Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 8, 136-144

Emotion Detection Of Students While Adopting E-Learning Approach

Sri Rama Murthy K[1] , Kamakshi Prasad V[2]

[1]Research Scholar, Department of CSE, Jawaharlal Nehru Technological University, Hyderabad, India, sriram.pavani@gmail.com
[2]Professor of CSE, College of Engineering, Jawaharlal Nehru Technological University , Hyderabad, India, kamakshiprasad@jntuh.ac.in

Abstract—Students in any type of learning have to utilize the time provided to them while teaching is going on. They shall concentrate and focus on the subject taught to them. This will increase the performance of the students. In this research paper, we are trying to identify the emotions of students, their behavior and analyze the diverted nature of the students while an online class is going on as a scope for the future. This study of the distracted nature of students can help the faculty members to keep track of such students and support those to improvise the performance. In order to study the diverted nature of students while online class is going on, one needs to have an idea of image processing, machine learning and deep learning algorithms.

Keywords—E-Learning, Mobile Learning, Emotion Detection, Distraction.

I. INTRODUCTION

With the massive evolution in the area of computer vision and image processing capabilities, artificial intelligence has many capabilities. For example, an artificial intelligence system can detect the movement of vehicles and the drivers can detect the drivers who are travelling without helmets on bikes and without a Seat belt in cars. Another example of an artificial intelligence system can be supported in the area of health by providing prescription support and vaccination advice. In the same way, the behavior of students in the online learning class can be analyzed, and it will be possible to conduct research in this area. Our research in this paper focuses on the emotion detection of students in the online class room while E-Learning and Mobile Learning is happening.

II. SOURCE OF THE DATA

A. About the Source of data

A Video that involves the activities of a student and a teacher can be a source of input for this research. A real-time video-based human activity recognition system can become an identical path in detecting the emotions of the students.

B. Algorithms in the area

There are several algorithms in artificial intelligence to detect the driver's activities while travelling. Here the source is a camera placed in front of the driver and an automatic alerting system that detect the emotions of the drivers and when a driver is in sleeping mode when his eyes get closed continuously for a few seconds, the alerting system alerts and save the vehicle and driver by automatically pausing the vehicle.[8]

III. CURRENT SCENARIO IN ONLINE EDUCATION

In the present day scenario, due to pandemic in various countries the students are suggested to attend online classes using any one of the gadgets. The teachers should monitor the student's attentiveness in the online classes.[4] As per the observation, there are various categories of teachers. The categories are 1. Some teachers not only teach, but they will also observe the behavior of students and keep on calling the students to make the students learn and they will keep them on their toes.[1] 2. Some teachers will teach, partially identify the students who are getting distracted and warn students, and continue teaching. 3. Some teachers will never bother whether the student is getting distracted or not and they continuously teach the classes. 4. Some teachers will create interest in students by continuously

creating zeal by imposing inquisitiveness among the students and allowing the students to participate in the activities and dragging the interest of students [5].

IV. Our Study of Students Distraction

Our study of the distraction of students helped us to carry on studying facial landmark detection, computing the change, rotation, and other facial landmarks in a video. There is splendid research work going on in the area of facial recognition, detection of landmarks on a face, emotion detection, driver drowsiness, etc,.[6] In our study, we noticed that students who are attending online classes are getting distracted due to various reasons. Distraction can be looking left or right, moving away from the class instead of sitting in front of the gadgets. If a learner instead of looking towards the gadget the movements of the neck and eyes left or right is too long or longer than a threshold, it is evidence of lack of interest and lacking concentration.

V. Why a Student Gets Distracted

There are chances for a student's inattentiveness due to sitting long hours in online classes that leads to frustration and may not show interest in listening to his or her class teacher. If the student is bright enough to learn by self-learning, there is no problem, but the problem comes with the performance of weaker students.

In an online classroom using zoom or Google meet, a learner will have some frequent set of observable attributes. Facial Landmark deviation will help in pinning down when there is a significant change in their attention level. [5] Example feeling heavy-eyed or yawning or having difficulty in concentration, etc. Emotional AI Algorithms will help to detect such behaviors. An alertness mechanism to the students or a log of the activities of the student's inattentiveness can help the teaching and student community. Our work proposes an Artificial Intelligent algorithm based on facial recognition, facial alignment, and divergent facial attribute analysis.

VI. Facial Emotions Detection

Our study in the research starts with Facial Detection. Object Detection is associated in computer science with various technologies such as computer vision and helps in identifying whether there are objects such as faces, buildings, trees, flowers, cars, etc.[7] The primary goal of face detection algorithms is to determine whether or not an image contains a face.

Face Detection is the basic step in face recognition, and it thus supports in finding a face or not, this is a subsystem and can provide assistance in a variety of areas, including security, biometrics, entertainment, and personal safety. Our main aim in this research is to identify the distraction of students in online classes. This needs Face Detection in the video frames.

Object Detection is a mix-up of technologies such as image processing and computer vision and these technologies interconnect with detecting instances such as an individual's face, building, flowers, etc. Face detection algorithms primary objective is to identify a face in the image or not.

VII. Methods in Facial Detection

There are various Face Detection methods as per our study. Yan, Kriegman, and Ahuja proposed a classification scheme for methods of face detection. These approaches are classified into four categories, and face detection algorithms may fall into two or more of these categories. These are the following categories: 1. Feature-based, Appearance-based, Knowledge-based, and Template-matching.

With the Feature-based classifier, the classifier is trained to identify the facial and non-facial regions. With the Knowledge-based classifier, the classifier should be trained based on rules. A face must have eyes, nose, and eyes, etc.

With the Template-matching classifier, the classifier uses templates that are pre-set or parameterized to detect faces by comparing the templates to the input photos. Once after detecting the face, it's the time to evenly catch the emotions on the face prior to the distraction of students in online class.

VIII. Emotion Detection

Facial expressions correspond to emotions; they are critical markers of human moods. The majority of the time (about 55% of the time), face expression is a nonverbal way of communicating emotion and may be used to detect whether or not someone is telling the truth. Improving learners' involvement with their educational programs is a crucial challenge in online learning. Learner involvement has been a hot topic in the education literature since the 1980s (Whitehill et al. 2014). Concerns about high drop-out rates in online courses may have sparked this interest (Rothkrantz 2016). It is well accepted that higher productivity and learning gain is connected to increased engagement and affect. According to some research, learner

involvement may be manipulated by pedagogical interventions, learning designs, and feedback. Detecting learners' engagement has become critical in online education in order to deliver tailored pedagogical support through intervention to online learners. Increased learner engagement benefits a variety of learning contexts, including traditional classrooms, instructional games, and intelligent tutoring systems. (Karumbaiah et al. 2017).

Different forms of learning engagements are useful to understand for designing individualized interventions to improve learners' experiences [3]. Learner engagement studies, on the other hand, require a method of measurement. Engagement theorists distinguish between two sorts of data: internal to the individual (cognitive and emotive) and external observable aspects (perceptible facial features, postures, speech, and actions). Measurement of engagement, according to several studies, necessitates combining observational data with data from within the individual.

IX. OUR METHODOLOGY

Convolution neural networks (CNNs) are the most widely used technique for image analysis. A CNN is distinguished from a multi-layer perceptron (MLP) by the presence of hidden layers referred to as convolution layers.[8] The suggested solution is predicated on a CNN framework with two levels. The first degree of abstraction proposed is backdrop removal, which is used to extract emotional content from an image. In this scenario, the primary expressional vectors are extracted using the conventional CNN network module (EV). The expressional vector (EV) is constructed by identifying significant face points. EV is closely tied to expression changes. The EV is calculated by applying a basic perceptron unit to a face image that has had the background removed. Additionally, in the suggested FERC model, a non-convolutional perceptron layer is included as the last step. Each convolutional layer accepts the input data (or picture), modifies it, and then passes it on to the subsequent level. This is a convolutional transformation fig 1. All of the convolutional layers employed are pattern-detection capabilities. Four filters were utilized inside each convolutional layer. Along with the face, the input picture sent to the first-part CNN (used to eliminate background) often contains forms, edges, textures, and objects. The edge detector, circle detector, and corner detector filters are used at the start of convolutional layer 1. The second CNN filter recognizes facial traits such as the eyes, ears, lips, nose, and cheeks after recognizing the face. This layer's edge detection filters. The second component of the CNN is composed of layers with a 33 kernel matrix, for example, [0.24, 0.18, 0.8; 0.79, 0.26, 0.83; 0.2, 0.28, 0.62]. These integers are originally chosen between 0 and 1. These values have been tuned for EV detection using the supervisory training dataset's ground truth. To optimize filter values in this case, we employed minimum error decoding. After tuning the filter using supervised learning, it is put in to the face with the background removed (i.e., the output image of the first-part CNN) in order to detect various facial components (e.g., eyes, lips, nose, ear, etc.).

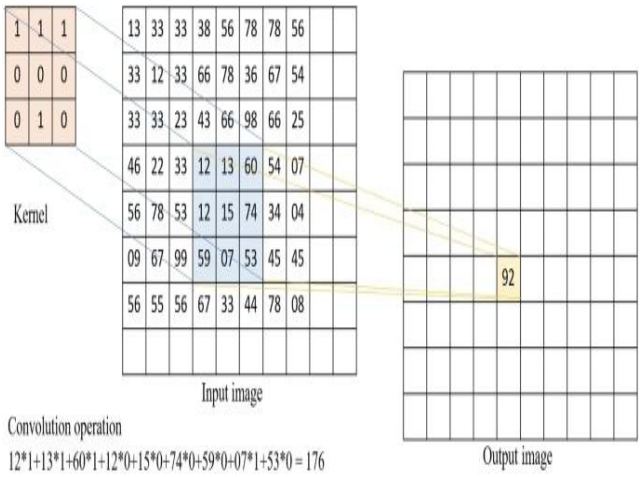

Fig 1: Operation of a convolution filters using the 3x3 kernel.

Each pixel in the input image and its eight neighbors are multiplied by the matching value in the kernel matrix, and the multiplied values are then added together to get the final output value.

X. KEYFRAME EXTRACTION FROM THE VIDEO

The network accepts both image and video inputs. When the FER model's input is video, the distinction between the corresponding frames is calculated. When the intra-frame difference is zero, the most stable frames occur. After applying a Canny edge detector to each of these steady frames, the aggregated number of white pixels was determined. After comparing the aggregated sums of all stable frames, the frame with the largest aggregated sum is picked since it has the most edge information (more edges more details). After then, this frame is chosen as a source of data for FERC. The rationale for selecting this image is that hazy photos have few or no edges.

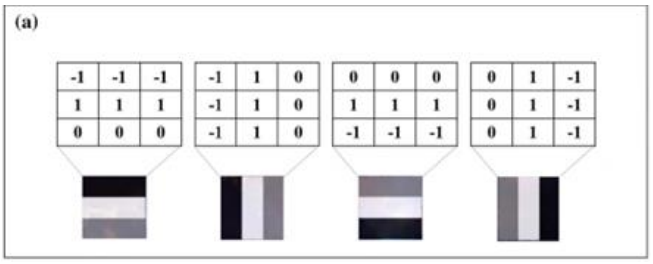

Fig 2: a. A vertical and horizontal edge detector filter matrix is used at layer 1 of the background removed CNN (first-part CNN).

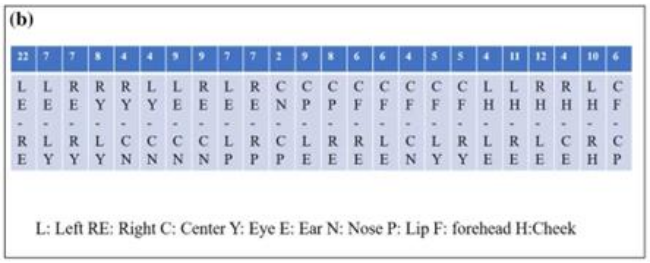

Fig 2: b. A representative EV matrix, including all 24 values in the pixel in the top row and the parameter measured in the bottom row.

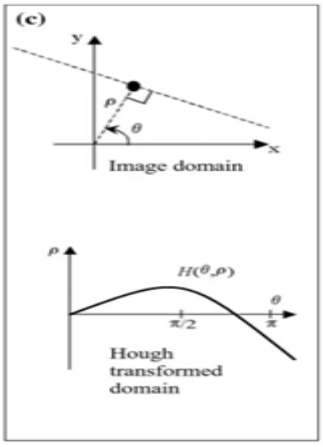

Fig 2: c. The Hough transform is used to convert a point in the Image domain (top panel) to the Hough transform domain (bottom panel).

XI. Removal of Background

After obtaining the input picture, Haar cascade features are used to extract human body components and positions from it. This output picture with identified features is a binary image that is utilized as the input for the first layer of CNN background removal (also referred to as the first-part CNN in this manuscript). This skin tone detection is image dependent. If the picture is one that has colors, the YCbCr color threshold might be utilized. The Y-value should be larger than 80, the Cb value should be between 85 and 140, and the Cr value should be between 135 and 200 for skin tone. The set of values given in the preceding paragraph was determined through trial and error and works for practically all of the accessible skin tones. We discovered that when an input image is monochrome, the skin tone recognition method performs really poorly. CNN also uses the circles-in-circle filter to enhance accuracy during background removal. This filter operation detects circles using Hough transform values. To ensure uniformity regardless of the kind of input picture, the Hough transform was always employed as the background removal CNN's second input feature. The Hough transform is calculated as shown in Eq. 1 shown below.

$$H(\theta,\rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x,y)\delta(\rho - x\cos\theta - y\sin\theta)\,dx\,dy$$

XII. Convolution Filter

As seen in images from results and research, for each convolution operation, the entire picture is partitioned into overlapping 3x3 matrices and then convolved with the matching 3x3 filters. The action of sliding and obtaining the dot product is referred to as 'convolution,' thus the term 'convolution filter.' During convolution, the dot product of both 3x3 matrixes is computed and placed in the appropriate spot, e.g., (1,1) at the output, as seen in Fig. 1. Once the entire output matrix has been calculated, it is transmitted to the next layer of the CNN for another loop of convolution. The last layer of the CNN used to extract face characteristics is a basic perception layer that optimizes the scale factor and exponent values in response to the deviation from the ground truth.

XIII. Details of Implementation Using Python

The project implementation is carried out by using Python3 in SpyderIDE (Operating System:Ubuntu 20.04 and Minimum 8gb RAM gives a better performance).

An IDE(Integrated Development Environment) simplifies and helps in developing new applications easily because multiple utilities don't need to be manually configured and integrated as part of the setup process. Spyder is an Integrated Developmenent Environment for python which is an open source and can be used across multiple platforms . It's available for Windows, Linux, and Mac OS thanks to Anaconda distribution has given an advantage of using it with different Operating systems . It has all the benefits of a comprehensive development tool with extra

competences of scientific packages. The IDE has many features, such as code completion i.e., helping users automatically complete familiar lines of code, syntax highlighting and style analysis.The feature 'go to definition' in Spyder helps users find component definitions.Spyder uses the PDB debugger which displays the line, file, and condition of each breakpoint.

As part of implementing the machine learning algorithm, firstly the machine should be capable of loading the dataset. Once the data gets loaded it should be cleaned and ready to get trained by implementing the algorithm. Then provide the test data to the algorithm and check the accuracies with various measures in machine learning algorithms. Primary consideration is all about basic measures such as precision, recall, accuracy, confusion matrix and then the Precision-Recall Curve, F score, ROC Curve(Receive Operating Characteristic) etc,.

Python is a high-level, interpreted, general purpose programming language with a high level capability which implements software as Multi-paradigm (Object-oriented, Procedural, functional, structured and reflective). It also accentuates code readability (Readability is a key factor in user experience and also is a software quality measure of how a piece of text is). The predefined methods and elements in various Packages of python helped us to implement and experiment the project such as Numpy, TensorFlow, opencv, cv2, matplotlib etc,.

NumPy is a package in Python Programming Language which basically helps in numerical and scientific computing. NumPy is a wallet of two words, coined by the blending of "Numerical" and "Python" which is very famous among data scientists and analysts, programming experts in machine learning, Artificial Intelligence, Hand Writing Recognition and Neural Networks etc. for its efficiency (run time speed) and the wide range of array operations.The project used Numpy for loading the numerical data, plot graphs of accuracy and loss in detecting the emotions.

Tensorflow is most popular framework of machine learning and deep learning. An open-source library which is released on 9 November 2015 and developed by Google Brain Team. Tensorflow is entirely based on Python programming language and used for numerical computation and data flow, which makes machine learning flexible and supports in building efficient machine learning algorithms. TensorFlow library integrates different API to create a large scale deep learning architecture like CNN (Convolutional Neural Network) or RNN (Recurrent Neural Network).

Tensorflow can train and run the deep neural networks for image recognition, handwritten digit classification, recurrent neural network, word embedding, natural language processing, video detection, and many more. Tensorflow run on multiple CPUs or GPUs and also mobile operating systems.

Keras, on the other hand, is a high-level API that runs on top of TensorFlow. Keras simplifies the implementation of complex neural networks with its easy to use framework. Using Keras, one can build a neural network model quickly and easily using minimal code, allowing for rapid prototyping. The keras modules particularly helped in building CNN model. A model will be trained and will get validated how good or bad on a particular data set, so as to improve accuracy score. These models group layers into objects. There are two types of models available in Keras. They are Sequential and Functional models. Keras Sequential Model is a model which is easy to use and help in adding various inputs to the functions in it. Sequential Model groups a linear stack of layers into a tf.keras.Model.

In the Keras Sequential Model, the data will flow from one layer to another layer. The flow of data is continued until the data reaches the final layer. Most of the Artificial Neural Networks, Convulutional Neural Networks and others use Sequential API Model.

Keras also provide Functional Model, which is an alternative model against the Sequential Model that enables to create models in a more complex manner which defines multiple inputs and multiple outputs. These models are sharing layers that help you to define direct acyclic graphs, multi layer output, sharing of the layer etc.

Conv2D, MaxPooling, Dropout, Flatten and Dense are several methods that acts as input to add method in Sequential model in Keras in Python which generates outputs.

The experiment carried out generated training and testing batches so that the model could be trained / validated on the test data. While implementing the code, the image size is defined to 48, so all images are reduced to 48 X 48. The training is conducted in the form of batches. Each batch is of the size of 64 units.

An epoch is a term used in machine learning which tells the number of passes of the entire training dataset the machine learning algorithm has completed. Datasets are generally grouped into batches (especially when the amount of data is huge). Some people use the term iteration loosely and refer to putting one batch through the model as iteration. In the experimentation, the batch size is of 64 units and number of epochs are

50. The total number of training images are around 28000.

Convolutions with OpenCV and Python is a way that helps to perform many operations on an image — an image is just a multi-dimensional matrix which has width (number of columns) and height (number of rows). Additionally images have depth to them which can be identified as number of channels in the image. A sampled RGB image contains a depth of 3 channels, each for RED, GREEN and BLUE respectively. An image is a big matrix and kernel or convolutional matrix is a tiny matrix that is used for blurring, sharpening, edge detection, and other image processing functions.Essentially, this tiny kernel sits on top of the big image applying a mathematical operation (i.e., a convolution) at each (x, y)-coordinate of the original image.

The most common type of convolution that is used is the 2D convolution layer and is usually condensed as conv2D. A filter or a kernel in a conv2D layer "slides" over the 2D input data, discharge element wise multiplication. As a result, it will be summing up the results into a single output pixel. The kernel will perform the same operation for every location it slides over, transforming a 2D matrix of features into a different 2D matrix of features.

Pooling layers provide an approach to down sampling feature maps by summarizing the presence of features in patches of the feature map. Two common pooling methods are average pooling and max pooling that summarizes the average presence of a feature and the most activated presence of a feature respectively. Maximum pooling, or max pooling, is a pooling operation that calculates the maximum, or largest, value in each patch of each feature map.

The results are down sampled or pooled as feature maps that highlight the most present feature in the patch, This has been found to work better in practice than average pooling for computer vision tasks like image classification. The image classification data while training helps in identifying the emotions while capturing the video.

Machine learning is classically used to predict outcomes from the given set of features. Therefore, anything we can do to widely applicable to the performance of our model is seen as a net gain. Dropout is a technique used to prevent a model from over fitting. Dropout works by randomly setting the outgoing edges of hidden units (neurons that make up hidden layers) to 0 at each update of the training phase. The Dropout helped to perform as mentioned above in the program implemented in python.

Flatten layers are used when you got a multidimensional output and you want to make it linear to pass it onto a Dense layer. The flatten method in keras or equivalent ravel method in numpy package is used for the purpose. An output from flatten layers is passed to a Machine Learning Program for classification or regression task.

The dense layer is a neural network layer that is connected deeply, which means each neuron in the dense layer receives input from all neurons of its previous layer. The dense layer is found to be the most commonly used layer in the models.In the background, the dense layer performs a matrix-vector multiplication. The values used in the matrix are actually parameters that can be trained and updated with the help of back propagation.

The output generated by the dense layer is an 'm' dimensional vector. Thus, dense layer is basically used for changing the dimensions of the vector. Dense layers also apply operations like rotation, scaling, translation on the vector. Implemented the Dense function also in identifying the emotions.

In order to train the model, the model is compiled with parameters such as loss, optimizer, and metrics with accuracy. A method named fit_generator of the Sequential model is invoked and several parameters such as number of iterators(epochs), paths, color mode which is gray scale, and class mode such as 'categorical', validation steps. These parameters will help in generating data of accuracy. This data will be helping in plotting the accuracy.

Artificial Neuron is the building block of artificial neural network. It is essentially a computational unit which performs the following steps – 1. It takes certain inputs and weights. 2. Applies dot product on respective inputs & weights and apply summation. 3. Apply some transformation using activation function on above summation. 4. Fires output.

In order to test the model, the camera will be turned on and captured and the video will be sent to the Cascade Classifier. The classifier instantly classifies and detects the emotions. The emotion detection will be displayed as shown in the results in the next section.

XIV. Results and Discussion

Our study is able to determine a person's emotional state based on their facial expressions. As illustrated in the

photos below, the machine learning technique is capable of recognizing a number of different emotions.

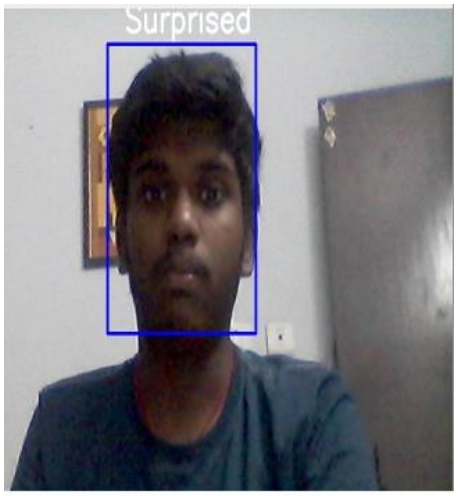

Fig 3: The system has discovered an image of a person who appears to be surprised.

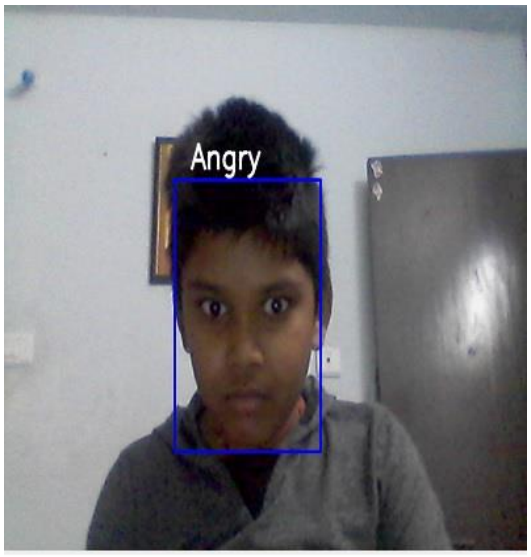

Fig 4: The system has discovered an image of a person who appears to be angry

Fig 5: The system has discovered an image of a person who appears to be fearful/ scared.

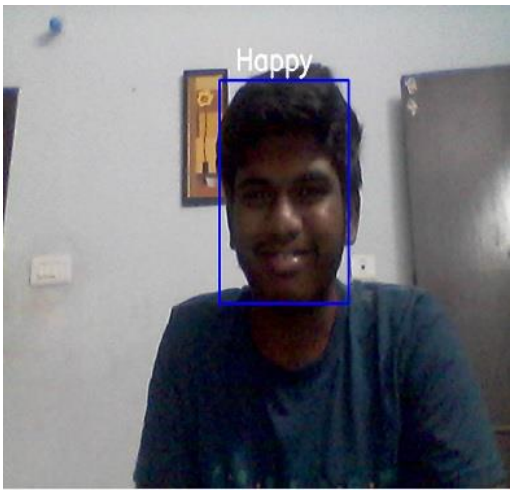

Fig 6: The system has discovered an image of a person who appears to be happy.

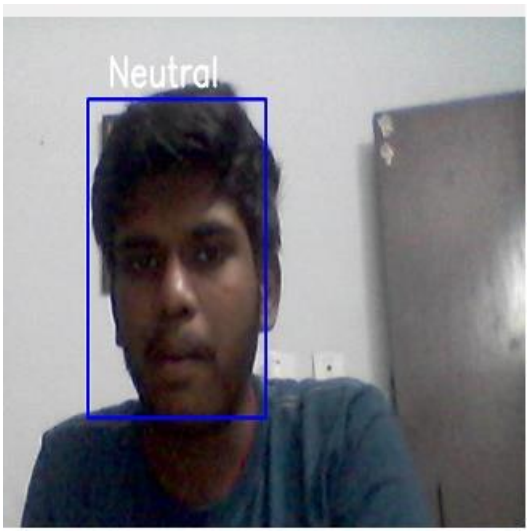

Fig 7: The system has discovered an image of a person who appears having no expression that is neutralitys on their face.

A model's accuracy is based on the correct predictions that are made for your model outcomes. Training files are bundled into a group, and then verified against algorithms to predict accuracy. 70% of the data set is used to generate the classifier, and 30% of the data set is compared to the classifier to measure how well the classifier is predicting the results. The percentages may vary, but accuracy is the number of correct predictions from all predictions made. Accuracy is calculated by using the number of correct predictions or all predictions made.

Our approach in identifying emotions resulted with the following accuracies.

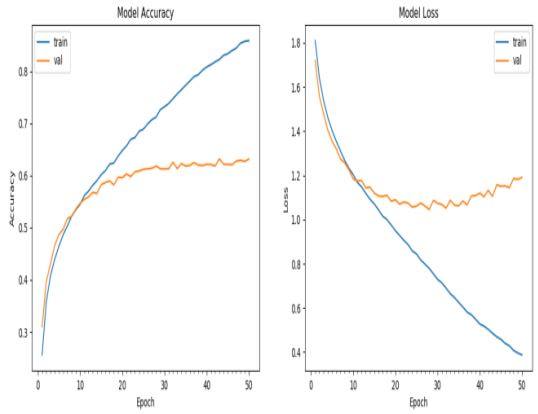

Fig 8 Image showing the accuracy and loss of the model built to the test the emotions.

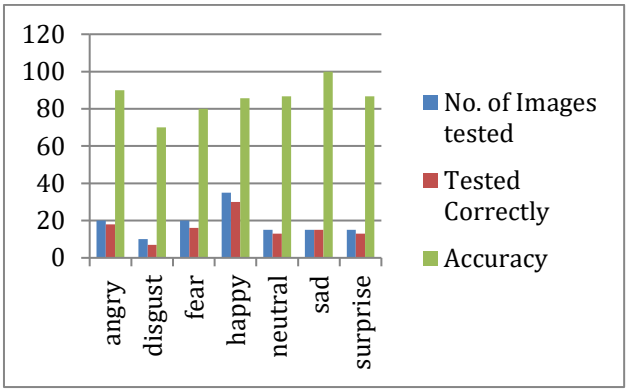

Fig 9 Image showing the accuracies of the emotions teste

XV. CONCLUSION

The carried out study is in its early phases which recognizes the emotions displayed on students' faces. Using this program, one may identify a wide range of emotions, including happiness, sorrow, neutrality, satisfaction, and fear among others. In the future, there is a wide scope of expansion to identify the feelings that students experience when they become distracted. Further the scope of research can be utilized to produce logs for teachers as to identify the distraction measure. These log records contain the student's id, the length of time the student was distracted, and the emotion the student was experiencing previous to the interruption.

These logs potentially may be used as a data source for future study. There is a wide scope to of research helping to identify and integration emotion detection features in e-class rooms.

ACKNOWLEDGEMENT

We thank and acknowledge Dr. Amjan Shaik, Head of the Department, Department of Computer Science Engineering, St. Peter's Engineering College, Mr.S.Veerabaraiah and Mrs.Amandeep Kaur, Faculty in the Department Computer Science & Engineering, St. Peter's Engineering College who provided insight and expertise that greatly assisted the research. We appreciate Pavan Srinivas Narayana, Arizona University, AZ, USA College for extending the support in identifying the solution for the problem.

We are very much thankful to all the authors in the reference list for making this research article in a brighter manner.

REFERENCES

[1] Tristan de Frondeville, "How to Keep Kids Engaged in Class," August 3, 2009, https://www.edutopia.org/classroom-student-participation-tips.

[2] Oya Celiktutan, Sezer Ulukaya & Bulent Sankur, "A comparative study of face landmarking techniques," 2013, Article number: 13 (2013).

[3] Martin, F. & Bolliger, D.U. (2018), Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. Online Learning 22(1), 205-222. doi:10.24059/olj.v22i1.1092

[4] Mehmet Ali Cicekci1 & Fatma Sadik2, "Teachers' and Students' Opinions About Students' Attention Problems During the Lesson", Journal of Education and Learning; Vol. 8, No. 6; 2019, ISSN 1927-5250 E-ISSN 1927-5269, Published by Canadian Center of Science and Education

[5] "Strengthening online learning when schools are closed: The role of families and teachers in supporting students during the COVID-19 crisis", "24 September 2020", https://read.oecd-ilibrary.org/view/?ref=136_136615-o13x4bkowa&title=Strengthening-online-learning-when-schools-are-closed.

[6] Feng You, Yunbo Gong, Haiqing Tu, Jianzhong Liang,and Haiwei Wang, "A Fatigue Driving Detection Algorithm Based on Facial Motion Information Entropy", Volume 2020, Article ID 8851485, https://doi.org/10.1155/2020/8851485.

[7] Chinthakindi Balaram Murthy, Mohammad Farukh Hashmi, Neeraj Dhanraj Bokde and Zong Woo Geem, "Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review," Article in Applied Sciences · May 2020

[8] Anil Kumar Biswall, Debabrata Singh, Binod Kumar Pattanayam, Debarata Samanta and Ming-Hour Yang, "IOT-Based Smart Alert System for Drowsy Driver Detection," Volume 2021, Article ID 6627217, https://doi.org/10.1155/2021/6627217

[9] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie & Laith Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data volume 8, Article number: 53 (2021).