Emotion Analysis of Students

[1]Amanda Judy Andrade, [2]Komal Kadam, [3]Nilam Kadam

[1,2,3]Students of Don Bosco Institute of Technology, Affiliated to University of Mumbai,

Department of Information Technology, Premier Automobile Road, Kurla-West, Mumbai – 400070, Maharashtra, India.

Abstract: *Mental health organizations embedded inside instructive frameworks can make a continuum of integrative thought that improves passionate prosperity and informative accomplishment for youngsters. A reconfiguration of preparing and dynamic health structures to help execution verification-based practice might be required to strengthen this continuum and ideal child improvement. Integrative methods that merge study lobby level and understudy level interventions have a ton of possibilities. Summarizing the proposed procedure by experimenting with it in group and collaborative learning stages, an unpretentious understudy responsibility examination can be used to make intelligent instructing structures more modified.*

Keywords: ***Emotion Recognition, Fully Connected Convolution Neural Network.***

I. Introduction

The student's emotional health and all-around performance in Academics, Extra-Curricular Activities are directly correlated to each other. Government bodies appropriate funds for education and [1], [2] human welfare in most developing countries to enhance their economy; they require a prolific educated workforce to act on demand. Tending to the issue of the student's emotional health by distinguishing it at the beginning phase and following up on the necessary changes can assist with expanding the overall performance.

It tends to be hard for instructors to distinguish tension and despondency because these issues frequently show up diversely for various individuals; however, this is the reason knowing the blends of practices to search for is critical. An understudy managing one of these issues can encounter negative consequences for consideration, translation, focus, memory, social interaction, and actual wellbeing. When somebody is facing uneasiness or wretchedness, most of their intellectual ability is utilized to make and deal with troubling contemplations. This can make it amazingly hard to zero in on particular musings and can be exceptionally debilitating for the understudy, which takes away from their learning capacities.

Psychological wellness administrations inserted inside educational systems can make a continuum of integrative consideration that improves kids' emotional wellbeing and instructive achievement. A reconfiguration of training and dynamic wellness frameworks to help execute proof-based practice may be required to fortify this continuum and for ideal kid improvement. Integrative techniques that consolidate study hall level and understudy level intercessions have a lot of potentials. A robust exploration plan is needed that centres around framework-level execution and support of interventions over the long run. Both moral and logical avocations exist for a mix of emotional wellness and training: coordination democratizes admittance to administrations and, whenever combined with utilization of proof-based practices, can advance the sound improvement of kids.

Summing up the proposed strategy by testing it in collective and social learning stages, a subtle understudy commitment investigation can be utilized to make wise coaching frameworks more customized.

This paper functions as precise documentation of the implemented model developed to simplify the student-teacher relationship process by capitalizing on the deep learning algorithms.

II. Related Work

Many researchers have done their work on the emotion recognition of students. Since emotional states play a fundamental role in human communication and social contact, automatic emotional state recognition has been an active research area these past few years. The authors have chosen these five papers as their primary focus of research.

Krithika L.B and Lakshmi Priya G.G. in [3] suggested a system that would detect emotions using Gaussian distance between the eyes, thus eliminating the requirement of a physical contact device for recognizing emotions and classifying learner involvement and interest in the subject. This helped in improvising the faculty feedback system to enhance the learning system. This serves as a stepping stone to our proposed approach.

T. S. Ashwin and Ram Mohana Reddy Guddeti in [4] put forward a system that would use non-verbal cues for analyzing student engagement in the classroom. They used 350 students present in a classroom as test subjects, utilizing faces, hand gestures, and body postures as training data. This project inspired the authors to deploy the planned approach in a classroom as all other projects presented were aimed at online learning systems, or they deployed it using Microsoft Kinect, a proprietary system that could only record 40 faces for a lecture of 40 minutes, which doesn't help in college/ university level wherein each subject has a slot of 60 minutes minimum allocated.

Jianzhu Guo, Zhen Lei, Jun Wan, Egils Avots et al. in [5] specified that pairs of multifaceted emotions were difficult to be categorized than seven basic emotions. The identification of compound emotions on the iCV-MEFED dataset was revealed to be very challenging, leaving a large room for improvement. This proved that the utilization of the iCV-MEFED dataset might not be fruitful. This prompted the authors to settle with the FER2013 dataset for data analysis, visualization.

The authors A. Sharma and V. Mansotra [6] proposed a classroom-based facial recognition using Deep Learning approaches. They took an entirely different approach from previous projects that selected SVM, RNN or ANN for training their models. This was the first project that implemented advanced techniques. They trained the FER2013 dataset and utilized the transfer learning technique to pre-train the Cohn-Kanade (CK+) dataset. They aimed to capture the mood of the entire classroom in general, which they were successful. This paper served as a beacon to the proposed project to implement it in a classroom using CCTV camera videos instead of a high-resolution digital video camera for capturing a student's emotions for a given lecture.

The authors Yuhao Tang, Qirong Mao, Hongjie Jia et al. in [7] presented an emotion-embedded visual attention model (EVAM) for learning emotion context data for forecasting affective dimension values from capture on tape episodes. They employed a gated recurrent unit (GRU) for contextual learning of facial feature sequences. They carried out their experiments on AVEC 2016 and AVEC 2017 datasets which is desirable. However, the panel members rejected the idea of using datasets that require permission to be used and suggested that we stick to open-source datasets like the FER2013 for data analysis. Hence, we had to drop the idea of utilizing the mentioned dataset even though it provided a better yield than the FER2013 dataset.

III. Aim and Objectives

The focal aim of the software is to predict and express the student emotion in real-time as fast and as accurate as possible. By automating the process that requires professionals from various fields, this project aims to increase the performance of students' test scores by detecting mood changes during a real-time classroom environment that can be addressed immediately.

a. Constraints:

1. **Latency:** Given an image, the software should foretell the expression instantaneously and transfer the result. Hence, there is a low latency requirement.
2. **Interpretability:** Interpretability is essential for still images but not in real-time. For still photos, the probability of predicted expressions can be given.
3. **Accuracy:** Our target is to foretell the expression of a face in the image as accurately as possible. The higher the test accuracy, the better our model will perform in the real world.

b. Performance Metrics

1. **Multi-Class Log-loss:** The authors have used a deep learning model with a cross-entropy layer in the end with seven SoftMax units, so, therefore, our goal is to reduce the multi-class log loss/cross-entropy loss.
2. **Accuracy:** This tells us how accurately our model performs in predicting the expressions.
3. **Confusion Metric:** Since our problem is multi-class classification, the confusion metric will help us know which classes are more dominant or, towards which category, the model is more biased. This gave us a clear illustration of the prediction score of the model.

IV. Proposed Work

The exploration started by investigating existing examination papers and characterizing our challenging assertion. Whenever this was accomplished, could we have gone to remain of choosing the last strategy for execution. The creators have proposed that carrying a solitary individual identifies whether the calculation works precisely. On the conversation with the board individuals, they needed to utilize the FER2013 dataset. They have effectively executed the proposed procedure in the ReLU 4- layered 2D R-CNN Design with the calculation's learning rate (α) to 0.0005, β_1=0.9, β_2=0.999, and epsilon ε=e-5 with the precision of the model at 66.7% and lower misfortune pace of 89% more than 50 epochs, in this way astounding past models.

A. Dataset Collection and Preprocessing

The dataset comprises 48 × 48 pixel grayscale pictures of countenances. The appearances have been consequently enrolled with the goal that the face is pretty much focused and possesses a similar measure of room in each image. The mission is to order each face reliant on the feeling that appeared outward into one of the emotion classes (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral).

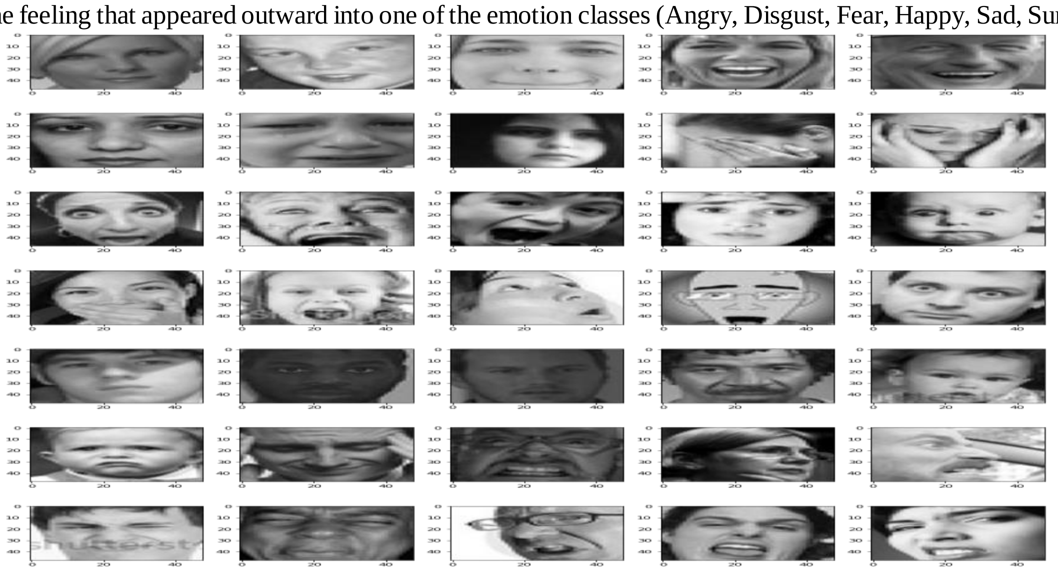

Figure 1: Sample Data Images from the FER2013 Dataset

The constituent of this string is space-isolated pixel regard in line with substantial request. The preparation set comprises 28,709 models. The public test collection utilized for the leader board shall consist of 3,589 models. The last test group, which was used to decide the medalist of the opposition, comprises other 3,589 models.

Data pre-processing includes grayscale conversion of RGB images, face detection and crop, image normalization, and image augmentation.

B. Proposed Model

The model used for this Application is Fully Connected Convolution Neural Network (FC-CNN). It is a popular method chosen for image classification with data augmentation for this research.

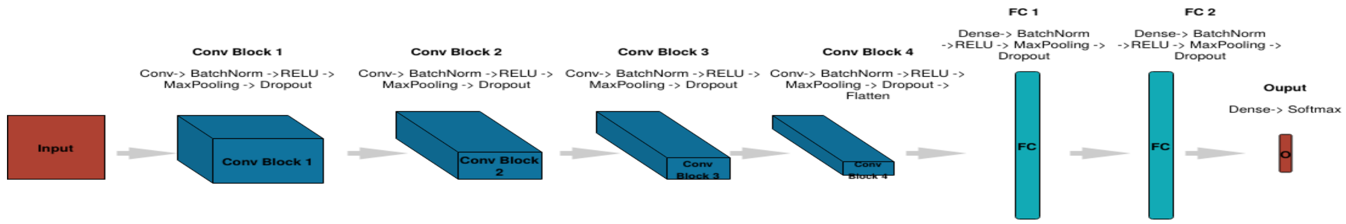

Figure 2: Inspired by Goodfellow, I.J., et al. (2013). Challenged in representation learning: A report of three machine learning contests. Neural Networks, 64, 59-63. DOI: 10.1016/j.neunet.2014.09.005

First, images from the dataset were fed into the face detection system using the Haar Cascade Classifier to detect faces in the input image. If looks are seen from the visual input, the input is passed to the ImageDataGenerator function from Keras API for further data augmentation such as flipping, rotating, shearing, etc. The face then gives into the CNN classifier to predict classes.

The proposed model used to classify facial expression consists of 4 convolution layers with 64, 128, 512, and 512 filters, respectively, with the 3 × 3 kernels. The conv2D is used to specify the convolution kernel, which is 3 × 3 kernel. The max-pooling layer is used for dimension reduction by finding the maximum value in the 2 × 2 kernel windows. The dropout layer is included to avoid overfitting problems. The batch-normalization layer performs the training time optimization. The activation function expended in all layers except the turnout layer is Rectified Linear Unit (ReLU) activation function. The ReLU is a function aiming to converge cost to zero faster and boost accurate results. The authors have used Categorical Cross-entropy as a loss function in the turnout layer, which is used for multimodal classification for distinguishing seven classes of emotion.

The Adaptive Moment Estimation (Adam) optimizer was deployed, which is used to update our neural network weights iteratively. The Adam optimization algorithm augments stochastic gradient descent, improving upon it by using the arithmetic mean of the second moments of the gradients instead of maintaining a single learning rate for all weight updates, where the learning rate does not change during training. It also has a comparatively more minor memory requirement and commendably reduces the training cost in large datasets over multiple iterations.

The number of batches corresponds directly to the number of variables processed before updating the model to account for the predicted values. Again, by exhaustive hyperparameter tuning, our model gave us the most accurate results when the batch size was 64 for the training set.

Once our data was pre-processed, reshaped, and tuned, we passed it to the model and evaluated its accuracy. We then saved and loaded our model to output reproducible results as in each run of the model, as the weights are arbitrarily assigned, leading to some variance in the production. Once this process was complete, our model was performing optimally and outputting near-perfect trends.

C. Flask RESTful API

Flask, a Python library, is an atomic web framework used to create application program interfaces, i.e., Flask has almost no dependency on external libraries to build a web-based application. It produces the user interface, and other functions such as database connectivity, form substantiation, upload regulation, and various certification technologies are backed by extension libraries available within Flask. We have built our entire software using the Flask library because of its similitude with database applications, cross-compatibility with Python programming language, and most importantly, practicality. To make our Application more plausible, we used the Flask-RESTful extension library. REST (Representational State Transfer) API is cacheable communication and client-server protocol that uses HTTP requests to GET, PUT, POST, and DELETE data. It works impeccably and follows stateless strictures. In simple words, when a client parses context, the server can stock and process this context without additional client inquiries.

Flask enables the use of basic structures like HTML, CSS, and JavaScript. HTML and CSS are used to control the presentation, formatting, and layout of the web-based Application. JavaScript is used to manage the behaviour of different elements of the design. In a Flask web application, there is a determinate structure. There is the main folder where all the Python files, contributing to the functionality of the entire software, are stored.

Along with those files, there are two additional folders, namely *templates* and *static*, in the same main folder. The templates folder is used to store the HTML, CSS, and JavaScript templates, and the static folder is a folder containing all the images used in the User Interface. In the main Python file, we have to link all the .html files with the help of '@app.route'. Therefore, it helps to map a function, with specific tasks, to a particular URL. The result of that function is presented on the user's screen. This is done for all the HTML files to link them with the python code functionality. The directory, which is the templates folder storing the HTML files, has a specific name. The label, beginning with a trailing slash (/), given to the @app.route should be the same as the name of the mapping function. This will reflect on the web application when a particular role is prompted. If the label given to @app.route is '/create' and the function's name is 'create,' the URL will appear https://127.0.0.1/create.

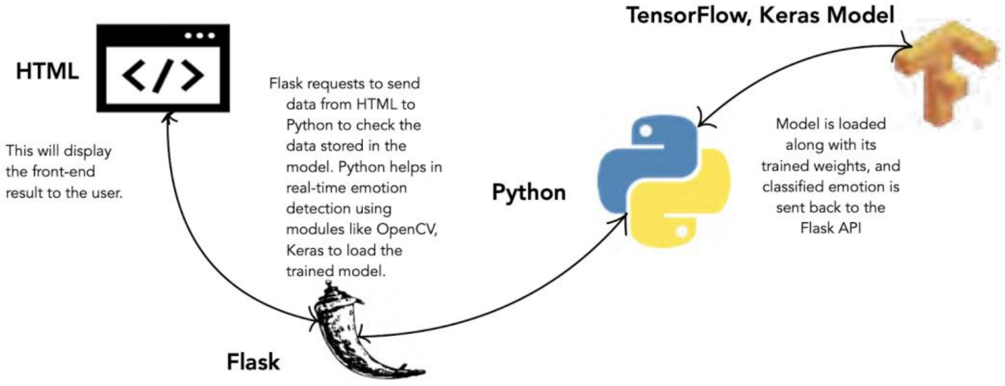

Figure 3: Proposed working of the Application

D. *Emotion Analysis*

Emotion analysis enables industries of all sizes to track how customers feel about their products and services. This ultimate guide dives into how the insights gained from understanding sentiment and emotion can deepen customer relationships and loyalty.

Emotion Analysis is a systematic process divided into four significant steps: Face Detection, Facial Feature Extraction, Analyzing the motion of Facial Features and finally, performing Emotion Analysis.

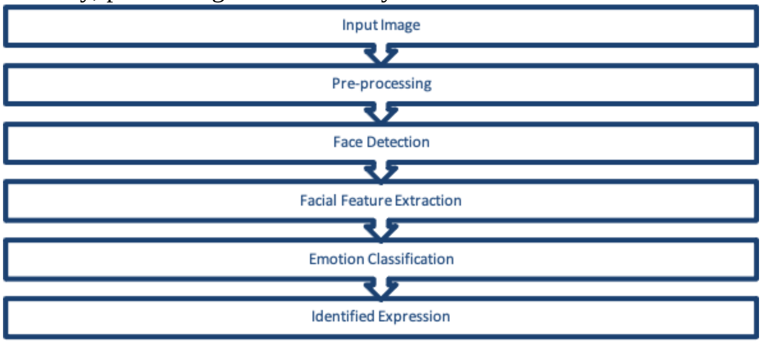

Figure 4: Project Structure

Step 1: Face Detection
The face is detected using the Viola-Jones face detection algorithm, an object-recognition framework that perceives image features in real-time. Despite being an outdated framework, Viola-Jones is quite prevailing, and its application has proven to be exceptionally remarkable in real-time face detection.

Step 2: Facial Feature Extraction
Extracting facial features from the perceived face region, i.e., detecting the structure of facial components or expressing the skin's texture in a facial area.

Step 3: Analyzing Facial Features Motion
Analyzing the movement of facial features or the variations in the appearance of facial features.

Step 4: Emotion Analysis
The software utilizes the Python module, OpenCV, for image processing, video analysis, camera collaboration, etc. It uses a Cascade classifier to classify faces in either real-time or stored videos. The Cascade Classifier works on the principle of the Viola-Jones Face Detection algorithm. The detected face is then matched with the stored weights, and finally, the emotion associated is depicted on the screen.

V. EXPERIMENTATION AND RESULT

The experiment was conducted over a sample of stored and real-time videos. The model was deployed and obtained the following results:

A. *Image Augmentation*

A good classifier requires loads of training datasets to accomplish a good outcome. Image Augmentation is an interaction to assist with artificial training images using various image processing methods. Flipping, pivoting, cropping, shading jittering, edge upgrade, and extravagant PCA are famous augmentation techniques. In this paper, cropping, pivot, shear, zoom, and flip have been utilized to improve the precision of the model.

B. *Model Implementation*

The emotion classification model has been written in the python programming language with Keras, TensorFlow, NumPy, PIL, OpenCV, and Matplotlib libraries. The Keras provides activation function, optimizers, layers, dropout, batch normalization, etc. The TensorFlow was used as a system backend to accept the inputs of a multidimensional array, which are the pixels of trained images. The OpenCV was used mainly to detect a face in the image or video streaming using Haar Cascade classifier, grayscale image conversion, and image normalization. The graphic user interface (GUI) was written in a python programming language to accept still images and real-time video streaming. The system then converts the input into a 48×48 grayscale image after Haar Cascade Classifier detects the face. After that, the cropped image has been passed into the proposed model to classify into seven distinct emotions.

C. *Overall Recognition Accuracy*

In the experiment, the overall precision and shortfall of the propositioned model against state-of-the-art, i.e., VGG-16 in the assignment of emotion classification based on the FER2013 dataset, has been performed. The optimized Convolutional Neural Network (CNN) with additional implementation of image augmentations, layer dropout, and normalization with less complexity of 5,786,247 parameters has proven to be more efficient equated to the state-of-the-art model with higher complexity of 5,790,727 parameters with the same dataset and numbers of training epochs by achieving a higher rate of model accuracy at 66.70% validation accuracy at 64.50% with the shorter training time at 50 epochs as shown below.

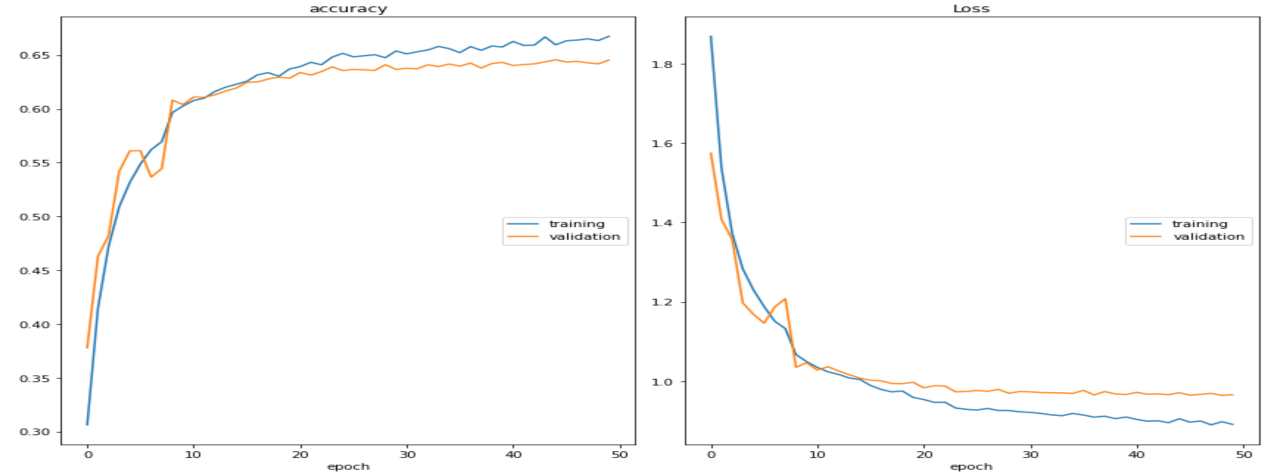

Figure 5: Model Accuracy and Model Loss of Proposed Model Used

Table 1: Model Loss and Accuracy Data

Data/ Metrics	Accuracy	Loss
Training	66.70%	89.10%
Validation	64.50%	96.60%

VI. CONCLUSION

The following figures depict emotion detected in real-time and stored videos (CCTV, news roll T.V.). Additionally, utilizing combined models is likewise a fascinating examination to improve model precision and diminish training time.

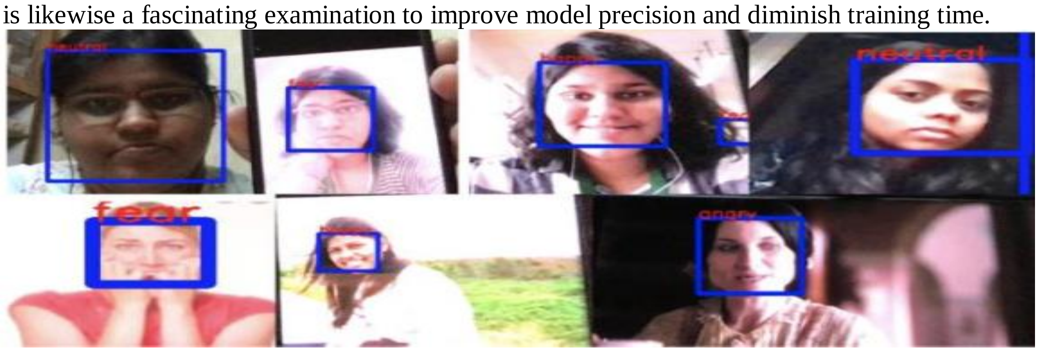

Figure 6: Real-Time Emotion Detection of Students

Figure 7: Emotion Detection over Stored Video. Presidential Debate 2020 (https://youtu.be/BIm_KpBshtA)

VII. DISCUSSION

The trained model can be deployed in Mobile-Based Applications using TensorFlow Lite and Firebase for identifying emotions on mobile devices through the usage of stored videos or by providing access to their device cameras to capture real-time emotions. One can improve the Face Recognition of the proposed project by deploying the face recognition module of Python instead of using the Viola-Jones Algorithm. The students can be assigned unique identification numbers using Face Recognition Module, and each student's emotion can be uniquely assessed and detected.

VIII. Acknowledgement

We state that the gist of this paper is entirely ours and does not show any virtuosity, as mentioned earlier. We would like to extend our sincerest gratitude to our mentor and project guide, Professor Vaishali Kavathekar, the Assistant Professor from Department of Information Technology, Don Bosco Institute of Technology, for her aid, assistance and priceless feedback.

IX. References

[1] International Board of Credentialing and Continuing Education Standards, "Impact of Anxiety and Depression on Student Academic Progress," International Board of Credentialing and Continuing Education Standards, [Online]. Available: https://ibcces.org/blog/2019/05/01/impact-anxiety-depression- student-progress/.

[2] H. K. S. S. F. T. F. M, "Mental Health Interventions in Schools in High-Income Countries," *Lancet Psychiatry,* vol. 1, no. 5, pp. 377- 387, 2014.

[3] Krithika. L.B. and Lakshmi Priya. G.G., "Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric," in *Procedia Computer Science*, 2016.

[4] T.S. Ashwin and R. M. R. Guddeti, "Unobtrusive Behavioral Analysis of Students in Classroom Environment Using Non-Verbal Cues," *IEEE Access*, vol. 7, pp. 150693-150709, 2019.

[5] Jianzhu Guo, Zhen Lei, Jun Wan, Egils Avots et al., "Dominant and Complementary Emotion Recognition from Still Images of Faces," *IEEE Access*, vol. 6, pp. 26391-26403, 2018.

[6] A. Sharma and V. Mansotra, "Deep Learning based Student Emotion Recognition from Facial Expressions in Classrooms," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 6, p. 2249 – 8958, 2019.

[7] Y. Tang, Q. Mao, H. Mia, H. Song and Y. Zhan, "An Emotion-Embedded Visual Attention Model for Dimensional Emotion Context Learning," *IEEE Access*, vol. 7, pp. 72457-72468, 2019.