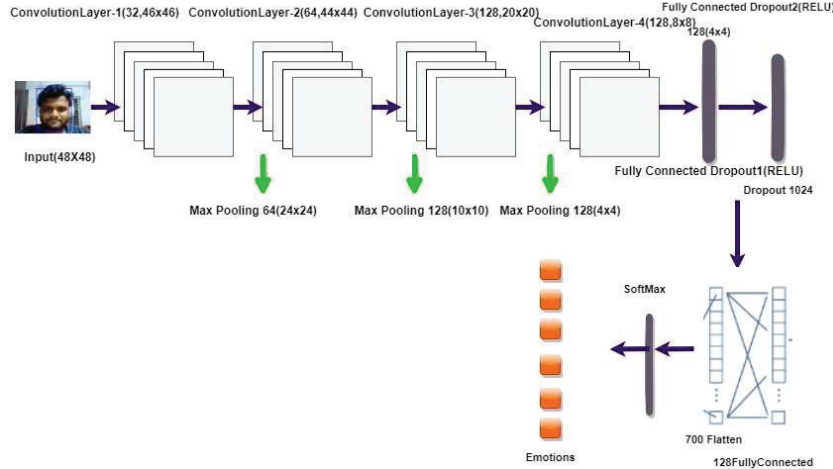


Figure 4 Proposed convolution neural network architecture.

$$H2 = (H1 - F_h) / S_h + 1 \tag{6}$$

$$D2 = D1 \tag{7}$$

Max pooling is a pooling operation that selects the most detail from the vicinity of the characteristic map included by the filter. For this reason, the output after max-pooling layer might be a function map containing the most outstanding functions of the previous function map. The maximum common shape is a pooling layer with filters of length 2×2 implemented with a stride of two down samples every intensity slice within the input by using 2 along width and height.

3.3.3 Dropout layer

To avoid over fitting, dropout layer will be randomly setting the input unit values to 0 and using a frequency of a particular rate during each and every iteration at training time. So, these layers are used multiple times until the system reaches the required number of features. Then, the 2D vector output is manipulated to a 1D vector. This vector is given as input to CNN model by using means of fully connected otherwise dense layers. The activation function used in the last layer of CNN model is softmax function by which we can perform multi-classification.

We have trained the Convolutional Neural Network model using 100 epochs with a 64-batch size. The model uses ReLu activation function in

Table 3 Hyper parameters

| Hyper-parameters | Values |
|------------------|---|
| Input Size | 48×48 |
| Convolution 1 | 32(46×46)Stride 2 Same Padding, RELU |
| MAX Pooling | Input Size(2×2) |
| Convolution 2 | 64(44×44) Stride 2 Same Padding, RELU |
| MAX Pooling | Input Size(2×2) |
| Convolution 3 | 128(20×20) Stride 2 Same Padding, RELU |
| MAX Pooling | Input Size(2×2) |
| Convolution 4 | 128(8×8) Stride 2 Same Padding, RELU |
| MAX Pooling | Input Size(2×2) |
| Flatten Vector | 700 neurons |
| Full connected | 128 neurons, RELU |
| Output | 7 neurons, Softmax |
| Epochs | 100 |
| Batch Size | 64 |
| Learning Rate | 0.0001 |
| Optimizer | Adams |

the hidden layers and softmax function in output layer. The model uses Adam optimizer with a learning rate of 0.0001. Table 3 depicts the hyper parameters and their corresponding values.

3.4 Learner Categorization

The final step is to classify the images using a network that is fully connected. The input images of learner data have been processed step by step and finally categorize the learners based on their emotions from facial expressions. A sample case of emotion detection is shown in Figure 5.

Fully connected layer allows classifying the feature vector generated by the previous layers based on a multilayer perception.

The model will be classifying the faces present in the image into 7 emotions those are happy, sad, angry, disgust, fear, surprise and neutral. Since, the purpose of detecting the emotions from faces of students is to integrate the output emotions into the e-learning system to tell the category of listener the student belongs to [28]. According to the earlier studies the

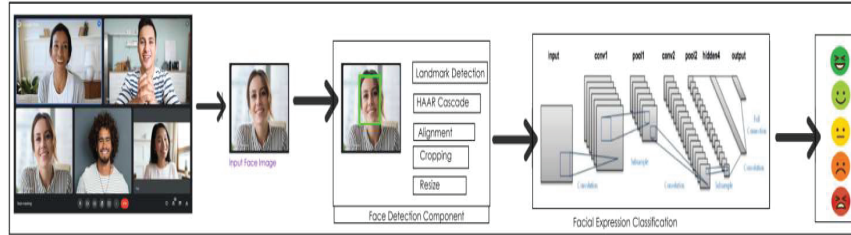


Figure 5 Illustration of sample case of emotion detection system.

students learning levels are broadly classified as four categories based on the emotion of the listener as Active Learner, Passive Learner, Non-Listener and Evaluative Listener. By the way of classifying students' category of listener will be assisting the teacher about which listener a particular student belongs to which in turn helps them to improve the strategy of teaching, sometimes to focus on the emotional side of students.

The emotion to listener category is as follows:

- Happy or Surprise: Active Learner
- Angry or Disgusted: Passive Learner
- Fearful or Sad: Non-Listener
- Neutral: Evaluative Listener

In order to assess the classification model performance, the metrics considered are accuracy and loss. Accuracy denotes the count of predictions and the predicted value needs to be true value. The accuracy is calculated during the training phase which will result in overall final accuracy.

To get a more nuanced view into how well the model is performing, loss function is used to consider the uncertainty of prediction based on the deviation from the true value. The main goal is to minimize the loss value.

The accuracy of a model is defined as number of correct predictions done by the model to the total number of predictions made by the model as shown in Equation (8)

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of predictions}} \quad (8)$$

The Loss of the proposed model is calculated using the categorical cross entropy. The formula is stated as shown in Equation (9).

$$\text{loss}_{\log} = (y \log(p) + (1 - y) \log(1 - p)) \quad (9)$$

4 Experimental Results

The proposed model has shown accuracy of 94.4% by using the following parameters of 100 epochs with batch size of 64. The performance of model is depicted in graphs are as follows. Python libraries such as OpenCV, NumPy, pandas, PIL and TensorFlow are used for pre-processing and feature extraction. The operating system used is Windows 10 with i5-8th gen processor and 8GB RAM on GPU NVIDIA GeForce GTX 1050Ti and used python 3.8. Following are the results conducted using a live webcam feed.

A GUI is developed using PyQt5 library in python as shown in Figures 6, 7, 8 and 9. The webcam feed is taken using OpenCV library. The image

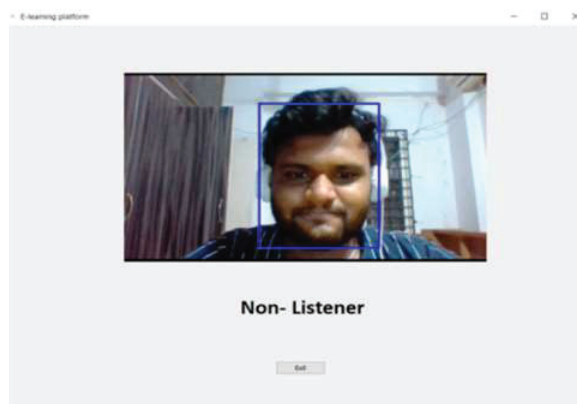


Figure 6 Non listener.

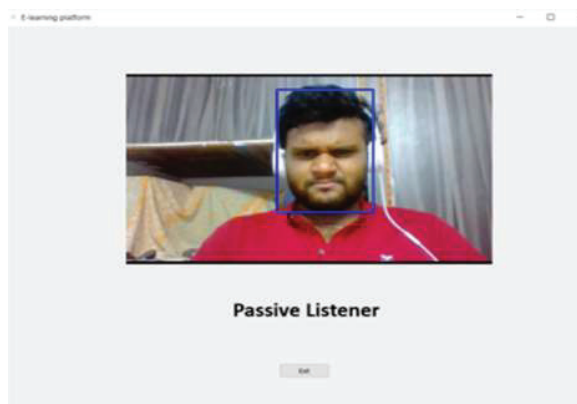


Figure 7 Passive listener.

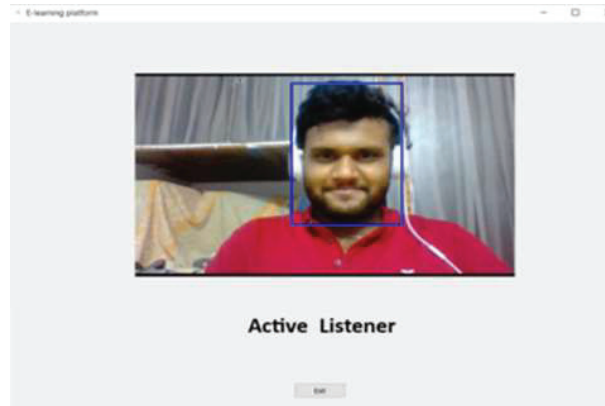


Figure 8 Active listener.

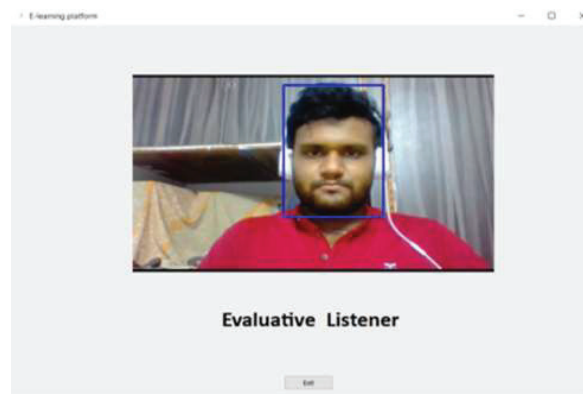
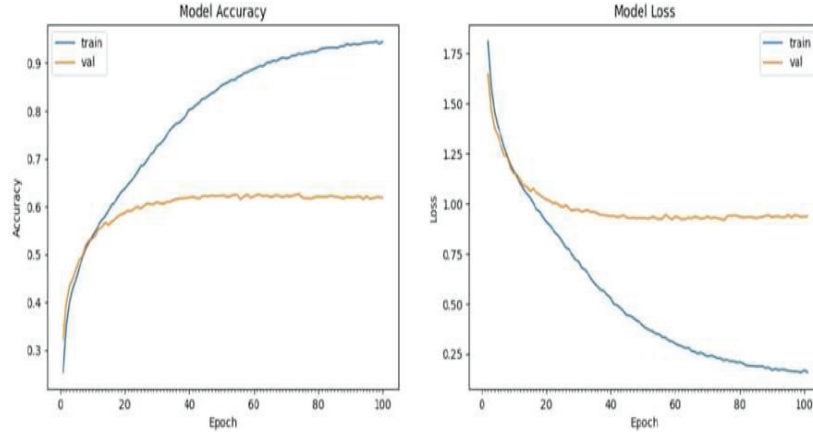


Figure 9 Evaluative listener.

is pre-processed and given to model which will be predicting the emotion and thereby the system automatically categorizes the type of listener and then output that text in GUI. Based on the training the model, the images are classified into Active, Evaluative, Passive and Non-Listener. The model which is proposed is trained using 100 epochs and achieved an accuracy of 94.44% in the training phase. The model is well trained and this model will be effectively recognizing the emotion and can be integrated to the e-learning system [29].

Figure 10 gives the accuracy and loss during the training and validation phases. The accuracy in training phase is 94.4% and during the validation

**Figure 10** Model accuracy and loss.**Table 4** Accuracy with different epochs

| Epoch | Accuracy |
|-------|----------|
| 100 | 94.4% |
| 150 | 96.81% |
| 200 | 97.40% |

Table 5 Comparison with other models

| Epoch | Accuracy |
|-------------------|----------|
| CNN [30] | 72.40% |
| CNN [31] | 65% |
| CNN [32] | 61.10% |
| Proposed Solution | 94.44% |

phase it reached to 58%. The loss value is nearly 1 and 0.12 during training and testing phases respectively.

If the data has got higher diversity, then greater number of epochs would be needed. As there is no diversity in the data we have considered we tested our model considering 100, 150, and 200 epoch. As the number of epoch increase the longer it needs to be trained and we arrived at the results in Table 4. We fixed and tested our model using 100 epochs in calculating the metrics to assess the classifiers performance.

Table 5 gives a comparison with other models which are stated in references. It describes the fact that the proposed CNN model has outperformed the previous CNN models in the case of accuracy by using the same FER2013

dataset. So it can be integrated to the learning platform so as to categorize the type of listener.

5 Conclusions

Emotions, in the field of online learning can stand as a distracter but will definitely serve as facilitator in support of decision making, directing in positive ways and stimulation. CNN was used to help in emotion recognition from faces. The test was successful and the system was perfectly able to detect faces and classify the type of learner. The model has done with an accuracy of 94.44% by using the FER2013 dataset. The outcomes show that type of learner is detected and that the system reached state of the art results. By interpreting the student emotion in the context of online class can result in valuable information. This article statistically demonstrated that students' facial expressions are the most commonly used nonverbal communication method in the classroom, and that students' expressions are significantly correlated to their emotions, which can aid in acknowledging their knowledge of the lecture in e-learning systems. In this experimentation the major focus is on identification of various types of the listeners based on their types of facial expressions. This can be extended in adaptive e learning system by monitoring at topic level.

References

- [1] C. Mega, L. Ronconi, and R. De Beni, What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement., *J. Educ. Psychol.*, vol. 106, no. 1, p. 121, 2014.
- [2] C. H. Wu, New technology for developing facial expression recognition in e-learning, in *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*, pp. 1719–1722, 2016.
- [3] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon, Emotion recognition in the wild challenge 2014: baseline, data and protocol, in *Proceedings of the 16th International Conference on Multimodal Interaction*, pp. 461–466, ACM, Istanbul Turkey, November 2014.
- [4] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, Attention mechanism-based CNN for facial expression recognition, *Neurocomputing*, vol. 411, pp. 340–350, 2020.

- [5] J. Shao and Y. Qian, Three convolutional neural network models for facial expression recognition in the wild, *Neurocomputing*, vol. 355, pp. 82–92, 2019.
- [6] F. Benmarrakchi, J. E. Kafi, A. Elhore, and S. Haie, Exploring the use of the ICT in supporting dyslexic students preferred learning styles : A preliminary evaluation, *Educ. Inf. Technol.*, pp. 1–19, Oct. 2016.
- [7] O. El Hammoumi, F. Benmarrakchi, N. Ouherrou, J. El Kafi and A. El Hore, Emotion Recognition in E-learning Systems, 2018 6th International Conference on Multimedia Computing and Systems (ICMCS), Rabat, Morocco, pp. 1–6, doi: 10.1109/ICMCS.2018.8525872, 2018.
- [8] Liu Y, Li Y, Ma X, Song R, Facial expression recognition with fusion features extracted from salient facial areas. *Sensors* 17:712, 2017.
- [9] Xing Y, Luo W, Facial expression recognition using local Gabor features and adaboost classifiers. In: 2016 international conference on progress in informatics and computing (pic). IEEE, pp. 228–232, 2016.
- [10] Gupta O, Raviv D, Raskar R, Multi velocity neural networks for facial expression recognition in videos. *IEEE Trans Affect Comput* 10: 290–296, 2017.
- [11] Lopes AT, De Aguiar E, Oliveira-Santos T, A facial expression recognition system using convolutional networks. In: 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE, pp. 273–280, 2015.
- [12] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). *SN Appl. Sci.* 2, 446, <https://doi.org/10.1007/s42452-020-2234-1>, 2020.
- [13] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., Challenges in representation learning: a report on three machine learning contests, *Neural Information Processing*, Springer, Berlin, Germany, pp. 117–124, 2013.
- [14] Shan, K., Guo, J., You, W., Lu, D., and Bie, R, Automatic facial expression recognition based on a deep convolutional-neural-network structure. 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 123–128, 2017.
- [15] Mohammed Megahed, Ammar Mohammed, Modeling adaptive e-learning environment using facial expressions and fuzzy logic, *Expert Systems with Applications*, Volume 157, 113460, ISSN 0957- 174, 2020.
- [16] J. M. Harley, S. P. Lajoie, C. Frasson, and N. C. Hall, An Integrated Emotion-Aware Framework for Intelligent Tutoring Systems, in *Artificial Intelligence in Education*, pp. 616–619, 2015.

- [17] Khalfallah, J., Slama, J. B. H., Facial expression recognition for intelligent tutoring systems in remote laboratories platform. *Procedia computer science*, 73, 274–281. doi: 10.1016/j.procs.2015.12.030, 2015.
- [18] Zatarain-Cabada, R., Barrón-Estrada, M. L., García-Lizárraga, J., Muñoz Sandoval, G., Ríos-Félix, J. M. Java tutoring system with facial and text emotion recognition. *Research in Computing Science*, 106, 49–58, 2015.
- [19] Phan-Xuan, H., Le-Tien, T., Nguyen-Tan, S., Fpga platform applied for facial expression recognition system using convolutional neural networks. *Procedia computer science*, 151, 651–658. doi: 10.1016/j.procs.2019.04.087, 2019.
- [20] Ayvaz, U., Gürüler, H., Devrim, M. O., Use of facial emotion recognition in e-learning systems. *Information Technologies and Learning Tools*, 60(4), 95–104. doi: 10.33407/itlt.v60i4.1743, 2017.
- [21] H.-D. Nguyen, S.-H. Kim, G.-S. Lee, H.-J. Yang, I.-S. Na and S.-H. Kim, “Facial Expression Recognition Using a Temporal Ensemble of Multi-Level Convolutional Neural Networks,” in *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 226–237, 1 Jan.–March 2022, doi: 10.1109/TAFFC.2019.2946540.
- [22] C. Shi, C. Tan and L. Wang, “A Facial Expression Recognition Method Based on a Multibranch Cross Connection Convolutional Neural Network,” in *IEEE Access*, vol. 9, pp. 39255–39274, 2021, doi: 10.1109/ACCESS.2021.3063493.
- [23] C. Liu, K. Hirota, J. Ma, Z. Jia and Y. Dai, “Facial Expression Recognition Using Hybrid Features of Pixel and Geometry,” in *IEEE Access*, vol. 9, pp. 18876–18889, 2021, doi: 10.1109/ACCESS.2021.3054332.
- [24] Agrawal, A., Mittal, N., Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. *The Visual computer*. doi: 10.1007/s00371-019-01630-9, 2019.
- [25] N.M. Deepika and Myneni Madhu Bala and Ravi Kumar, “Design and implementation of intelligent virtual laboratory using RASA framework”, *Materials Today: Proceedings 2021* ISSN 2214-7853. doi.org/10.1016/j.matpr.2021.01.226, 2021.
- [26] Monika Dubey, Prof. Lokesh Singh, Automatic Emotion Recognition Using Facial Expression: A Review, *International Research Journal of Engineering and Technology*, Volume: 03 e-ISSN: 2395-0056, 2016.
- [27] C. Jain, K. Sawant, M. Rehman and R. Kumar, “Emotion Detection and Characterization using Facial Features,” 2018 3rd International

- Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, pp. 1–6, doi: 10.1109/ICRAIE.2018.8710406, 2018.
- [28] Krithika, L. B., Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. *Procedia computer science*, 85, 767–776. doi: 10.1016/j.procs.2016.05.264, 2016.
- [29] Dandamudi, Rohit and Myneni, Madhu Bala, A Smart Personalized Learning Environment Through Social QA System (December 4, 2020). e-journal – First Pan IIT International Management Conference 2018.
- [30] Breuer R, Kimmel R, A deep learning perspective on the origin of facial expressions. *arXivPrepr arXiv170501842*, 2017.
- [31] Fan Y, Lu X, Li D, Liu Y, Video-based emotion recognition using CNN-RNN and C3D hybrid networks. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. pp. 445–450, 2016.
- [32] Mollahosseini A, Chan D, Mahoor MH, Going deeper in facial expression recognition using deep neural networks. In: *2016 IEEE Winter conference on applications of computer vision (WACV)*. IEEE, pp. 1–10, 2016.

Biographies



Myneni Madhu Bala is currently working as a professor and Dean of computational studies at Institute of Aeronautical Engineering, Hyderabad. She received her Ph.D in Computer Science and Engineering from JNTUH. She has Twenty one years of academic and research experience Her research interests are Data Science frameworks, Image Mining, Text mining, Machine learning, Artificial Intelligence, Deep Learning and Data Analytics. She has published about 57 articles in reputed Journals indexed in SCOPUS, SCI etc.

She has published 2 patents. She has received grants from AICTE for organizing Short Term Training Programs and Faculty Development Programs. She is a reviewer for Elsevier, Springer and more indexed journals. She acted as sessionchair, organising member, and advisory member for various International Conferences. She delivered various invited talks on Data Modelling, Data Science and Analytics. She is a Life member of professional bodies like CSI and ISTE, Sr. Member for IEEE, WIE & International association IAENG, ICST, SDIWC.



Haritha Akkineni is currently an associate professor in Information Technology at PVP Siddhartha Institute of Technology, Vijayawada. She received her Ph.D in Computer Science and Engineering. She is working in the area of Opinion Mining and Data Sciences. She has twelve years of academic and research experience Her research interests are Data Science, Image Mining, Artificial Intelligence, Data Analytics, Deep Learning and Machine Learning. She has published about 35 papers in reputed Journals like SCOPUS UGC etc. She has published 2 patents. She has received grants from AICTE for organizing Short Term Training Programs. She is a reviewer for SCOPUS indexed journals. She authored a book on Opinion Mining. She acted as Workshop/tutorial chair for various International Conferences. She delivered various invited talks.



Chennupalli Srinivasulu is working as Professor of CSE at Institute of Aeronautical Engineering (IARE). He has 24 Years of experience in Teaching and Industry. His Area of Interest is Parallel Computing, Soft Computing, Software Engineering etc. He has published more than 38 Publications in reputed International Journals.