AUTOMATED STUDENT ENGAGEMENT MONITORING DURING ONLINE CLASSES

Dr.S.VANITHA SIVAGAMI, B.MAHASIVAPRIYA, R.AISWERYAA
svanitha@mepcoeng.ac.in, mahasivapriya_cs@mepcoeng.ac.in, raisweryaa_cs@mepcoeng.ac.in
Mepco Schlenk Engineering College, Sivakasi.

KEYWORDS
Focus detection
Emotion recognition
Emotion weight
Concentration index
Engagement classification

ABSTRACT

E-learning has become a widely accepted mode of education system, especially in the current scenario of the COVID-19 pandemic. However, the lack of face to face interaction between students and instructors in e-learning environments can lead to reduced student engagement, which in turn can negatively impact learning outcomes. In our work, we propose deep learning methods to address this issue. Specifically, we develop a deep learning model that analyzes student engagement data and provides insights into how to improve student engagement. We experimentally evaluate the proposed model on a dataset of student engagement data from a e-learning platform and also collected dataset by our own which show that it can accurately predict student engagement levels. Our results suggest deep learning techniques can provide valuable insights into improving engagement level of students in online platforms which ultimately leading to better learning outcomes. Our project highlights the importance of student engagement in online learning and challenges associated with it. The project also describes the experimental evaluation of the proposed model on a real-world dataset and identifies factors that affect student engagement. The findings of the study suggest that deep learning techniques have the potential to improve engagement level of the students in online classes and effectively lead to better learning outcomes.

I. INTRODUCTION

In the olden days, humans couldn't live without basic needs which included water, air, shelter, food, and clothes. Nowadays, education plays an increasingly significant role in our lives. In terms of academics, they are now upgrading to virtual classes and examinations. The spread of covid-19 forces the education system to follow E-learning as a safety measure. This methodology had a serious impact on student's individual learning. It is possible for students to get distracted by several activities while attending online classes. Their emotions will vary with respect to time. Students cannot be attentive for the whole session. Teachers should monitor the activities of students while attending online classes. Due to this pandemic situation, it will be challenging for the system to maintain the same learning standards for students as well as for teachers. In order to offer individualized educational assistance to online students, it is crucial for online educators to accurately and effectively identify the level of engagement exhibited by their students. It is still a challenge for monitoring the student's engagement level during online classes. Therefore, we have presented this paper in order to track and monitor the student's engagement level throughout the classes. This will increase their dedication and engagement level by supporting the disengaged and nominally engaged students.

II. RELATED WORKS

A. Engagement and emotion detection

Dewan et al.[1] developed a deep learning system to detect students' engagement through their emotions. They trained their approach using a dataset from an online learning platform and utilized CNN and RNN to analyze student interactions. The authors found that their deep learning approach outperformed traditional machine learning techniques in detecting engagement levels of online learners. [2]S. Dash et.al proposed a two-stage algorithm for detecting student engagement levels in online learning environments. They developed a system that used machine learning models to analyze student activities and classified them into engaged and disengaged categories. The authors evaluated the performance of their model and found that it outperformed traditional machine learning techniques in detecting engagement levels of online learners. Buono p et al.[3] developed a system that analyzed students' facial expressions and head movements to detect their level of engagement during a lesson. Their method achieved high accuracy in detecting student engagement levels. [4] N Monkeresi et al investigated how computer vision techniques could identify student engagement during a structured writing exercise. They used a combination of facial expressions and heart rate to develop an automated system for detecting engagement in individuals. The system used a webcam and a remote photoplethysmography (rPPG) sensor to capture data, which was then processed and analyzed to classify engagement levels using a support vector machine (SVM). The study was conducted on a dataset of 120 videos of 40 participants and published by the authors. [5] Grafsgaard, Joseph et al aimed to create an automated system for recognizing facial indicators of frustration in individuals. The authors used machine learning techniques to analyze facial expressions and identified patterns of movement indicative of frustration. They used a dataset of videos of individuals performing a task designed to induce frustration and used OpenCV to analyze the videos and track facial landmarks. Decision trees and SVMs were used to classify facial expressions as indicative of frustration or not.[6] M Mushred et al aimed to analyze the usability of Convolutional Neural Networks (CNNs) was examined to identify student engagement in online learning. CNNs were used to create an automated system for detecting involvement in e-learning environments, which provided personalized feedback based on the student's engagement level. Video and audio data were analyzed to extract features related to the student's facial expressions, head movements, speech patterns, and vocal tone. The extracted features were used to train a model to classify the student's engagement level. The study used a dataset of videos of students engaged in e-learning activities. [7]A study was conducted to develop a framework for categorizing levels of student participation and classifying them using various classifiers. The proposed method was applied in a group meeting to detect the engagement levels of participants. Non-intrusive sensing technologies were used to capture audio, video, and physiological signals to determine the level of engagement. A multi-modal approach was employed to collect data from participants during meetings using wearable sensors, cameras, and microphones. Machine learning algorithms were then used to classify the data and determine each participant's level of engagement.[8]proposed framework with five main modules, namely detection, feature extraction, tracking, classification, and decision, was presented for real-time video stream analysis. The algorithm aimed to detect learner ROIs through segmentation and tracked and categorized them by extracting features. A decision module was then used to combine the classification scores to output a list of engagement levels. The authors conducted a literature review from 2000 to 2018 to identify engagement detection methods, data types, and performance metrics used in online learning environments. They aimed to identify research gaps and future directions in the field.[9] A research study was conducted to develop a deep learning model for recognizing fundamental facial expressions and engagement in students. They trained a convolutional neural network (CNN) on the FER-2013 dataset to create a rich facial representation model with state-of-the-art performance. A separate CNN was built and initialized to train an engagement recognition model on a new dataset. The goal was to build an automated system that provides real-time feedback to teachers and students about individual engagement levels. The authors used a dataset of videos of students engaged in classroom activities and used deep learning techniques to analyze the video data and extract features related to facial expressions. The extracted features were used to train a model to classify the student's engagement level, and the FER2013 dataset was utilized for facial expression analysis with labeled images of human faces showing different expressions..[10] A deep convolutional neural network was utilized to recognize facial expressions and analyze students' behavior during lectures. The emotions displayed by the students' faces were used to identify their level of engagement. The use of facial expressions enabled the examination of students' engagement. The study concluded that facial emotion recognition is an effective tool for identifying students' engagement during lectures. The deep convolutional neural network demonstrated its usefulness in recognizing and interpreting facial expressions. Gupta et al.[11]was proposed a real time learner engagement detection system that used deep learning models for emotion recognition in real-time. The system consisted of three components: facial expression detection, emotion recognition, and engagement level determination. The system was trained on a dataset of online learning sessions and was tested and evaluated using accuracy and F1-score metrics..[12]Hasnine M. N. proposed a system that used facial expression recognition and sentiment analysis techniques to extract emotions from students' faces and text-based inputs. The system analyzed the collected data using machine learning algorithms to identify specific emotions such as happiness, sadness, and boredom. Finally, the data was visualized in real-time through a web interface that provided engagement feedback to students and instructors..[13]. The proposed methodology by Sarthak Batra involved collecting a dataset of video screengrabs and their engagement scores. The data was preprocessed by resizing and normalizing the input images, and split into training, validation, and testing sets. The DMCNet architecture was trained using mean squared error and mean absolute error as loss functions, optimized with the Adam optimizer, and evaluated on the test set, outperforming several baseline models in terms of accuracy and robustness to noise..[14]Alruwais et.al. presented a methodology for detecting student engagement in an online class setting using machine learning techniques. The methodology involved two modules: feature extraction and classification. The algorithm processed video feed from a webcam to detect facial landmarks, extracted features, and used an SVM to predict student engagement.[15] Sarker et.al. developed a methodology to detect distracted students in virtual reality educational platforms by analyzing eye gaze data using a machine learning algorithm. Features extracted from the data included pupil diameter, fixation duration, saccade velocity, and fixation count. The SVM model achieved an accuracy of 82.5% in detecting distracted students, as evaluated using precision, recall, accuracy, and F1-score..[16] A study conducted by Walsh et al. surveyed 188 undergraduate students and held focus groups with 12 students to gather demographic information and measure their engagement with videos. Thematic analysis and descriptive statistics were used to analyze the data collected from video analytics, which revealed that intentional content positively impacted student engagement. Additionally, students valued concise, visually appealing, and organized videos.

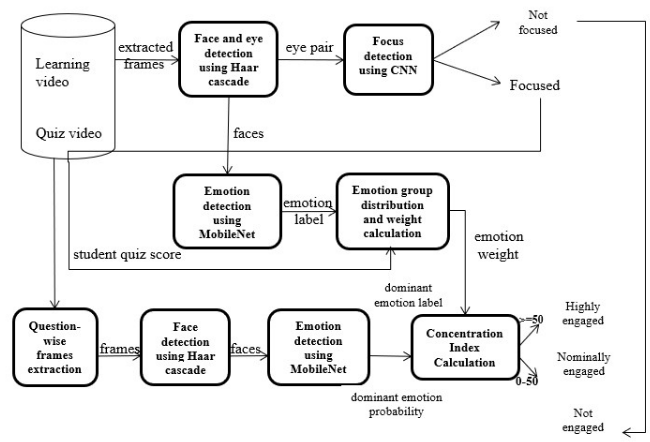

Figure.1. Architecture Diagram for Student Engagement Monitoring.

III. Proposed Work

In online learning situations, it is important to ensure that students are engaged and focused on the learning content. Haar Cascade algorithm, is a fast and efficient method for object detection. In this case, it is used to detect the frontal face of the student in the input image, and then locate the eye region within the face. This makes it an ideal approach for use in online learning environments where rapid feedback is required to ensure that students remain engaged in the learning content. The eye pairs are manually cropped from the detected image, and based on the segmented eye pair, the student is labeled as either "Focused" or "Not-Focused". If the student is labeled as "Focused", the emotions analysis is implemented. This involves using a deep learning model capable of identifying the dominant feeling conveyed by the student's facial expression during the session. The emotions that can be classified are Happy, Sad, Surprise, Angry, Disgust, Fear or Neutral. By analyzing the student's emotions, engagement detection can be achieved. For example, if a student is labeled as "Focused" and their facial expression is consistently happy, it can be inferred that the student is engaged and enjoying the learning content. On the other hand, if a student is labeled as "Focused" but consistently shows a sad or neutral facial expression, it may indicate that the student is not fully engaged or is having difficulty with the content. The use of facial expression analysis in conjunction with object detection techniques such as the Haar cascade classifier algorithm can provide a more comprehensive understanding of student engagement in online learning environments. This approach can assist educators in identifying potential areas of difficulty for individual students and adjust the teaching style accordingly to enhance their learning experience.

I. **Dataset collection**

For our work, we have used 3 dataset named FER 2013, Emo-Detect dataset and Own Dataset.

FER-2013 DATASET

It is collected from Kaggle which is especially famous for facial expression recognition. It covers about seven categories of emotions such as happy, neutral, sad, angry, disgust, fear, and surprise with emotion labels. Each image from every emotion which is shown in figure 2.

Figure.2. FER-2013 dataset

It includes train and test folders. These consists of 7 subfolders and each subfolder consists of approximately 1500 images. There are totally about 35,887 images. The labels for emotions consist of numerals and the labelled values are 0 for Angry, 1 for Disgust, 2 for Fear, 3 for Happy, 4 for Sad, 5 for Surprise, and 6 for Neutral.

EMO-DETECT DATASET

Emo-detect dataset has been collected which is used for emotion classification. It includes recordings of the students who were attending 10 programming questions. It contains 28 video clips in total. Each video takes different duration of time. There may be 45 minutes video or up-to 1 hour duration. It depends on the students who were undertaking the test. Sample video frame will be shown in figure 3.

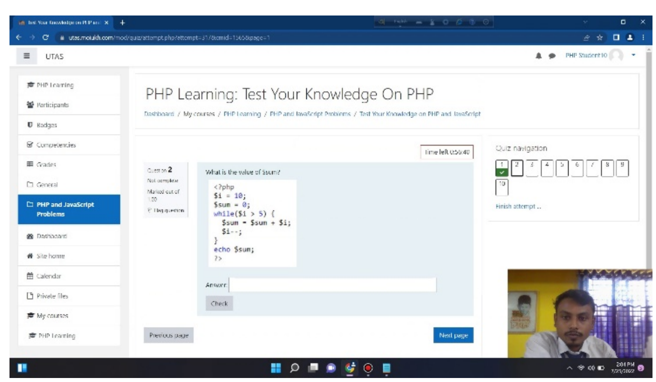

Figure 3.video frame of Emo-detect dataset.

OWN DATASET

A dataset of 20 students' recordings are collected. It contains video recordings of students' faces while listening informative video followed by attending quiz. 4a shows sample video frame while listening informative video and Figure 4b shows sample video frame while attending quiz. And some of the quiz questions are listed below.

1. which topic did you learned today?

   a.Stack b.Queue c.linked list d.Data Structures

2. Which Data structure follows LIFO?

   a.queue b.array c.stack d.Stack and Queue

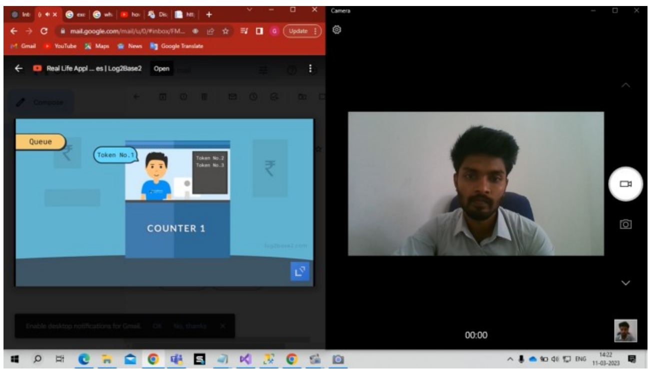

Figure. 4a Own Dataset while listening video

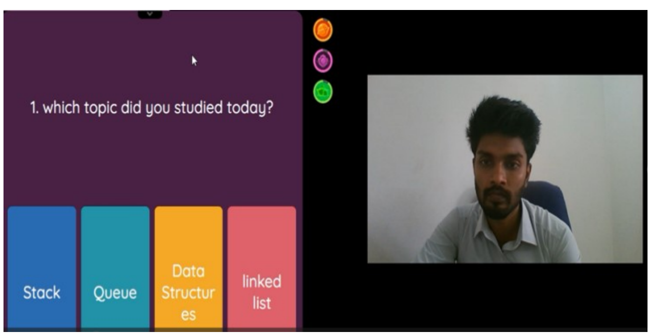

Figure. 4b Own Dataset while attending quiz

II. **Haar cascade algorithm**

Haar Cascade algorithm is a popular method for object detection that quickly and efficiently extract features from images. The diagrammatic representation of this algorithm is shown in the figure 5.Haar Cascade Classifier is a technique based on machine learning for detecting objects in images and videos, which includes face and eye detection. The classifier is trained using positive and negative sample images, where the positive samples depict faces or eyes, and the negative samples show random images without any faces or eyes. During the detection phase, the algorithm evaluates the input image by scanning each window, computing a feature vector for each window using Haar-like features, and comparing it to the trained classifier to detect the presence of a face or eye. Once the region of interest is identified, a rectangle is drawn around the detected face or eye in the image. We used this method to locate the eye region within the face and identify the student's frontal face in the image for the purposes of this work. From the extracted frames, faces and eyes are detected.

.

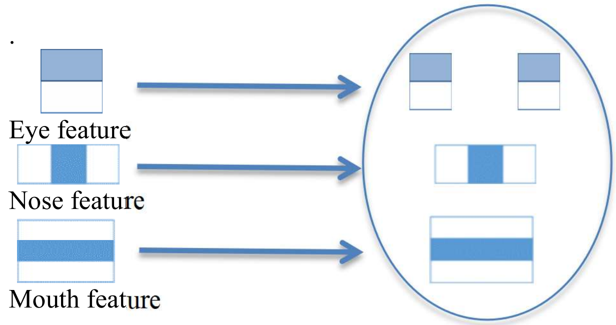

Figure 5. Structure of Haar cascade classifier

III. **Focus Detection using Convolution Neural Network(CNN)**

Convolutional Neural Networks (CNNs) are a widely used deep learning technique for analyzing images and identifying features within them. CNNs can differentiate and classify them into various categories, making them an efficient approach for processing images. CNNs have become popular in many fields due to their ability to extract important information from images and accurately classify them which is shown in table 1. In this study, eye images were obtained using the Haar Cascade model on facial images from the FER-2013 dataset. These images were then labeled by a group of annotators as either "Focused" or "Not-Focused," based on the following criteria.

- ✓ Focused -"looking towards or around the screen".
- ✓ Not-Focused-"eyes closed", "looking away from screen".

The network used for this task was also trained on FER-2013 dataset.

Table 1. CNN architecture

Layer (type)	Input Size	Filter Shape	Output Shape
Input layer	**(64, 64, 3)**	**N/A**	**(None, 64, 64, 3)**
Conv2D	**(64, 64, 3)**	**(3, 3, 3, 32)**	**(None, 62, 62, 32)**
Activation_1 (ReLU)	**(None, 62, 62, 32)**	**N/A**	**(None, 62, 62, 32)**
Conv2D	**(None, 62, 62, 32)**	**(3, 3, 32, 64)**	**(None, 60, 60, 64)**
Activation_2 (ReLU)	**(None, 60, 60, 64)**	**N/A**	**(None, 60, 60, 64)**
MaxPooling2D	**(None, 60, 60, 64)**	**N/A**	**(None, 30, 30, 64)**

Flatten	**(None, 30, 30, 64)**	**N/A**	**(None, 57600)**
Dense	**(None, 57600)**	**N/A**	**(None, 2)**

•The input layer is designed to hold the raw pixel values of an image in a 64x64 format.

•The convolutional layers utilize a set of 3x3 filters to calculate the output of neurons that are connected to specific regions in the image.

•A pooling layer with a 2x2 size is used to decrease the spatial size of the data representation.

•Finally, a fully connected layer is implemented to compute the scores for the two classes.

•Binary classification task was carried out to distinguish between two categories: 'Focused' or 'Not engaged'.

Fig 6a and fig 6b shows the results of focus detection for Emodetect dataset. Student images with focus labelling is done by using CNN. Fig 6c and 6d shows the result of focus detection using CNN for our own dataset.

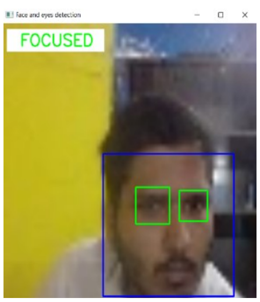

Fig 6a Focused

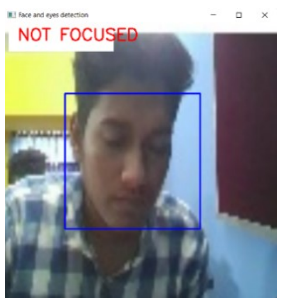

Fig 6b Not-Focused

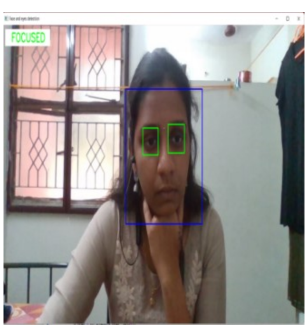

Fig 6c Focused(Own Dataset)

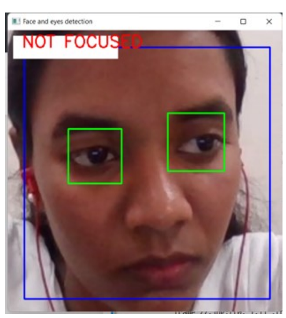

Fig 6d NotFocused

IV. Emotion Detection Using Mobilenet

A well-known convolutional neural network (CNN) architecture called MobileNet was created to be compact and effective while still achieving high accuracy on image categorization tasks. In order to do this, it uses depthwise separable convolutions, which are depthwise convolutions that apply a single filter to each input channel before switching to pointwise convolutions. (which applies a 1x1 filter to combine the outputs of the depthwise convolution). We used the FER-2013 dataset for this study, which consists of 35,887 48x48 pixel images that have each been assigned one of seven facial expressions: anger, disgust, fear, happiness, sorrow, surprise, and neutral. By starting our model with a MobileNet architecture that has already been trained and optimized on the FER-2013 dataset, we applied transfer learning. We utilized the Adam model and a categorical cross-entropy loss function.

Table 2 Emotion weight calculation for Emodetect dataset

Dominant Emotion	Emotion Weight
Anger	0.5
Disgust	-
Fear	0.12
Happy	0.68
Neutral	0.75
Sad	0.628
Surprise	-

Table 3 Emotion Weight Calculation for own Dataset

Dominant Emotion	Emotion Weight
Anger	0.5
Disgust	-
Fear	0.8
Happy	0.68
Neutral	0.75
Sad	0.628
Surprise	-

Model with a 64x64 input layer, convolutional layers with 3x3 filters, and 2x2 pooling layers. Instead, it is using the MobileNet architecture as a base model and adding some layers on top of it to fine-tune it for a specific task.

Table 4 MobileNet Architecture

Type/Stride	Filter Shape	Input Size
Convolution	3x3x3x32	224x224x3
Conv depthwise	3x3x32 dw	112x112x32
Convolution	1x1x32x64	112x112x32
Conv depthwise	3x3x64 dw	112x112x64
Convolution	1x1x64x128	56x56x64
Conv depthwise	3x3x128 dw	56x56x128
Convolution	1x1x128x128	56x56x128
Conv depthwise	3x3x128 dw	56x56x128
Convolution	1x1x128x256	28x28x128
Conv depthwise	3x3x256 dw	28x28x256
mConvolution	1x1x256x256	28x28x256
Conv depthwise	3x3x256 dw	28x28x256
Convolution	1x1x256x512	14x14x256
5 x (conv depthwise) (convolution)	3x3x512 dw 1x1x512x512	14x14x512 14x14x512
Conv depthwise	3x3x512 dw	14x14x512

Convolution	1x1x512x1024	7x7x512
Conv depthwise	3x3x1024 dw	7x7x1024
Convolution	1x1x1024x1024	7x7x1024
Average pooling	Pool 7x7	7x7x10241
Fully Connected	1024x1000	1x1x1024
Softmax	Classifier	1x1x1000

In order to predict the emotion that is contained in the input data, the output layer of the network classifies it into seven categories: angry, disgusted, fear, happy sad, surprised, and neutral.

V. **Emotion Group Distribution and Weight Calculation**

The value used to represent how much a particular emotional state affects a student's focus at that moment is known as the "emotion weight." Its value is between 0 and 1. Frames are extracted, faces are identified using a Haar-cascade classifier, and students are categorized according to the main emotion they displayed in order to obtain the weights related to each emotion.

For example, a student was classified as belonging to the happy emotion group if they expressed a happy expression for more than 50% of the duration of the film.
The formula used to determine the mean score each group received on a quiz,

$$\textbf{Emotion weight, } EW_g \textbf{= Total score / N}$$

where ,

EW_g -> Emotion weight belonging to a group

Total score -> The sum of all the scores achieved by the students in the quiz.

N -> The total number of students in that particular group who took the quiz.

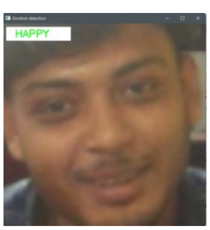

Fig 8a HAPPY

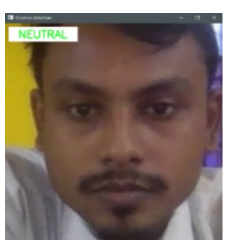

Fig 8b NEUTRAL

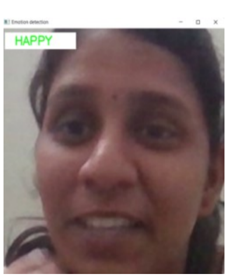

Fig 8c HAPPY

Fig 8d NEUTRAL

Face images with emotion label is predicted using mobilenet. Multiple emotions can also be expressed by a student. Fig 8a, Fig 8b shows the results of emotion detection for Emodetect datatset. Fig 8c, Fig 8d shows the results of emotion detection for our own dataset using mobilenet.

Emotion weight is calculated based on the mean score achieved in the quiz. Students are grouped according to their dominant emotion which is expressed over the time. Table 2 shows the dominant emotion and corresponding weight for the grouped students in Emodetect dataset.

Table 3 shows the dominant emotion and its emotion weight for the students who are grouped based on their dominant emotion label for our own dataset.

VI. **Concentration Index Calculation**

A Dominant Emotion Probability (DEP) score is calculated after facial emotion data has been classified using a Mobilenet. There are seven different categories of emotions that people can experience: neutral, happy, surprised, sad, disgusted, angry, and fearful.

As stated in equation 1, the DEP value is multiplied by the relevant Emotion Weight (EW) that corresponds with table I to determine the Concentration Index (CI):

CI = DEP × EW ------------------------(1)

where, DEP is the dominant emotion that the student expressed over a time. EW is the weight of the emotion calculated from the mean score achieved.

VII. **Engagement Classification**

Based on the index value, the estimated CI will be broken down to three levels.

When a student's concentration index value from their facial emotion is between 50% and 100% and they remain attentive, they fall under the category of highly engaged.

When a student is focused and the concentration index value from the facial emotion is lower than 50%, they are considered to be nominally engaged.

Not engaged: When a student is not focused, they fall under the not engaged category.

Table 5 presents the quiz's overall results as well as the student and own dataset questions' partial concentration indices.

S_id	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	Total CI
s1	67.95	67.49	67.14	67.98	67.96	67.96	67.83	67.68	67.22	67.8	67.7
s2	67.8	67.9	67.9	67.9	67.9	67.5	67.9	67.9	67.9	67.9	67.8
s3	37.5	52.8	35.2	65.2	55.4	57.9	61	9.1	59.5	56.7	49
s4	67.3	8.4	67.9	50.1	10.9	9.4	10.4	10.1	7.3	56.8	29.8
s7	74.5	73.8	73.2	71.3	66.6	72	72.6	73.9	69.3	70.8	71.8
s11	11.36	47.84	10.07	10.46	10.94	10.91	11.23	57.62	60.09	58.69	28.92
s12	67.7	65.9	67.9	67.1	67.4	67.9	67.1	67.2	67.6	67.9	66.8
s15	67.7	67.9	67.8	67.8	67.8	67.9	67.8	67.9	67.9	67.9	67.8
s16	67.25	67.87	64.82	67.47	66.8	67.69	67.85	67.83	67.25	67.57	67.24
s17	9.08	9.61	9.94	9.47	11.13	64.67	10.06	55.55	62.15	62.62	30.33
s18	61.8	60.3	62.2	60.4	60.2	62.3	60.9	58.3	62.1	62.1	61.1
s20	62.72	66.59	61.72	62.18	61.03	59.52	61.59	62.28	62.57	62.63	62.38

Table 6 presents the overall quiz results, student partial concentration indices, and quiz questions for the Emodetect Dataset.

s_id	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	Total CI
S2	34.2	22.8	36.3	16	17.5	10	28.5	26.7	22	14.5	22.9
S3	16.2	16.2	16.2	16.2	16.2	16.2	16.2	26.9	16.2	162	17.2
S4	14.6	27.7	27.6	23.3	24.9	25.8	25.8	23.3	14.6	23.3	23.1
S5	20.5	23.8	23.8	12.7	23.8	23.8	20.2	13.7	25.2	20.5	20.8
S6	34.9	14.9	12.9	27.2	12.9	15.5	15.5	14.9	34.5	12.5	19.6
S7	25.2	15	35.3	34.9	32	36.2	37.2	36.8	37.9	35.3	31.6
S8	34.8	34.8	34.8	34.8	22.6	34.8	34.8	34.8	34.8	34.8	33.5
S9	29.8	27	34.2	34.8	39.8	32.5	39.8	34.2	36.8	33.4	31.2
S11	34.8	34.8	34.2	36.3	21.7	35	34.2	34	36.3	36	33.7
S12	46.8	46	46.6	46.6	46.8	45.8	16.4	46.8	46	8.4	39.6
S13	16.2	16.2	17.5	16.3	18.2	43.3	22.2	16	16.5	16.4	19.9
S14	17.8	18.4	15	18.3	17	16.5	36.7	17.8	14.7	23	19.5
S15	27.6	27.6	11.4	27.6	27.6	27.6	27.6	27.6	27.6	27.6	25.9
S16	23.6	22.2	45.2	25.1	41.4	23.7	25	25.9	45.5	21.7	29.9
S17	31	34.6	36.6	34.6	36.1	35.8	36.9	37.2	34.3	33.9	35.1
S18	21.7	17.2	17.9	18.5	17.5	17.1	27	28.5	28.1	28.1	19.2
S19	37.3	36.9	37.2	46.2	37.1	37	37.1	27	21.4	26.8	35.4
S20	18.6	18.6	18.6	18.2	19.1	17.7	18	19.2	24.5	37	20.9
S22	17.8	26.6	14.4	14	17.8	14	17	18	19.1	17.8	17.6
S23	26.6	26.6	26.6	26.6	15.2	27.5	26.6	26.6	26.6	27.5	25.7
S24	24.7	20	19.4	20	19.4	19.4	17.8	19.4	20	24.7	20.5
S26	24.5	33.6	18.5	42.3	46.9	36.6	46.8	13.5	44	35.2	36.2
S27	14.3	36.5	14.3	14.3	16	36.5	36.5	36.5	36.5	36.5	27.8
S28	36.8	34.6	34.7	20.1	36.8	36.8	34.6	36.8	33.9	36.8	34.2

Figure 9 Engagement levels(Emodetect)

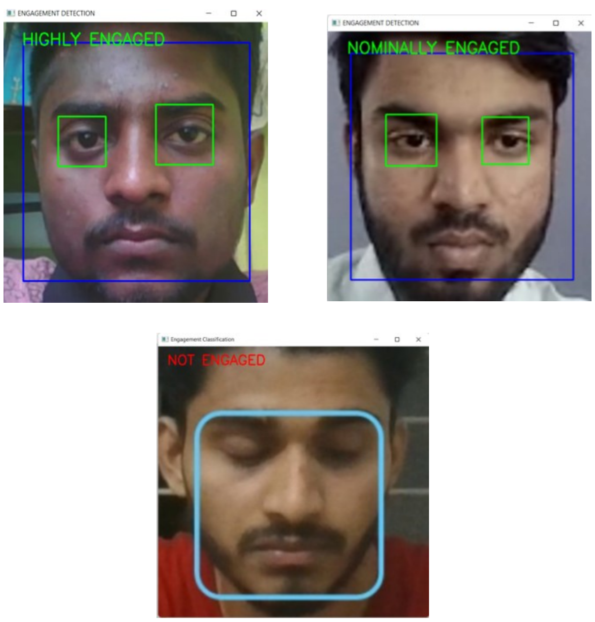

Figure 10 Engagement levels(Own dataset)

Experimental results and Discussion

<u>Emodetect Dataset:</u>

We have tested with 25 students among them **12** of them were fall under **Focused** category and **13** of them were **Not-focused.** Table 6 shows the result of concentration indexes, for each student and questions. Fig 7e, Fig 9 shows the result of engagement level.

<u>Own Dataset:</u>

We have tested with 21 students among them **12** of them were fall under **Focused** category and **9** of them were **Not-focused.** Table 5 shows the result of concentration indexes, for each student and questions. Fig 10 shows the result of engagement level.

In this work, for focus detection using CNN, we have achieved 70% of accuracy. For emotion detection using mobilenet, we have obtained 96% of accuracy.

REFERENCES

[1]M. A. A. Dewan, F. Lin, D. Wen, M. Murshed and Z. Uddin, "A Deep Learning Approach to Detecting Engagement of Online Learners," 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, 2018, pp. 1895-1902, doi: 10.1109/SmartWorld.2018.00318

[2]S. Dash, M. A. Akber Dewan, M. Murshed, F. Lin, M. Abdullah-Al-Wadud and A. Das, "A Two-Stage Algorithm for Engagement Detection in Online Learning," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 2019, pp. 1-4, doi: 10.1109/STI47673.2019.9068054.

[3]Buono, P., De Carolis, B., D'Errico, F. et al. Assessing student engagement from facial behavior in on-line learning. Multimed Tools Appl (2022). https://doi.org/10.1007/s11042-022-14048-8

[4]H. Monkaresi, N. Bosch, R. A. Calvo and S. K. D'Mello, "Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate," in IEEE Transactions on Affective Computing, vol. 8, no. 1, pp. 15-28, 1 Jan.-March 2017, doi: 10.1109/TAFFC.2016.2515084.

[5]Grafsgaard, Joseph & Wiggins, Joseph & Boyer, Kristy & Wiebe, Eric & Lester, James. (2013). Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis. Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013. 159-165. 10.1109/ACII.2013.33.

[6]M. Murshed, M. A. A. Dewan, F. Lin and D. Wen, "Engagement Detection in e-Learning Environments using Convolutional Neural Networks," 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, 2019, pp. 80-86,doi:10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00028.

[7]Frank, Maria & Tofighi, Ghassem & Gu, Haisong & Fruchter, Renate. (2016). Engagement Detection in Meetings.

[8]Dewan, M. & Murshed, Mahbub & Lin, Fuhua. (2019). Engagement detection in online learning: a review. Smart Learning Environments. 6. 10.1186/s40561-018-0080-z.

[9]Nezami, O. M., Dras, M., Hamey, L., Richards, D., Wan, S., & Paris, C. (2018). Automatic Recognition of Student Engagement using Deep Learning and Facial Expression. arXiv. https://doi.org/10.48550/arXiv.1808.02324

[10]Gagana S, Sheba Selvam,Priya G, Preethi H, Seema D (2021) Student Behaviour Detection in Eductation Training Institution. Volume 12,Issue 7 July2021:11708 –11719

[11]Gupta, S., Kumar, P. & Tekchandani, R.K. Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. Multimed Tools Appl (2022). https://doi.org/10.1007/s11042-022-13558-9

[12]Hasnine, M. N. (2021). Students' emotion extraction and visualization for engagement detection in online learning. International Journal of Human-Computer Interaction, 37(7), 663-678. https://doi.org/10.1080/10447318.2020.1832449

[13]Batra, S. (2019). DMCNet: Diversified Model Combination Network for Understanding Engagement from Video Screengrabs. In 2019 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1030-1035). IEEE. https://doi.org/10.1109/ICME.2019.00208

[14]Alruwais, N. (2018). Student-Engagement Detection in Classroom Using Machine Learning Algorithm. In 2018 4th International Conference on Information Management (ICIM) (pp.1317).IEEE.https://doi.org/10.1109/INFOMAN.2018.8462996

[15]Sarker, R. (2022). Detecting distracted students in educational VR environments using machine learning on eye gaze data. Journal of Educational Computing Research, 60(5), 799-819. https://doi.org/10.1177/07356331211072908

[16]Walsh, J. N. (2014). Investigating student engagement with intentional content: An exploratory study of instructional videos. Journal of Computing in Higher Education, 26(1), 39-60. https://doi.org/10.1007/s12528-014-9079-z