IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.106107

# Detection of Student's Affective States in Classroom using CNN

**Neha Pawar[1], Shubhangi Funde[2], Revati Kshirsagar[3], Vaishnavi Kaulagi[4]**

Progressive Education Society's Modern College of Engineering, Pune[1-4]

**Abstract:** Predicting the student's emotional engagements using Computer vision techniques are a challenging task. There are several works on computer vision based affective state recognition of students in the e-learning environment, there are limited works on affective state recognition of students in the classroom environment where more than one Student are present in a single image frame. Face recognition has become an attractive field in computer-based application development in the last few decades. The learning process has also evolved a lot. However, the emotion of students is usually neglected in the learning process. This project is mainly concerned about using facial expression to detect emotion in the learning environment. There are many algorithms for facial recognition and emotion capturing out of which we have used Convolutional neural network (CNN).The captured facial expression will be used in the Learning Environment for analyzing the learner mood. The proposed architecture uses the students' facial expressions for analyzing their affective states. The experimental results will predict the probability of affective states of the faces detected in learning environment for understanding of emotions during learning process in order to enhance the learning and feedback achieving process.

**Keywords:** Face Detection, Face emotion recognition, Convolution neural network, OpenCV, Machine learning.

## I. INTRODUCTION

Humans interact with each other mainly through speech, but also through body gestures, to emphasize certain parts of their speech and to display emotions. One of the important ways humans display emotions is through facial expressions which are a very important part of communication. Facial expressions are the facial changes in response to a person's internal emotional states, intentions, or social communications. Facial emotion recognition is the process of detecting human emotions from facial expressions. The human brain recognizes emotions automatically, and software has now been developed that can recognize emotions as well. This technology is becoming more accurate all the time, and will eventually be able to read emotions as well as our brains do. The automatic prediction of the student's affective states were less explored in the classroom environment. The existing studies predict the students' emotional and behavioral engagement separately in both e-learning and classroom environments. The use of multi-modality for recognizing the students' affective states was very less explored. An automatic prediction of group-level students' engagement was not explored. Finally, there exists no standard dataset to train, test, and validate the machine learning/deep learning models in classroom environments. This motivated us to propose a convolutional neural network architecture to automatically predict the students' affective states in the classroom. In this paper, we present an approach based on Convolutional Neural Networks (CNN) for facial expression recognition. The input into our system is an image; then, we use CNN to predict the facial expression label which should be one these labels: anger, happiness, fear, sadness, disgust and neutral.

## II. RELATED WORK

Abhilash Dubbaka And Anandha Gopalan [1] proposes a system which will uses webcam to monitor faces of learner watching MOOC (Massive Open Online Course). This paper categorize facial expressions and translate their expressions into learners engagement level in online learning environment. This paper explores the use of webcams to record students' facial expressions whilst they watched educational video material to analyse their Learner Engagement levels. Convolutional neural networks (CNNs) were trained to detect facial action units, which were mapped onto two psychological measurements, valence (emotional state).

NIGEL BOSCH and SIDNEY K. D'MELLO, University of Notre Dame JACLYN OCUMPAUGH and RYAN S. BAKER, Teachers College, Columbia University VALERIE SHUTE, Florida State University [2] uses computer vision and machine learning techniques to detect student's affective state from both facial expression and gross body movement during interaction with an educational physics game.

"Automatic detection of students' affective states in classroom environment using hybrid CNN "by Ashwin
T.S. , Ram Mohana Reddy Guddeti. [3] This paper describes how Convolution Neural Network can be used to predict overall affective state of entire class. This paper uses student's facial expressions with hand gesture and body posture to

analyze their affective states.

In 2016, Pramerdorfer and Kampel obtained state-of-the art, which is 75.2% accuracy on the FER2013, using Convolutional Neural Networks (CNNs) [4]. The authors used an ensemble of CNNs using VGG, Inception, and ResNet with depths of 10, 16, and 33, with parameters of 1.8m, 1.6m, and 5.3m, respectively. The authors used the face images as given in the dataset, and for illumination correction, they used histogram equalization. They performed horizontal mirroring for training data augmentation and randomly cropped images to the size of 48 x 48 pixels. They also trained the architecture for up to 300 epochs and used stochastic gradient descent to optimize the cross-entropy loss, with a momentum value 0.9. The other parameters were fixed, like learning rate with 0.1, batch size with 128, and weight decay with 0.0001.

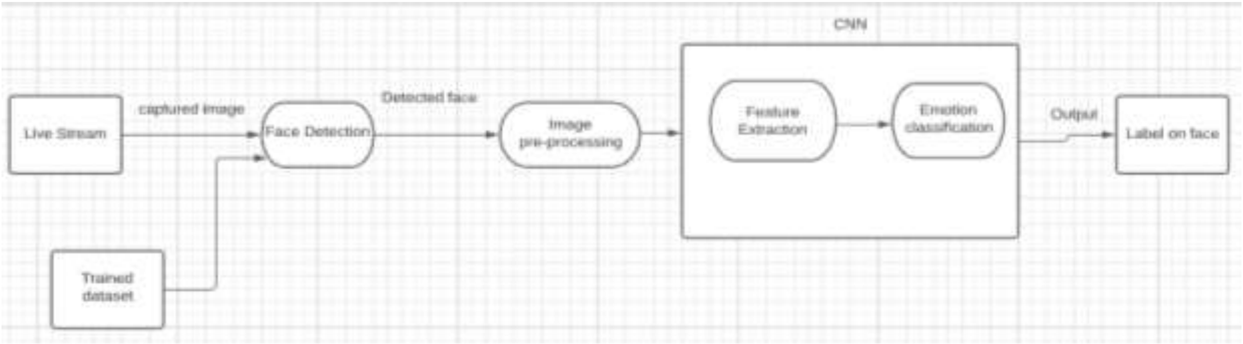
Zhang et al. [5] used a Siamese Network to introduce a method for understanding social relation behaviors from images and achieved a test accuracy of 75.1% on the challenging Kaggle facial expression dataset. The authors used multiple datasets, with various labels, to increase the training data; they also introduced a feature extraction method and patch-based registration, as well as working on feature integration via early fusion.

Kim et al. [6] proposed an ensemble of CNNs and demonstrated that during training and testing it is advantageous to use both registered and unregistered forms of given face images. The authors achieved a test accuracy of 73.73% on the FER2013 dataset. They also conducted Intraface for a conventional 2-D alignment, which is publicly available for landmark detector, and performed illumination normalization. To avoid the registration error, they performed registration selectively, based on the results of facial landmark detection.

## III. DATASET USED

The dataset is downloaded from website: https://www.kaggle.com/msambare/fer2013. The data consists of 48x48 pixel grayscale images of faces. The Dataset provides seven categories: [Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral]. The training set consists of 28,709 examples. The public test set consists of 3,589 examples.

## IV. SYSTEM ARCHITECTURE



Architecture diagram

Above architecture diagram shows proposed model for affective states of student in classroom environment. From live stream the data will be taken as input. The image will be given as input to next process i.e face detection**.** From the entire captured image only face will be detected after scanning. Detected face will be input for next process where image will be resized and the rgb values will be converted into grey values. These values will be used by CNN to extract features and classify emotions. For the pre trained model the values will be matched and the category for which the values are having most inclination towards will be taken as the emotion state.

### A. Pre-processing

Pre-processing can be used to enhance FER system performance and can be done previous to the feature extraction process. Image pre-processing has various processes, such as the detection and alignment of faces, correction of illumination, pose, occlusion, and data augmentation.

In the FER2013 dataset, the faces are registered automatically, so that they have similar space requirements and are more or less centered in the images. face detection is done using the Haar Cascade classifier [11]. When images are captured in various types of light, expression features are sometimes inaccurately detected, and therefore, the expression recognition rate can be low and make feature extraction more difficult

### B. Feature Extraction

The extraction of facial features requires translating the input data into a set of features. By using feature extraction,

DOI 10.17148/IJARCCE.2021.106107

researchers can reduce an immense amount of data down to a relatively small set, which allows for faster computation. We applied dlib facial landmark detector pre-trained on iBUG 300-W dataset [12], [13], [14] for feature extraction, and extracted the eight most prominent parts of a face, including both eyebrows, both eyes, the nose, the inner and outer outlines of the mouth, and the jaw. In which we extracted the right and left eyes, nose, and inner and outer outline of the mouth, which are marked with yellow color.

### C. CNN Architecture

CNNs have been widely used in a variety of computer vision applications, including FER. Early in the 21st century, several studies of FER literature [15], [16] determined that CNNs work well on changes of face location, as well as variations in scale. They were also found to work better than multilayer perceptron (MLP) when looking at face pose variations not seen previously. Researchers used CNN to help solve various facial expression recognition problems, such as translation, rotation, subject independence, and scale invariance [17]. Our model was trained using the following characteristics:

- six convolutional layers using “RELU” as an activation function;
- three max-pooling: out of which the first two using pool size (3,3) and stride (2,2), and the third using pool size (2,2) and stride (2,2); every max pooling is followed by every two convolutional layers; two drop out with value 0.2;
- one flattened layer and two dense layers: one dense layer with “RELU,” and the other with “Softmax” as an activation function;
- total parameters and trainable parameters are 1.2 million, respectively.

## V. CONCLUSION

The project will explored the student’s affective states in the classroom environment as both emotional and Behavioural engagement are considered to predict student’s affective states on the basis of emotions like engages, bored, happy, neutral, etc.

## REFERENCES

[1] IEEE Conference by Abhilash Dubbaka Anandha Gopalan, Department of Computing Imperial College London London, United Kingdom abhilash., “Detecting Learner Engagement in MOOCs using Automatic Facial Expression Recognition.”

[2] ACM Transactions by Nigel bosch Sidney K. D mello Ryan Baker. “Using video to automatically detect learner affect in computer enabled classrooms.”

[3] ScienceDirect by Ashwin T.S., Ram Mohana Reddy Guddeti. “Automatic detection of students’ affective states in classroom environment using hybrid CNN.”

[4] Another website used for dataset collection: https://www.kaggle.com/c/challenges-in representation-learning-facial-expression recognition-challenge/data