AUTOMATED STUDENT ENGAGEMENT MONITORING DURING ONLINE CLASSES

A PROJECT REPORT

Submitted by

AISWERYAA R (Reg. No. 201904007)
MAHASIVAPRIYA B (Reg. No. 201904083)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MEPCO SCHLENK ENGINEERING COLLEGE, SIVAKASI

(An Autonomous Institution affiliated to Anna University Chennai)

April 2023

BONAFIDE CERTIFICATE

Certified that this project report titled **"AUTOMATED STUDENT ENGAGEMENT MONITORING DURING ONLINE CLASSES"** is the bonafide work of **Ms. R. AISWERYAA (Reg. No.: 201904007), Ms. B. MAHASIVAPRIYA (Reg. No.: 201904083)** who carried out the research under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Internal Guide	Head of the Department
Dr.S.VANITHA SIVAGAMI, M.E.,Ph.D., Associate Professor(Senior Grade), Dept. of Computer Science and Engg, Mepco Schlenk Engg College (Autonomous) Sivakasi.	**Dr. J. RAJA SEKAR,** M.E., Ph.D., Professor and Head Dept. of Computer Science and Engg. Mepco Schlenk Engg College (Autonomous) Sivakasi.

Submitted for Viva-Voce Examination held at **Mepco Schlenk Engineering College (Autonomous), Sivakasi** on ________/ / 20__ .

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ABSTRACT

E-learning has become a widely accepted mode of education, 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 this project, deep learning techniques are used to address this issue. Specifically, a deep learning model is developed that analyzes student engagement data and provides insights into how to improve student engagement. The proposed model is evaluated on a dataset of student engagement data from a e-learning platform and also Auto-Engage-Detect dataset collected which show that it can accurately predict student engagement levels. This project highlights the importance of student engagement in e-learning environments and the 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 student engagement in e-learning environments and ultimately lead to better learning outcomes.

ACKNOWLEDGEMENT

First and foremost, we thank the Almighty for his abundant blessings that is showered upon our past, present, future successful endeavors.

At the appellation, we would like to extend our sincere gratitude to our reverent principal, **Dr. S. Arivazhagan**, M.E., Ph.D., for providing sufficient working environment such as system and library facilities. We also thank him very much for providing us with adequate lab facilities, which enable us to complete our project.

We are grateful to our charismatic Head of the Computer Science Engineering and Department **Dr. J. Raja Sekar,** M.E., Ph.D., for giving us this golden opportunity to undertake a project of this nature and for his most valuable guidance given at every phase of our work.

We would like to extend our gratitude to **Dr. S. Vanitha Sivagami**, M.E., Ph.D., Associate professor, Department of Computer Science and Engineering, Mepco Schlenk Engineering College for being our project guide and directing us through our project.

We extend our sincere thanks to our project coordinator, **Dr. M. S. Bhuvaneswari,** M.E., Ph.D., Associate Professor for his encouragement throughout the project and helping us schedule our works.

We would like to thank our lab technicians for their tremendous help and providing us with the necessary components required for making this project.

We are very grateful to our beloved parents and friends who afforded the necessary help for us at the right time for making our project a grand success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS		EXPANSION
CNN	-	**C**onvolutional **N**eural **N**etwork
DNN	-	**D**eep **N**eural **N**etwork
DBN	-	**D**eep **B**elief **N**etwork
LSTM	-	**L**ong **S**hort-**T**erm **M**emory
ROI	-	**R**egion **O**f **I**nterest
LDP	-	**L**ocal **D**irectional **P**attern
NiN	-	**N**etwork **i**n **N**etwork
CI	-	**C**oncentration **I**ndex

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

In this chapter, the overall description of the project, objectives, outcome and organization of the project are discussed.

1.2 PROBLEM DESCRIPTION

In modern scenario, E-learning platforms come-up with students' disengagement problem. Some of the reasons include lack of student-teacher relationship and several distractions in learning environment. It becomes more challenging for teachers to monitor students' engagement and maintain the right-level of interactions. While attending online classes, student can't be in same emotion throughout the sessions. It seems problematic to monitor the engagement of students.

The main purpose of this project is to detect the focus, emotion and engagement level of the student using deep learning algorithms. Convolutional Neural Network (CNN) is used for focus detection using face of the students. Binary classification task is carried out to distinguish between two categories: 'Focused' or 'Not engaged'. Haar cascade classifier algorithm is used to detect face and eyes from the given input images. Mobilenet is used to predict the emotion of the student using the labels as 0- angry, 1- disgust, 2- fear, 3- happy, 4- neutral, 5- sad, 6- surprise. Engagement detection is done in order to find the level of engagement of the student. It is implemented based on grouping the students according to their emotions and concentration index calculation.

1.3 OBJECTIVE

The main objective of our project is:

- To track the students' engagement based on eye focus detection and emotion detection.

1.4 OUTCOMES

The outcome of our proposed work is to

- Detect eye focus using CNN
- Track student engagement by relating student emotion and quiz score.

1.5 ORGANIZATION OF THE PROJECT REPORT

Our project “Automated Student Engagement Monitoring during Online Classes” detects student engagement levels in real-time, based on various factors such as the student's behaviour, interaction with the learning content, and their academic performance. The model will be trained on large datasets of student engagement data and will use various deep learning techniques such as convolutional neural networks (CNNs) and MobileNet to analyze the data.

This chapter deals with the importance of detecting the engagement level of the student in E-Learning environments using the video recordings.

Chapter 2 deals with the techniques used in focus detection and emotion detection.

Chapter 3 deals with detailed description of the proposed system.

Chapter 4 deals with the methods and modules involved in the design of the proposed system.

Chapter 5 deals with the implementation methods used in various methods of the proposed system.

Chapter 6 deals with the results obtained from the proposed system and discussions on the obtained results.

Chapter 7 deals with the conclusions of the work done and the future enhancements for the system.

1.6 SUMMARY

This chapter gives the brief description about the purpose, objectives, outcomes and organization of the report. In the next chapter we will be discussing about the detailed description on literature survey of various existing techniques.

CHAPTER 2

LITERATURE SURVEY

2.1 OVERVIEW

This chapter deals with the literature survey of various existing techniques. At the end of this chapter, the drawbacks identified in the existing methodologies and gap analysis is done. The next chapter describes the proposed methodology used to overcome the drawbacks.

2.2 A Deep Learning Approach to Detecting Engagement of Online Learners

M. A. A. Dewan et al. proposed a deep learning approach for detect students' engagement through their emotions. Facial expression is an important and widely used cue for automatically detecting students' behavioral and emotional engagement. The authors developed a system that uses deep learning algorithms to analyze student interactions with online learning platforms and predict their level of engagement. The dataset was collected from an online learning platform that included data on student interactions such as clicks, time spent on each page, and quiz attempts. The dataset aws used to train their deep learning approach to identify student engagement levels.

Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN) are used to analyze student interactions and make predictions about engagement levels. They also used feature selection and data normalization techniques to optimize their model. The authors presented the performance of their model using various metrics such as recall, accuracy, precision, and F1-score. The outcomes showed that their deep learning approach outperformed traditional machine learning techniques in detecting engagement levels of online learners.

Merits :

- Deep Belief Network (DBN) was successfully used in many image classification applications

- The experiment result achieved a high accuracy (88%) in detecting engagement levels.

Demerits :

- Dataset used to train and test the model was limited to a specific online learning platform and may not generalize to other platforms
- The authors noted that their method didn't able to detect more complex forms of engagement

2.3 A Two-Stage Algorithm for Engagement Detection in Online Learning

S. Dash et al proposed that students' behavioural. (ON task and OFF task) and emotional (satisfied, bored, and confused) information for engagement detection. According to experimental findings, engagement detection that is based on behavioural and emotional characteristics is highly accurate. They proposed a two-stage algorithm for detecting student engagement levels in online learning environments. Machine learning models were used to analyze student activities and classify them into engaged and disengaged categories.

The dataset was collected from an online learning platform that included data on student activities such as login times, quiz attempts, and forum posts. Machine learning model was trained to classify student activities as engaged or disengaged. Two stages of classification were used for detecting student engagement levels. The first stage involved feature selection and feature engineering techniques to extract relevant features from the dataset. The second stage involved classification using Support Vector Machines (SVM) and Random Forest (RF). Model performance was evaluated using various metrics such as accuracy, precision, recall, and F1-score. The results showed that their two-stage algorithm outperformed traditional machine learning techniques in detecting engagement levels of online learners.

Merits :

- AllConv model was simple to build, reliable and gives high accuracy in pattern classification

- CMConv model achieved the highest accuracy over the other models. The reason because of using the advantageous features from the AllConv, VDConv, NiNConv, and CPConv models in the CMConv model

Demerits :

- Need to improve the performance of engagement detection in online learning using multiple traits such as facial expressions, gesture and posture.
- Algorithm relies heavily on the data collected from students, which may not always be accurate.

2.4 Assessing student engagement from facial behavior in on-line learning

In 2022, a study was conducted by Buono et al. to investigate how student engagement level could be monitored from facial expression and the engagement level of the students could be detected. In this work, a novel system for assessing learning engagement based on face expression recognition was presented. A method for assessing student engagement in online learning environments using facial behavior analysis was proposed. A system was developed that analyzed students' facial expressions and head movements to detect their level of engagement during a lesson.

A dataset of videos of students attending online lessons was used and manually labeled them as either engaged or disengaged. Then, a deep learning algorithm was used to automatically analyze the facial expressions and head movements of the students in the videos and classify them into engaged or disengaged categories. A Convolutional Neural Network (CNN) architecture was used for facial expression analysis and a Recurrent Neural Network (RNN) for head movement analysis. Transfer learning was used to fine-tune the pre-trained CNN model on their dataset.

The model performance of their system was evaluated using various metrics such as accuracy, precision, recall, and F1-score. The results showed that their method had achieved high accuracy in detecting student engagement levels.

Merits :

- Fusions of models output was used to achieve better performance.
- The best results was reached using Long Short-Term Memory (LSTM) networks

Demerits :

- The learning tasks should be more challenging, in order to better capture both physiological and psychological engagement
- It relies on facial behaviour data, which was not always accurately reflect the student's level of engagements as some students may not feel comfortable being monitored via webcam

2.5 Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate

H. Monkaresi et al(2017) used computer vision techniques to extract three sets of features from videos, heart rate, Animation Units. They explored how computer vision techniques could be used to detect engagement while students (N= 22) completed a structured writing activity similar to activities encountered in educational settings. This paper proposed a novelty that automatic estimation of attention of students during lectures in the E-Learning environment. The proposed work of this work was to develop an automated system for the detection of engagement in individuals using a combination of video-based estimation of facial expressions and heart rate. They used a webcam to capture facial expressions and a remote photoplethysmography (rPPG) sensor to estimate heart rate. The collected data is then processed and analyzed to extract features, which were used to train a machine learning model to classify the level of engagement of the individual. The authors used a dataset of 120 videos of 40 participants performing engagement-inducing tasks. They used the OpenFace software to extract facial expression features from the videos and the rPPG sensor to estimate heart rate. They used a support vector machine (SVM) to classify the level of engagement based on the extracted features.

Merits :

- Feature-level fusion was more successful for the concurrent models.

- It can capture both facial expressions and physiological signals, which was more accurate and comprehensive measure of engagement

Demerits :

Due to limits in analyzing head motions, and frequent face occlusions, their methods were not able to extract features from some video segments, thereby leading to data loss.

2.6 Automatic Recognition of Student Engagement using Deep Learning and Facial Expression

In this proposal, Nezami et al (2018) presented a deep learning model to improve basic facial expression recognition and engagement recognition. In the first step, a convolutional neural network (CNN) was trained on the dataset of the Facial Expression Recognition Challenge 2013 (FER-2013) to provide a rich facial representation model, achieving state-of the-art performance. In the next step, the model was applied to initialize our engagement recognition model, designed using a separate CNN, and learned on our newly collected dataset in the engagement recognition domain.

Merits :

- Usage of large dataset and manually annotated labels, which could improve the performance.
- Using deep learning techniques such as CNNs and MLPs could potentially provide more accurate and reliable results

Demerit :

The dataset used in the study was limited to a small number of undergraduate students attending a lecture which need more improvement.

2.7 Engagement Detection in e-Learning Environments using Convolutional Neural Networks

M. Murshed et al (2019) presented a framework that defined a spectrum of engagement states and an array of classifiers aimed to detect the engagement state of students. In group meeting, their proposed method is applied to detect engagement levels of the person involved throughout the meeting. The proposed method for detecting and classifying levels of engagement in group meetings using non-intrusive sensing technologies. They used a combination of audio, video, and physiological signals, such as heart rate and skin conductance, to determine the level of engagement of participants in meetings. The authors used a multi-modal approach to capture data from participants during meetings. They used wearable sensors to collect physiological signals, and cameras and microphones to capture audio and video data. They then used machine learning algorithms to classify the data and determine the level of engagement of each participant.

Merits :

- The authors used a large dataset and pre-processed visual features, which can improve the performance of the CNN model.
- The proposed CNN model achieved higher accuracy than All-CNN, VD-CNN, and NiN-CNN models

Demerits:

- Application of CNN models have not been widely investigated yet for this purpose.
- The dataset used in the study was limited to one MOOC platform, which may not be representative of all e-learning environments.

2.8 Engagement Detection in Meetings

Frank et al(2016) proposed framework consisting of five different modules that include detection, feature extraction, tracking, classification, and decision. The system was designed to detect (ROIs) of the learners in the live video stream. It first performed segmentation to isolate the ROIs. For each ROI, features were then extracted in a feature extraction module then to initiate tracking and classification. A tracking module was designed for tracking the movement or changes in the ROIs Finally, a decision module combineed classification scores to output a list of engagement levels. The proposed work was to review the existing literature on engagement detection in online learning environments. The work aimed to identify the current state of the art in engagement detection and to highlight the research gaps and future directions in this field. A comprehensive literature review of articles published between 2000 and 2018 on engagement detection in online learning environments was conducted. They analyzed the articles to identify the methods used for engagement detection, the types of data collected, and the performance metrics used to evaluate the engagement detection systems.

Merits :

- The authors used a combination of audio and video features, which could provide a more comprehensive view of audience engagement
- Engagement detection method used an engagement table to track each participants potential engagement level with each object.

Demerits :

- They did not discuss the evaluation of engagement between humans in normal interaction.
- The study did not compare the proposed methodology with other engagement detection methods, which limits the assessment of its performance.

2.9 Engagement detection in online learning: a review

Dewan et al (2019) presented a deep learning model to improve basic facial expression recognition and engagement recognition. In the first step, a convolutional neural network (CNN) was trained on the dataset of the Facial Expression Recognition Challenge 2013 (FER-2013) to provide a rich facial representation model, achieving state-of the-art performance. In their work, engagement recognition model got initialized, built by using a separate CNN, and trained on new dataset. The proposed work was to develop an automated system for recognizing student engagement in classroom settings using deep learning and facial expression analysis. The system aimed to improve teaching and learning outcomes by providing real-time feedback to teachers and students about the engagement levels of individual students. The authors used a dataset of videos of students engaged in classroom activities. They used deep learning techniques to analyze the video data and extract features related to the student's facial expressions. The extracted features were used to train a machine learning model to classify the student's engagement level. The authors used the FER2013 dataset for facial expression analysis, which contains labeled images of human faces showing different expressions.

Merits :

- The accuracy of the facial expression-based channels was found to be higher than the heart-rate channel.
- The authors highlighted the strengths and limitations of each methodology, which can inform the selection of appropriate methods

Demerits :

Some of the challenges include illumination variation, occlusions, head poses, errors in speech detection, objects appearing too far or close.

2.10 Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models

Gupta et al. proposed a real-time learner engagement detection system that uses facial emotion recognition based on deep learning models. The system was designed to used in the context of online learning and aims to provide accurate and timely feedback to instructors about the engagement levels of learners. The system was composed of three main components: facial expression detection, emotion recognition, and engagement level determination. The system used a camera to capture the facial expressions of learners and applied deep learning models to recognize emotions and determine the level of learner engagement. The system was trained on a dataset of online learning sessions to achieve high accuracy in detecting learner engagement levels. The proposed system used deep learning models for facial expression detection and emotion recognition. The system used a pre-trained Convolutional Neural Network (CNN) for facial expression detection and a pre-trained Recurrent Neural Network (RNN) for emotion recognition. The engagement level determination component used a rule-based system that combines the results of facial expression detection and emotion recognition to determine the level of learner engagement. The proposed system was tested on a dataset of online learning sessions and evaluated using accuracy and F1-score metrics.

Merits :

- The system provided immediate feedback to instructors on the engagement levels of their students, allowing for more effective teaching and learning.
- The use of deep learning models allowed for more accurate and precise detection of emotions, increasing the reliability and effectiveness of the system.

Demerits:

- The system relied heavily on the accuracy and consistency of facial recognition technology, which can be influenced by factors such as lighting, camera angle, and facial expressions.
- The use of such a system raised concerns around privacy and data security, as it involved with personal data without the explicit consent of all parties involved.

2.11 Students' emotion extraction and visualization for engagement detection in online learning

Hasnine, M. N. et.al proposed system uses facial expression recognition and sentiment analysis techniques to extract emotions from students' faces and text-based inputs, respectively. The system then visualizes the extracted emotions in real-time through a web interface that provides engagement feedback to students and instructors. Emotion extraction involves collecting data on students' emotions using sensors such as cameras, microphones, and wearable devices. The data was then analyzed using machine learning algorithms to identify specific emotions such as happiness, sadness, and boredom. The emotion analysis step involves using these emotions to determine the level of engagement of the student, which can be measured using a scale from 1 to 5. Finally, the visualization step involved with presenting the data in a way that is easy to understand and interpret. This could be done using various visualization techniques such as graphs, charts, and heatmaps.

Merits :

- They proposed a novel approach for detecting students' engagement levels in online learning by extracting and visualizing emotions.
- The study showed promising results in accurately detecting students' emotions and engagement levels using various machine learning algorithms and visualization techniques.

Demerits :

- It only focused on the analysis of textual data, which did not provide a comprehensive picture of student emotions during online learning.
- It did not discuss potential ethical concerns or privacy issues related to the collection and analysis of students' emotional data in online learning environments.

2.12 Diversified Model Combination Network for Understanding Engagement from Video Screengrab

Sarthak Batra et.al proposed methodology involved collecting a dataset of video screengrabs and their engagement scores, preprocessing the data by resizing and normalizing the images and splitting the dataset into training, validation, and testing sets. They proposed a DMCNet architecture consisting of several parallel sub-networks with different combinations of convolutional and pooling layers, and an attention mechanism to weight the contribution of each sub- network based on the input image. The authors trained the DMCNet using a combination of mean squared error and mean absolute error as loss functions, optimized the network using the Adam optimizer, and used data augmentation techniques to prevent overfitting. Finally, they evaluated the performance of the DMCNet on the test set using several evaluation metrics and showed that it outperformed several baseline models in terms of accuracy and robustness to noise.

Merits :

- DMCN proposed in this work was effectively combine multiple visual and audio features to improve the accuracy of engagement prediction from video screengrabs.
- DMCN was designed to address the issue of overfitting in engagement prediction models, which can lead to poor generalization performance on new data.

Demerits :

- The proposed DMCN requires a large amount of training data and computational resources to train and evaluate, which may be a challenge for some applications.
- It did not provide a detailed analysis of the interpretability of the DMCN model, which may limit its utility in some domains.

2.13 Student-Engagement Detection inClassroom Using Machine Learning Algorithm

Alruwais et.al. presented a methodology for detecting student engagement in a classroom setting using a machine learning algorithm. The methodology consists of two main stages: feature extraction and classification. In the feature extraction stage, the algorithm processes a video feed from a webcam and detects facial landmarks such as eye gaze, head pose, and facial expression. These features were then extracted from the video frames and used as input to the classification stage. In the classification stage, a support vector machine (SVM) was trained on the extracted features to predict whether the student is engaged or not.

Merits :

- The use of machine learning algorithms can provide an objective and reliable way to detect student engagement in the classroom.
- The system proposed has the potential to provide real-time feedback to instructors which improve student learning outcomes.

Demerits :

- There may be ethical concerns around the use of technology to monitor student behavior that violate student privacy.
- The accuracy of the machine learning algorithm may be affected by factors such as the quality of the data collected and the variability in human behavior.

2.14 Detecting distracted students in educational VR environments using machine learning on eye gaze data

Sarker et.al. proposed a methodology to detect distracted students in virtual reality (VR) educational environments using machine learning applied to eye gaze data. The study collected eye gaze data of students in a simulated VR classroom setting and used support vector machine (SVM) algorithm to train a model that can distinguish between attentive and distracted students. The extracted features from the eye gaze data include pupil diameter, fixation duration, saccade velocity, and fixation count. The proposed methodology was evaluated using accuracy, precision, recall, and F1-score, and the results showed that the SVM model was effective in detecting distracted students with an accuracy of 82.5%

Merits :

- The use of machine learning on eye gaze data provides a non-intrusive and objective method for detecting distracted students in VR educational environments.
- The detection of distracted students can help improve the effectiveness of VR educational environments by providing timely interventions and support.

Demerits:

- The study was conducted in a controlled laboratory environment, which may not fully represent the distractions that can occur in real-world educational settings.
- The sample size of the study was relatively small, which may limit the generalizability of the results to larger populations.

2.15 Investigating student engagement with intentional content: An exploratory study of instructional videos.

Walsh, J. N. et.al. surveyed 188 undergraduate students to gather demographic information and measure their engagement with the videos. Focus groups were conducted with 12 students to provide more in-depth information about their engagement with the videos. Video analytics were used to collect data on the students' viewing behavior, such as how long they spent watching each video and how many times they watched each video. The data collected were analyzed using thematic analysis and descriptive statistics to identify common themes and patterns. The study found that intentional content has a positive impact on student engagement and that students value videos that are concise, visually appealing, and present information in a clear and organized manner.

Merits :

- The study provides valuable insights into student engagement with instructional videos, which can inform the development of effective online learning materials.
- The use of mixed-methods approach, including both quantitative and qualitative data collection, enhances the reliability and validity of the study findings.

Demerits:

- The sample size of the study is relatively small, which may limit the generalizability of the findings to a larger population.
- The study is limited to a specific context and may not be applicable to other settings or disciplines, which may require different types of instructional videos.

Limitations of existing systems

Most of the existing works consider only the facial images of the students. They failed to discuss issues related to environmental constraints of a student such as head poses. Video captures for engagement detection in a learning context are not described. Data loss due to limits in analyzing head motions, and frequent face occlusions. Data loss due to limits in analyzing head motions, and frequent face occlusions.

2.16 SUMMARY

In this chapter, the existing projects have proposed models for engagement of the students in E-learning environments. There are several techniques to monitor the students during online classes. In the next chapter study of the system describes the procedures which are taken in this work. It will provide a detailed study and explaining how these limitations are overcome in the proposed system.

CHAPTER 3

SYSTEM STUDY

3.1 OVERVIEW

In this chapter, the technologies used in this project which includes Haar cascade classifier, CNN and Mobilenet are discussed. It deals with the detailed description of our proposed system. The section 3.2 explains Haar cascade classifier algorithm. The section 3.3 describes about Convolution Neural Network. The section 3.4 contains the detailed information about another convolution neural network called Mobilenet.

3.2 FOCUS DETECTION

Focus detection is done for the students based on their eye focus. It makes use of face and eye pairs which is detected from the video frames. The technology analyzes images of the students captured by their webcams during the online class session. Haar cascade classifier algorithm is used for detecting the faces and eyes of the students.

3.2.1 HAAR CASCADE CLASSIFIER

Haar Cascade algorithm is a popular method for object detection that quickly and efficiently extract features from images. To achieve accurate detection, the algorithm requires extensive training with numerous positive and negative images.

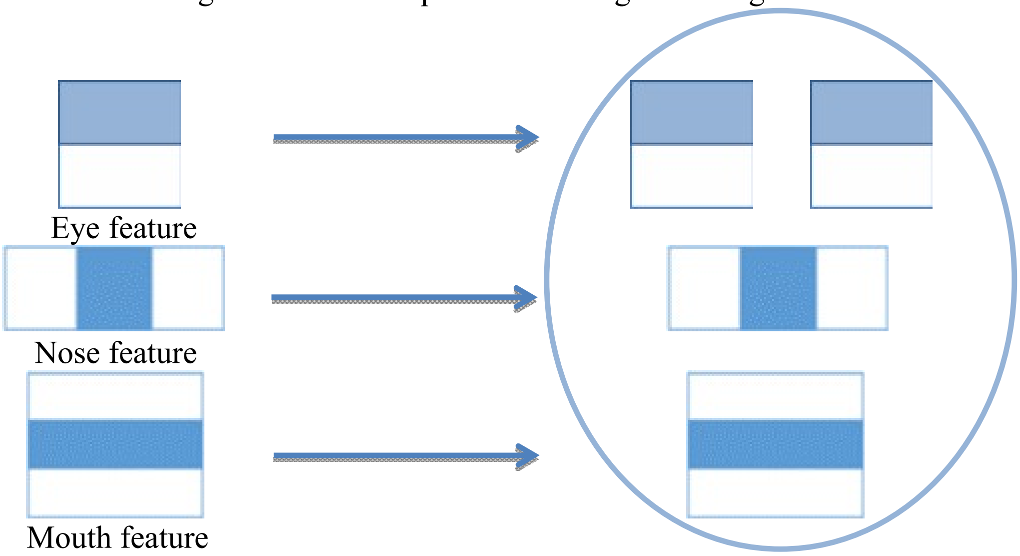


Figure 3.1 Structure of Haar cascade classifier

It is a classifier that is trained to detect face, nose and mouth in input images. It uses both positive (images with faces) and negative (images without faces) samples to train the classifier on different feature sets like full-body, lower-body, eye, frontal-face, etc. The trained model is saved as .xml files and can be used to detect objects in other input images. Figure 3.1 shows the structure of this algorithm.

The classifier has four main stages:

Haar feature selection, which calculates features in subsections of the input image by measuring the difference between the sum of pixel intensities of adjacent rectangular regions. Creating an integral image to reduce computation time by using only four pixels instead of all pixels.

AdaBoost algorithm is used to select relevant Haar-like features for classification and make a strong classifier using a mix of all weak classification functions.

Cascading classifiers are used to classify facial regions in the image by grouping features into different stages of the classifier, rather than using all features at once on all regions of the image.

Once trained, the algorithm can detect similar images based on the previously learned features. This algorithm to detect the student’s frontal face in the image and locate the eyes region within the face. From the extracted frames, faces and eyes are detected.

3.2.2 CONVOLUTIONAL NEURAL NETWORK(CNN)

CNN or Convolutional Neural Network, is a type of neural network commonly used for image and video processing tasks, such as image classification, object detection, and facial recognition. CNN consists of multiple layers that perform different operations on the input data, such as convolution, pooling, and activation.

Input layer: This layer is responsible for taking the input data, which is usually an image or a sequence of images. The input data is represented as a matrix of pixel values.

Convolutional layer: This layer performs a convolution operation on the input data, which involves sliding a small filter over the input image and computing the dot product between the filter and the corresponding pixels in the input. This operation helps to extract features from the input image.

ReLU layer: This layer applies a non-linear activation function called Rectified Linear Unit (ReLU) to the output of the convolutional layer. ReLU helps to introduce non-linearity into the model, which is important for capturing complex relationships between the input and output.

Pooling layer: This layer reduces the spatial dimensions of the output from the previous layer by taking the maximum or average value of a small region in the input. This operation helps to reduce the number of parameters in the model and also makes the model more robust to small translations and distortions in the input, shown in fig 3.2.

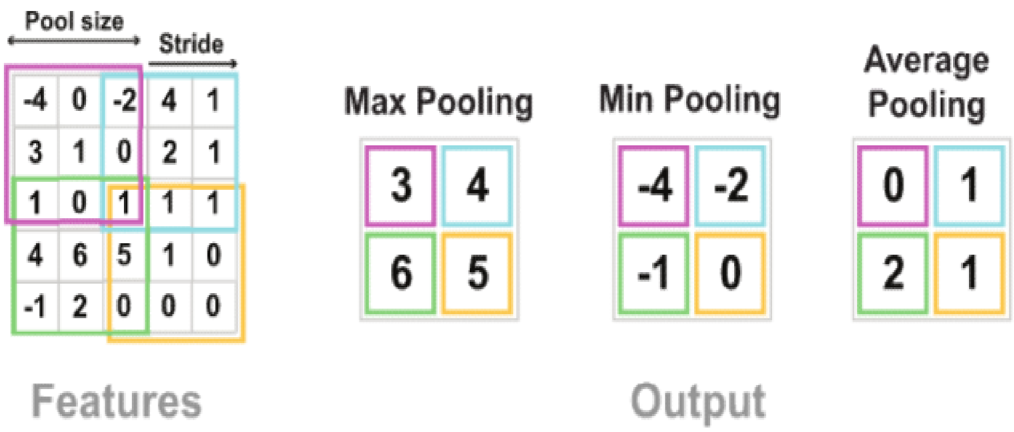


Fig 3.2 pooling layers

Source : https://epynn.net/Pooling.html

Dropout layer: This layer randomly drops out a fraction of the output from the previous layer during training. This helps to prevent overfitting and improves the generalization ability of the model.

Fully-connected layer: This layer connects all the neurons from the previous layer to all the neurons in the current layer, which allows the model to learn complex, non-linear relationships between the input and output. This layer is often followed by a softmax layer, which converts the output to a probability distribution over the different classes.

The layers in a CNN are typically arranged in a sequential order, with the input layer followed by several convolutional and pooling layers, followed by one or more fully-connected layers, and finally the output layer. The number of layers and the size of each layer can vary depending on the specific task and the complexity of the input data.

3.3 EMOTION DETECTION

Emotion detection is the measure of emotional state of the students as they participate in virtual learning environments. It can be used by educators to monitor the engagement levels and emotional well-being of students in real-time. One of the popular Convolution Neural Networks named Mobilenet is used here for detecting the students' emotion.

3.3.1 MOBILENET

MobileNet is a popular convolutional neural network (CNN) architecture that is specifically designed for efficient computation on mobile and embedded devices. It achieves this by using a lightweight network structure that reduces the number of parameters and computations required for inference, while maintaining high accuracy on image classification tasks. It has 27 Convolutions layers which includes 13 depthwise Convolution, 1 Average Pool layer, 1 Fully Connected layer and 1 Softmax Layer. In terms of Convolution layers, which are shown in table 3.1.

Table 3.1 Mobilenet Convolution layers

13	3x3	Depthwise Convolution
1	3x3	Convolution
13	1x1	Convolution

Input Layer: This layer is the first layer in the network and takes in the input image data. MobileNet typically takes in RGB images of size 224 x 224 pixels as input.

Convolutional Layers: MobileNet uses a series of convolutional layers to extract features from the input image. These layers use 3x3 convolutional filters with a stride of 1 and same padding to maintain the input spatial dimensions. The depth of each convolutional layer gradually increases as we move deeper into the network, allowing it to capture more complex features.

Depthwise Separable Convolution Layers: These layers are the key component of the MobileNet architecture, which helps to reduce the number of computations required for inference. Instead of using traditional convolutional layers, MobileNet uses depthwise separable convolution layers that split the convolution operation into two separate operations: a depthwise convolution and a pointwise convolution. The depthwise convolution applies a single filter per input channel, while the pointwise convolution applies a 1x1 convolution to combine the output of the depthwise convolution. This reduces the number of parameters and computations required for the convolution operation.

Table 3.2 Mobilenet architecture

Type / Stride	**Filter Shape**	**Input Size**
Conv / s2	3 x 3 x 3 x 32	224 x 224 x 3
Conv dw / sl	3 x 3 x 32 dw	112 x 112 x 32
Conv / s l	1x 1 x 32 x 64	112 x 112 x 32
Conv dw / s2	3 x 3 x 64 dw	112 x 112 x 64
Conv / sl	1x 1 x 64 x 128	56 x 56 x 64
Conv dw / sl	3 x 3 x 128 dw	56 x 56 x 128
Conv / sl	1x 1x 128 x 128	56 x 56 x 128
Conv dw / s2	3 x 3 x 128 dw	56 x 56 x 128
Conv / sl	1x 1 x 128 x 256	28 x 28 x 128
Conv dw / s l	3 x 3 x 256 dw	28 x 28 x 256
Conv / sl	1x 1 x 256 x 256	28 x 28 x 256
Conv dw / s2	3 x 3 x 256 dw	28 x 28 x 256
Conv / sl	1x 1 x 256 x 512	14 x 14 x 256
5x Conv dw / s l Conv / s l	3 x 3 x 512 dw 1x 1 x 512 x 512	14 x 14 x 512 14 x 14 x 512
Conv dw / s2	3 x 3 x 512 dw	14 x 14 x 512
Conv / sl	1x 1 x 512 x 1024	7 x 7 x 512
Conv dw / s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv / sl	1x 1x 1024 x 1024	7 x 7 x 1024
Avg Pool / s l	Pool 7 x 7	7 x 7 x 1024
FC / sl	1024 x 1000	1 x 1 x 1024
Softmax / sl	Classifier	1 x 1 x 1000

Batch Normalization Layers: MobileNet uses batch normalization layers after each convolutional and depthwise separable convolution layer. This helps to reduce the internal covariate shift in the network and speeds up the training process.

ReLU Activation Layers: Rectified Linear Unit (ReLU) activation layers are used after each convolutional and depthwise separable convolution layer. This introduces non-linearity into the network and helps to capture complex features.

Global Average Pooling Layer: This layer is used to reduce the spatial dimensions of the output of the convolutional layers. It takes the average of each feature map and outputs a single value for each channel. This reduces the number of parameters in the network and helps to prevent overfitting.

Fully Connected Layer: Finally, the global average pooling layer is connected to a fully connected layer with a softmax activation function, which outputs the predicted class probabilities which is clearly shown in table 3.2.

3.4 SUMMARY

This chapter gives an overview about the algorithm and techniques used in our project. In the next chapter we will be discussing about working flow, modules and its description. The system design will help to understand how the system works and the outcomes of the project.

CHAPTER 4

SYSTEM DESIGN

This chapter deals with the modules used in the project and explains the sequence in which data flows. This flowchart describing the entire flow of the proposed system is shown in the Figure 4.1.

4.1 OVERVIEW

The detailed description of the modules and procedures to implement those modules are described in this chapter. In order to propose the system work, design should be made available. System design will help us to understand what will be done and how the system works. It includes various modules which are combined together to produce the result. It is the working flow of the system represented as a flowchart in Figure 4.1.

4.2 SYSTEM ARCHITECTURAL DESIGN

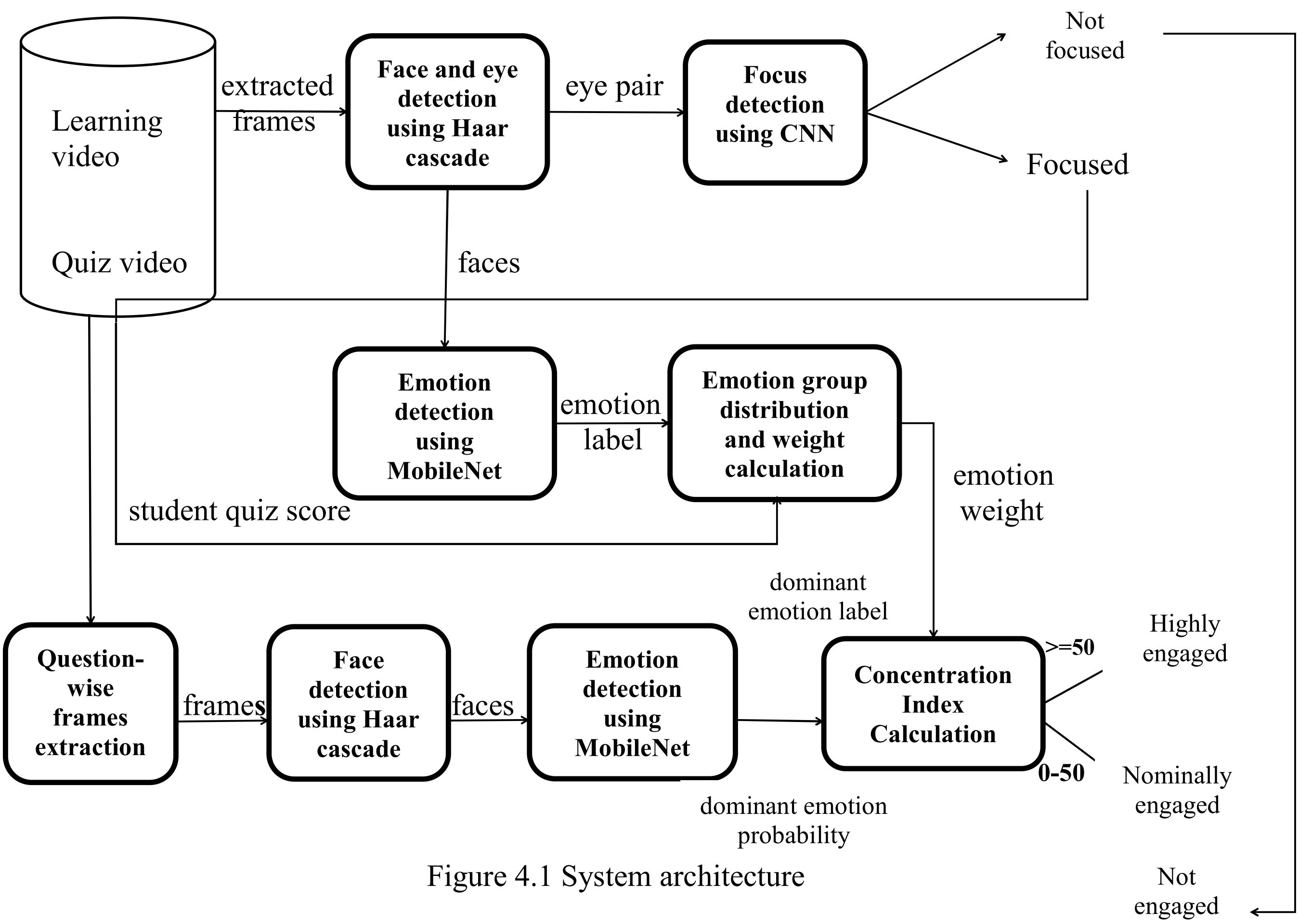


Figure 4.1 System architecture

4.3 MODULE DESCRIPTION

The modules present in the proposed model are as follows:

1. Face and Eye detection
2. Focus detection using CNN
3. Emotion detection using MobileNet
4. Emotion Group distribution and weight calculation
5. Concentration Index Calculation and Engagement Classification

4.3.1. Face and Eye detection

It means detecting human face and eyes by identifying the features of a human face from images or video streams. From the input image, face and eye regions are detected using haar cascade classifier.

In order to find whether a student is focused or not focused, their faces and eyes will be detected. It makes use of the students' eye focus. It refers to the direction and position of a student's eyes. Eye focus can provide valuable information about a student's attention.

4.3.2. Focus Detection using CNN

Focus detection is a technology that helps teachers or online class instructors to monitor the attention and engagement of students during online classes. A CNN model is used for Focus Detection whose architecture is shown in Table 4.1. FER2013 dataset is used for building CNN model which consists of gray scale images of face. Using Haar cascade classifier, face and eyes of the students can be detected. Their eye regions are cropped for further classification process.

The Convolutional Neural Network (CNN) is a popular deep learning algorithm that can efficiently analyze images and identify features within them. By assigning adjustable weights and biases to various components of an image, the CNN can differentiate and classify them into various categories. This approach is particularly useful because it reduces the computational burden of processing images, making it faster and more efficient. CNNs have found widespread use in various fields due to their ability to extract important information from images and classify them accurately. In this work, eye images that were obtained after applying the Haar Cascade model on facial images obtained from the FER-2013 Dataset. The dataset was annotated a group of labelers into "Focused", "Not-Focused" categories. The following constraints are used to label the image:

• Focused -"looking towards or around the screen".

• Not-Focused-"eyes closed", "looking away from screen".

Table 4.1 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)

FER 2013 Dataset is used to train the CNN. The image is resized into 224 x 224 pixels. Fig. represents our CNN model. It contains 3 convolutional2D layers with 3 pooling layers to extract features. Pooling layers are used to reduce the dimensions of the feature maps. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network. Input image is a grayscale image with 7 emotion classes. In the sequential model, Input layer 32×32 holds the pixel values of input images. Convolutional layers with a set of 3×3 filters computes the low-level features from the input images.

Filters helps in extracting specific features from input data. Max Pooling layer of pool size 2×2 reduces the spatial size of convolved features. It also helps reduce over-fitting by providing an abstracted representation. Classification through a Fully connected layer computes the two class scores using Relu as the activation function. It doesn't allow for the activation of all of the neurons at the same time. i.e., if any input is negative, ReLU converts it to zero and doesn't allow the neuron to get activated. This means that only a few neurons are activated, making the network easy for computation.

It has 2 dense layers and sigmoid function is chosen as the last activation function to predict the probability as an output. Since probability of anything exists only between the range of 0 and 1, sigmoid function will be preferred. Using the probability value, it then classifies whether the student is focused or not focused which will be represented using the labels. Label 0 represents that the student is focused and Label 1 represents that the student is not focused. The sigmoid function is shown in equation 1.

$$\mathbf{f\,(x)} = \frac{\mathbf{1}}{\mathbf{1} + e^{-x}} \qquad \mathbf{(1)}$$

In total, there are 5 layers in neural network: 1 input layer, 2 convolutional layers, 2 pooling layers, 1 flatten layer, and 1 dense layer The output layer consists of a single neuron with an activation function, which outputs a probability score between 0 and 1 that represents the likelihood that the student is in focus.

4.3.3. Emotion Detection using MobileNet

MobileNet is a popular convolutional neural network (CNN) architecture designed to be lightweight and efficient, while still maintaining high accuracy on image classification tasks. It achieves this by using depthwise separable convolutions, which consist of a depthwise convolution (which applies a single filter per input channel) followed by a pointwise convolution (which applies a 1x1 filter to combine the outputs of the depthwise convolution). They are based on a streamlined architecture that uses depthwise separable convolutions to build lightweight deep neural networks. It uses depthwise separable convolutions. It significantly reduces the number of parameters when compared to the network with regular convolutions with the same depth in the nets.

For this work, FER-2013 dataset is used , which consists of 35,887 images of size 48x48 pixels, labeled with one of seven facial expressions: anger, disgust, fear, happiness, sadness, surprise, and neutral. Transfer learning is used by initializing our model with a pre-trained MobileNet architecture, which was fine-tuned on FER-2013 dataset which is shown in table 4.2. Categorical cross-entropy loss function and the Adam optimizer with a learning rate of 0.001. 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 layerson top of it to fine-tune it for a specific task.

The output layer of the network predicts the emotion of the input data, classifying it into seven categories, namely Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

Table 4.2 Mobilnet 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
Convolution	1x1x256x256	28x28x256
Conv depthwise	3x3x256 dw	28x28x256
Convolution	1x1x256x512	14x14x256
5 x(conv depthwise) (convolution)	3x3x512 dw 1x1x512x512	14x14x512 14x14x512
Convolution 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	1x1x7

For every frame, their emotion will be identified using MobileNet. Various emotions such as angry, disgust, fear, happy, neutral, sad and surprise can be classified for the students using this architecture. MobileNet classifies the facial emotions data, generating the Emotion Probability for every frames. The emotion with the highest probability score will be the one that the model predicts as the most likely emotion for that student.

4.3.3.1 Dominant Emotion Probability (DEP)

N frames are extracted for a minute from the input video. Emotion is recognized for each frame. Emotion probability and emotion class will be predicted which is mostly expressed by a student. Mode value is used for determining the dominant emotion.
From that dominant emotion, Maximum value of that particular emotion probability is considered as Dominant Emotion Probability(DEP) for the particular student. Class label and emotion label are shown in table 4.3.

Table 4.3 Class and emotion labels

CLASS	EMOTION
0	Angry
1	Disgust
2	Fear
3	Happy
4	Neutral
5	Sad
6	Surprise

4.3.4 Emotion Group Distribution and Weight Calculation

While listening the video, if the student expresses particular emotion for majority of the time (more than 50%) means that particular student falls under that particular emotion group. For Example, if the student expresses surprise emotion for more than 50% time means the student will come under surprise emotion group.

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." It is calculated as the value that describes how much a specific emotion expressed over a time which is calculated using mean score of the student. It’s value ranges from 0 to 1. Such calculation is done for each group.

Emotion weight is calculated by taking the average of these scores for each emotion group. It is the mean score of quiz which is calculated for each student. The formula for calculating the mean score achieved in a quiz for each emotion group is given in equation 2.

$$\text{Mean score, } EW_g = \text{Total score}_g / N_g \qquad (2)$$

where,

EW_g -> Emotion weight belonging to a group

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

N_g -> The total number of students who took the quiz.

The quiz score can be of 3 categories which includes 0,1 and -1
where,

0 - wrong answers, not attended questions

1 - correct answers

-1 - referring browser, answer discussing with others

Based on these constraints score will be awarded to the students who are attending the quiz.

4.3.5 Concentration Index Calculation and Engagement Classification

For calculating Concentration Index(CI) ,frames are extracted while attending each questions, From sequence of frames, faces are detected by using Haar Cascade Classifer. Sequence of face frames are fed into the mobilenet architecture, from that Dominant

Emotion and its probability is calculated. The resulting Concentration Index (CI) is determined by multiplying the Dominant Emotion Probability (DEP) value by the corresponding Emotion Weight (EW), which is expressed by equation 3.

$$CI = DEP \times EW_g \tag{3}$$

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.

Engagement Classification :

After CI is calculated, its values will be grouped into 2 levels based on the index value. The two levels which includes highly engaged and nominally engaged.

- ✓ 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**.

- ✓ When a student is not focused, they fall under the **not engaged** category.

4.4 SUMMARY

This chapter gave an overview about all the modules and techniques used in our project. The next chapter will give the detailed description of implementation process for each module. It will provide details about how this design is implemented to detect the engagement level of the student.

CHAPTER 5

SYSTEM IMPLEMENTATION

5.1 OVERVIEW

In this chapter, the algorithms and formula for implementing the modules of the proposed system has been discussed. The following are the modules which are present in this proposed model that consists of focus detection, concentration index calculation and engagement classification are as follows.

5.2 Focus Detection:

Based on the students' eye focus, focus detection is performed. It uses the faces and eyes pairs of students that are identified from video frames. During the online lesson, the students' webcams record the session that the system then analyzes. The Haar cascade classifier technique is utilized to identify the students' faces and eyes.

```
Input : video frames of the collected datasets,CNN model
Output : Eye focus detection(fvs)
Process : from the given input image, focus detection is done
fvs=[ ]
frames= frameSeparation(videoframe)
for each frame:
   frg = rgbtoGray(frame)
   face = DetectFaces(frg)
   eyes=DetectEyes_Haarcascade(face)
   fv=EyeFocusDetection(eyes,CNNmodel)
   fvs=fvs ∪ fv
```

5.3 Emotion Detection

Student faces are taken as input for detecting their emotion. A student can express multiple emotions. Among this, dominant emotion is considered as his\her emotion class. Convolution Neural Network namely Mobilenet is used as the model for emotion detection.

Input : student class listening video,Mobilenet model
Output : Emotion weight(fes)
Process : from the given input class listening video, emotion detection is done

```
fep=[ ]
frames= frameSeparation(video)
for each frame:
   frg = rgbtoGray(frame)
   face = detectFaces(frg)
   (fp,fe)=emotionDetection(face,mobnetModel) // fe - emotion & fp - probability
   fps=fps ∪ fp
   fes= fes  ∪ fe
de=mode(fes)
dep=max(fps)
```

5.4 Groupwise Emotion Distribution and Weight Calculation

Input: dominant emotion(de)
Output: Emotion weight(EWg)
Process: Using dominant emotion, emotion weight is calculated for the grouped students

```
students = [ ]
groups = [ ]
for each student:
   for each emotion:
         students= GroupStudentsByEmotion(de_i)
```

groups[emotion]=groups[emotion] $\cup$ students

for each group:

EW_{group} = Total $score_{group}$ / N_{group}

5.5 Concentration Index Calculation and Engagement Classification

Input : Dominant Emotion Probability (DEP), Emotion weight (EW_g)

Output : Concentration Index (CI)

Process : Using the input values, concentration index will be computed

for each student:

$CI = DEP \times EW_g$ // g is the emotion group the student belongs

```
begin;
 if CI >= 50 then
   EC ← highly engaged;
 elseif 0<=CI<50
   EC ← nominally engaged ;
 else
   EC ← not engaged;
end
```

5.6 SUMMARY

This chapter gave the algorithms and formula for implementing the modules of the proposed work. In the next chapter we will discuss about the results, conclusions, social impact and applicability of the proposed work in detail.

CHAPTER 6

RESULTS AND DISCUSSION

6.1 Dataset Description

In this section, we are going to discuss about the dataset used. For our work, we have used 3 dataset named FER 2013, Emo-Detect dataset and Auto-Engage-Detect dataset.

6.1.1 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 angry, disgust, fear, happy, neutral, sad and surprise with emotion labels. Each image from every emotion is shown in figure 6.1.

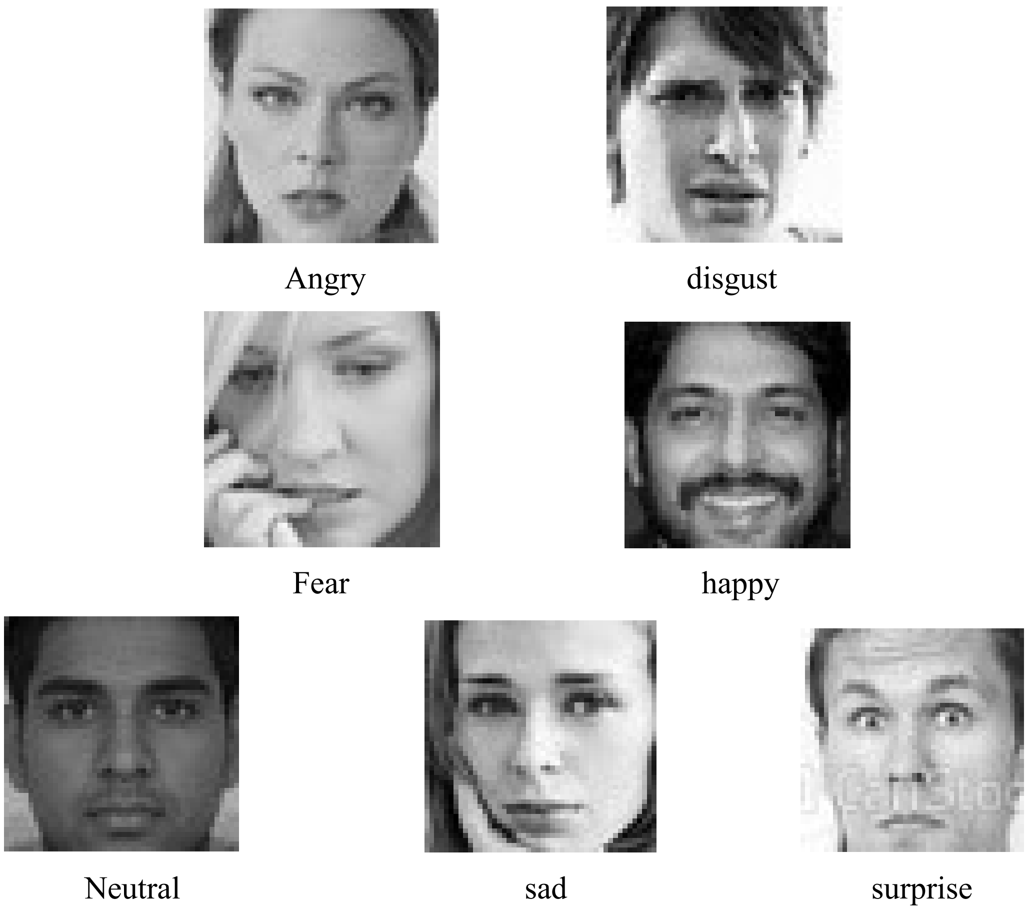


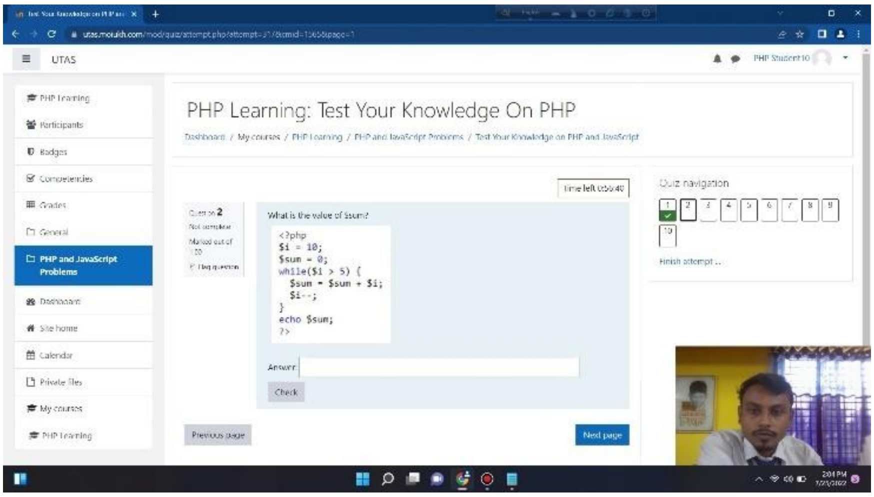
Figure6.1. Sample images of FER 2013 Dataset

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 emotions label consists a numeric code for each student and represents values 0 as Angry, 1 as Disgust, 2 as Fear, 3 as Happy, 4 as Sad, 5 as Surprise, and 6 as Neutral.

In this project, this dataset is used for training the focus detection model and emotion detection model. While testing student class listening videos are used.

6.1.2 EMODETECT 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. Samplevideo frame will be shown in figure 6.2.



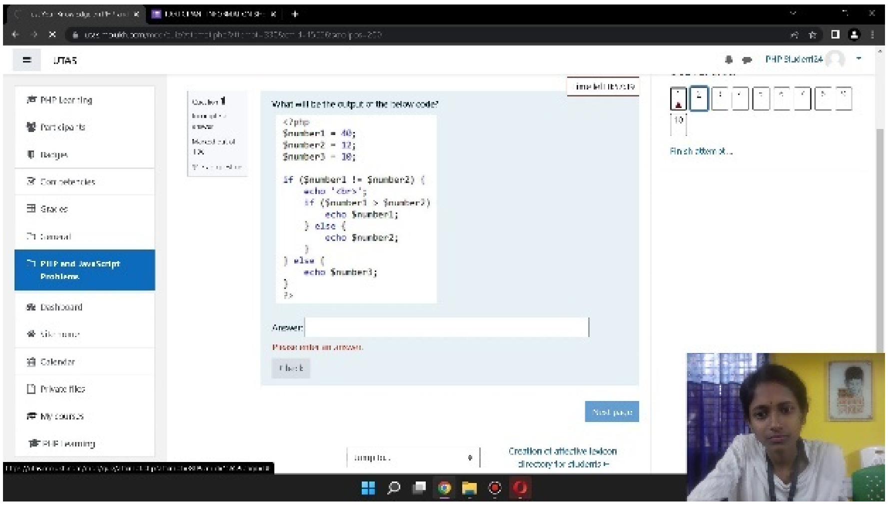


Figure 6.2. Video frame of Emodetect dataset

6.1.3 Auto-Engage-Detect dataset

This dataset is collected from 21 students. It contains video recordings of students' faces while listening informative video followed by attending quiz. It holds 10 minutes in total for each student. i.e, 5 minutes for listening the video related to data structures and 5 minutes for attending the quiz(approx.). They are involved with various emotions throughout the session. There are 10 quiz in total related to data structures. Figure.6.3.1.shows sample video frame while listening informative video and Figure.6.3.2. shows sample video frame while attending quiz. And the quiz questions are listed below.

1.Which topic did you studied 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

3.Which of the following is a linear data structure?

a.array b.binary tree c.queue d.stack and queue

4.  

This image denotes which data structure in real life?

a.queue b.list c.stack d.array

5. 

This image denotes which data structure in real life ?

a.queue b.list c.stack d.array

6.What is the arrangement followed in Stack DataStructure?

a.FIFO b.LIFO c.FILO d.LILO

7.Queue follows the ____________principle in arranging the data

a.LIFO b.FIFO c.LILO d.FILO

8.What is an array?

a.A group of elements of same data type.

b. An array contains more than one element.

c. Array elements are stored in memory in continuous or contiguous locations.

d.All of the above.

9.If the elements "A", "B", "C" and "D" are placed in a queue and are deleted one at a time, in what order will they be removed?

a.ABCD b.DCBA c.DCAB d.ABDC

10.Select linear data structures

a.queue b.array c.list d.tree

A sample video frame of the student while attending class is shown in figure 6.3.1.

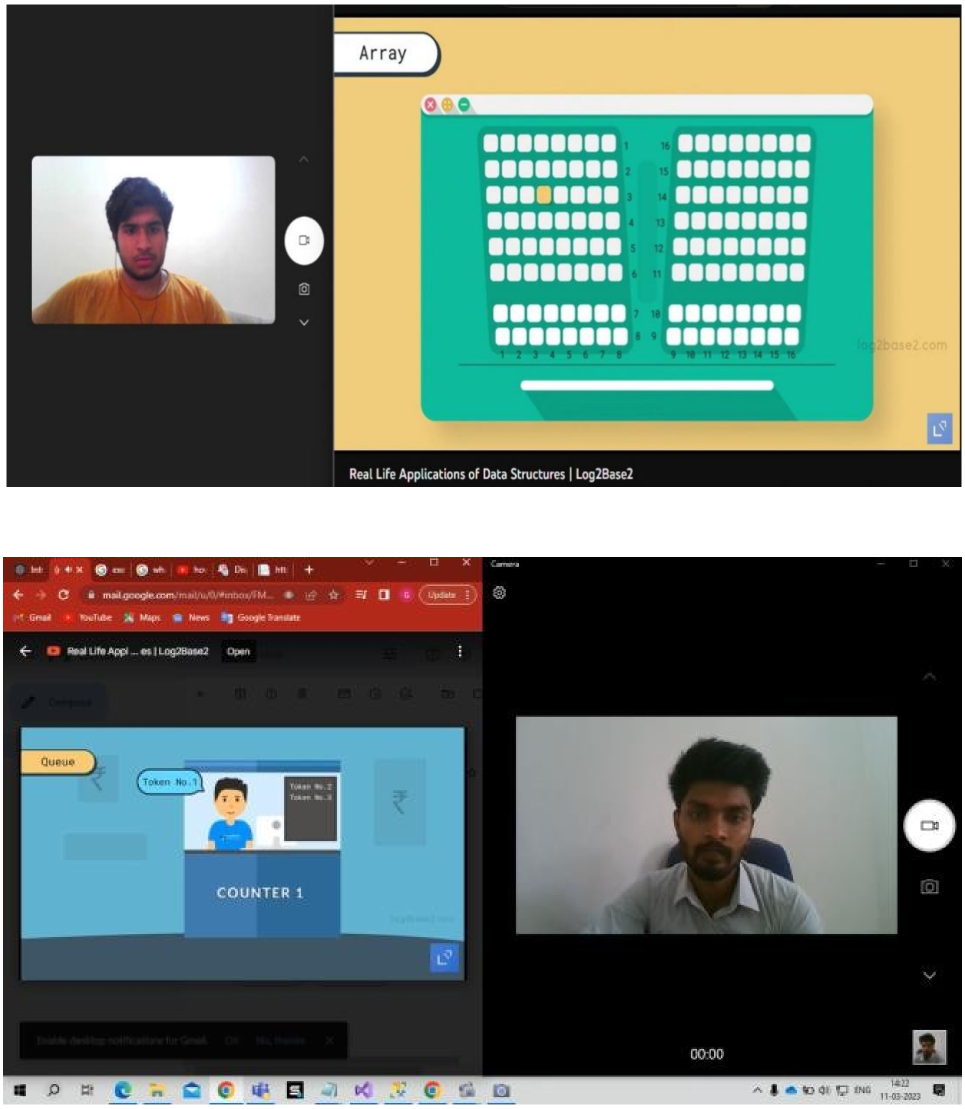


Figure.6.3.1 Video frame while listening video

The dataset consists of 21 video recordings in total in which the students are listening to an informative video. Webcam will record the faces of the students as shown in the figure.6.3.1. Here we have taken the topic under Data Structures which is about 5 minutes. It includes the basics and useful contents relevant to the topic.

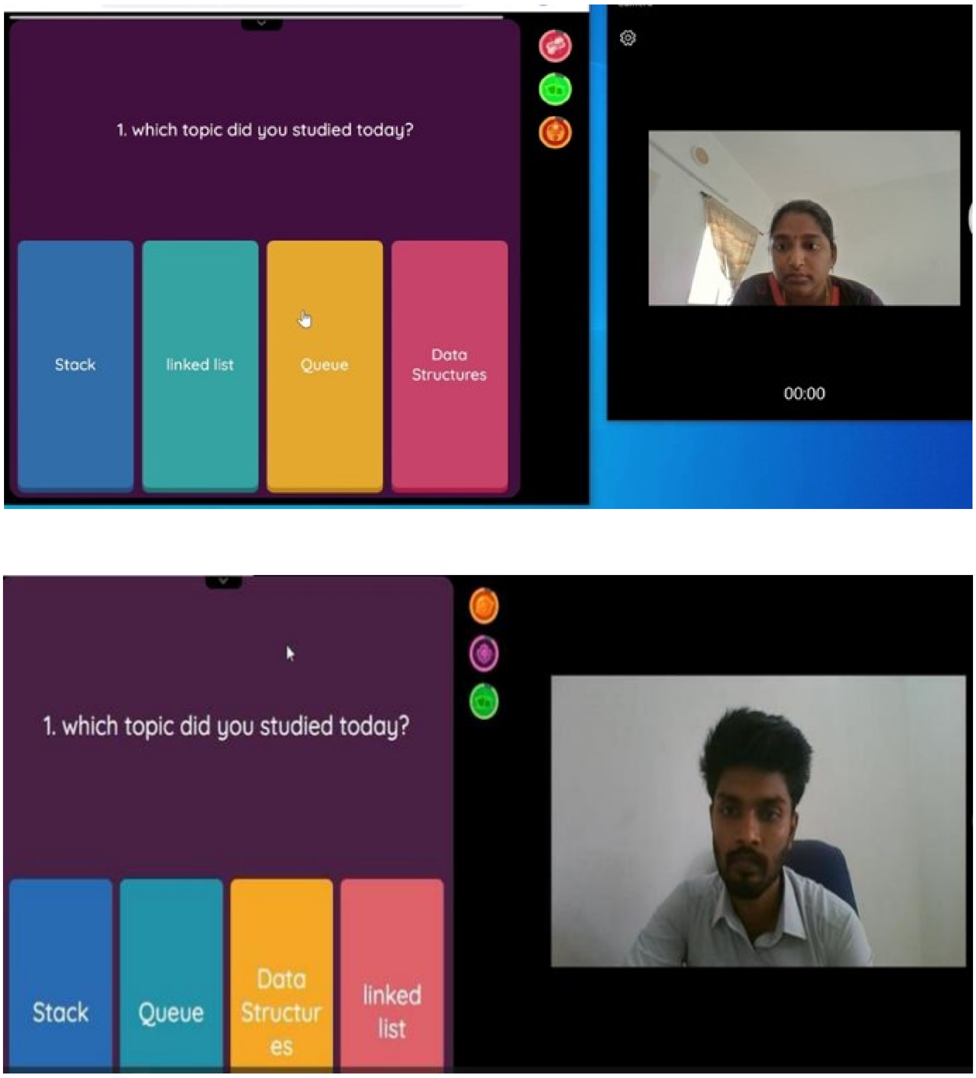


Figure.6.3.2 Video frame while attending video

In this sample, the students are attending the quiz related to the watched informative video. Here also webcam will record the faces of the students as shown in the figure.6.3.2. Here we have taken 10 questions under Data Structures. The time limit for attending each quiz is about 30 seconds. It includes the basics questions relevant to the topic.

6.2 FACE AND EYE DETECTION USING HAAR CASCADE CLASSIFIER

Using Haar cascade classifier, face and eye regions are detected from the input video frames. The below given figure 6.4.1 is the sample video frames of the Emodetect dataset.

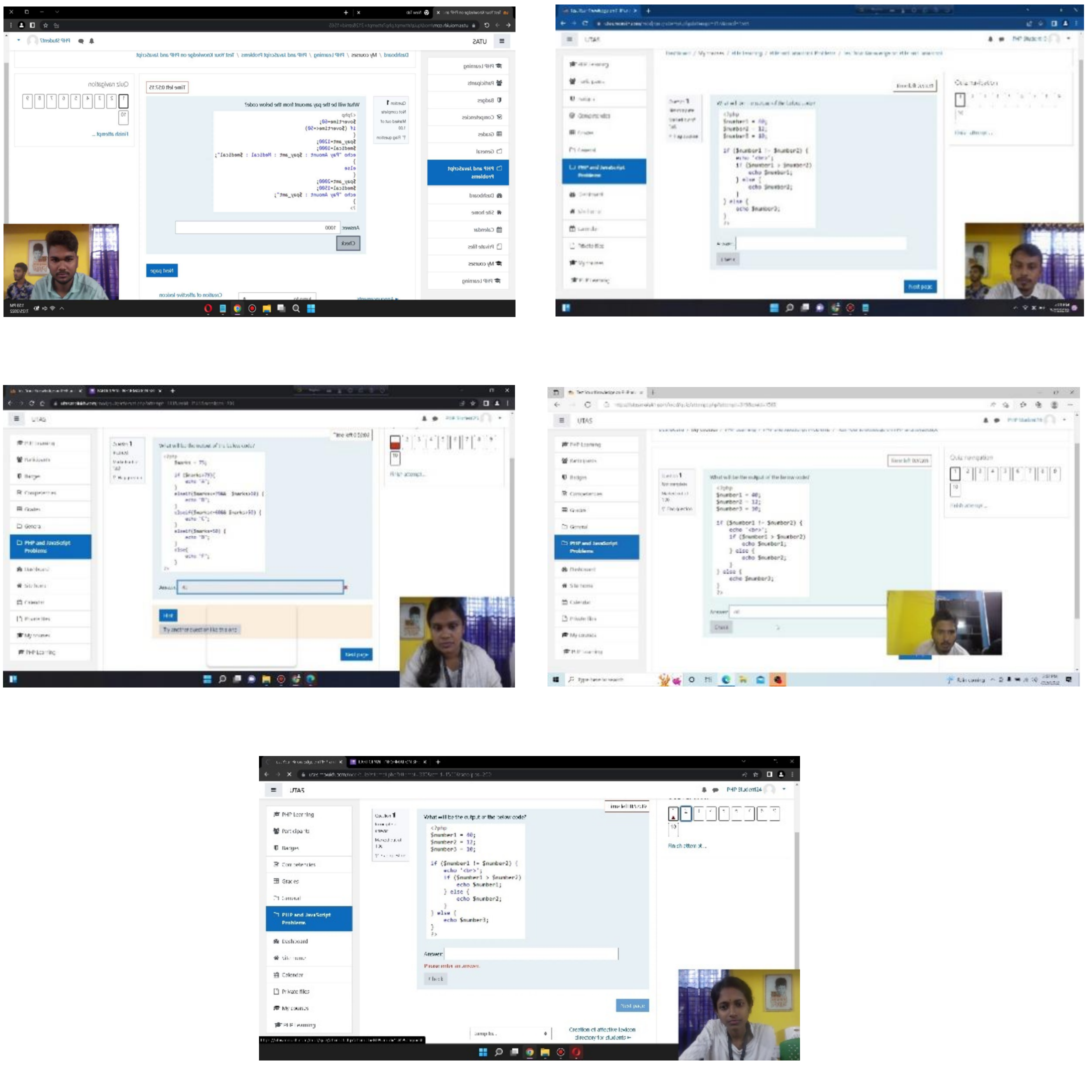


Figure.6.4.1 Video frames of Emodetect dataset

The below figure 6.4.2 is the sample video frames of the Auto-Engage-Detect dataset.

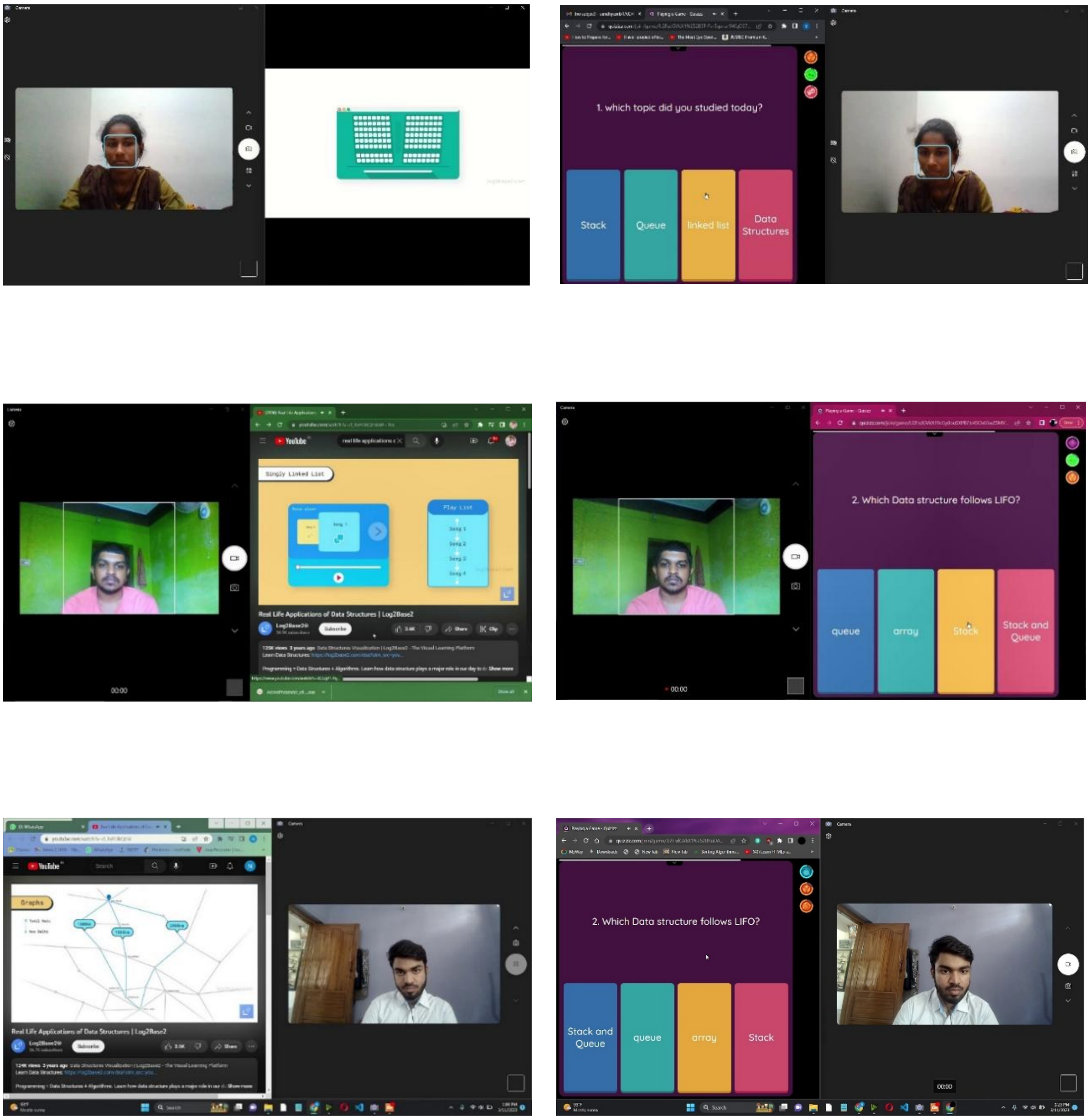
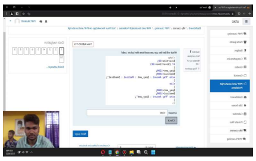
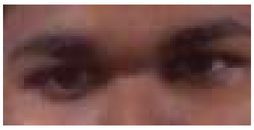
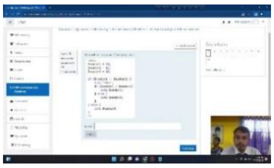
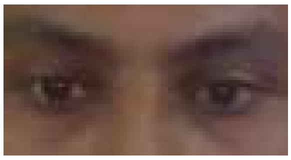
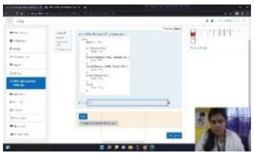
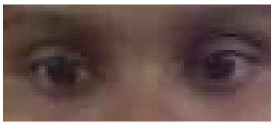
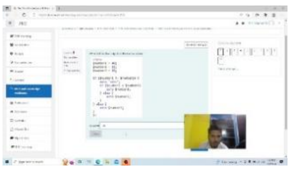
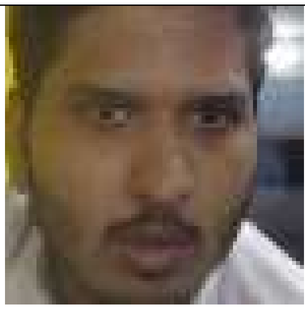
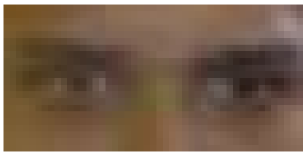


Figure.6.4.2 Extracted video frames for listening and attending quiz

The below tables 6.1 and 6.2 represents the results of face and eyes detection for Emodetect and Auto-Engage-Detect dataset.

Table 6.1 face and eyes detection for Emodetect dataset

Sample no	Input frames	Detected faces	Detected eyes
Sample 1			
Sample 2			
Sample 3			
Sample 4			

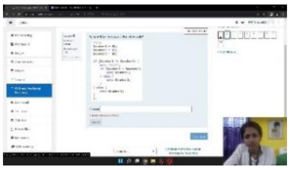
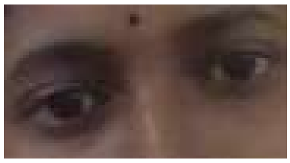
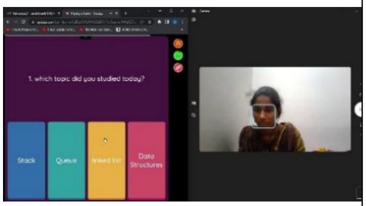
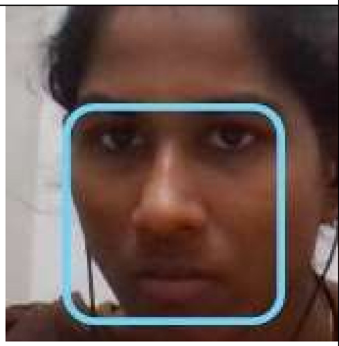
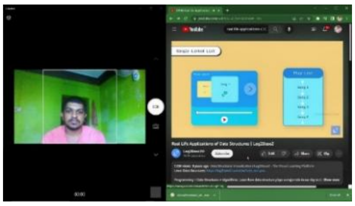
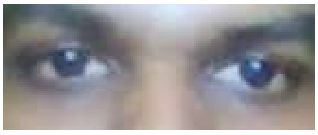
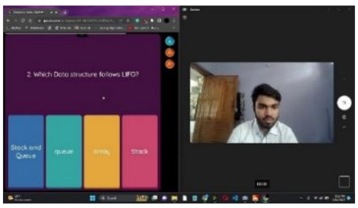
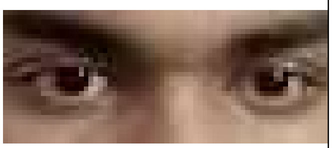
Sample 5			

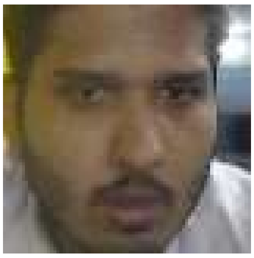
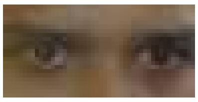
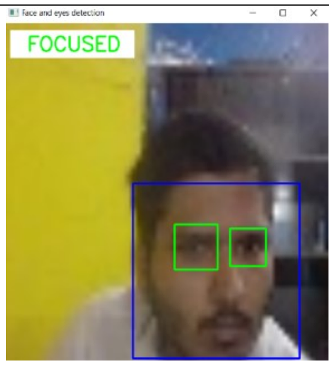
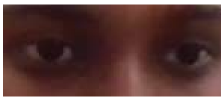
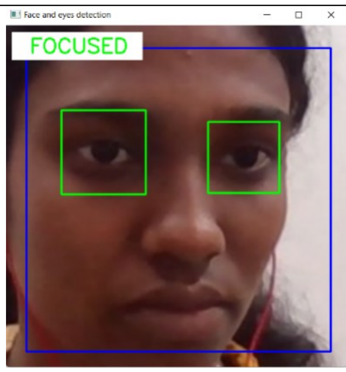
Table 6.2 face and eyes detection for Auto-Engage-Detect dataset

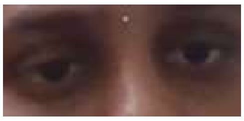
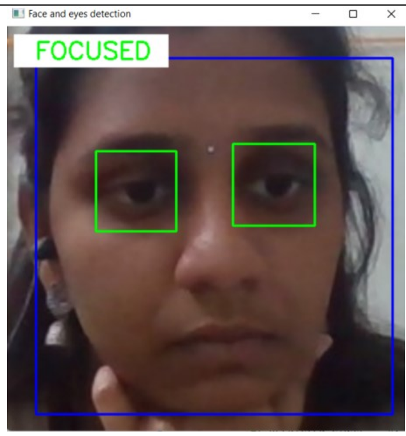
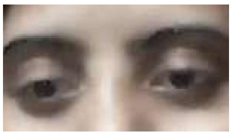
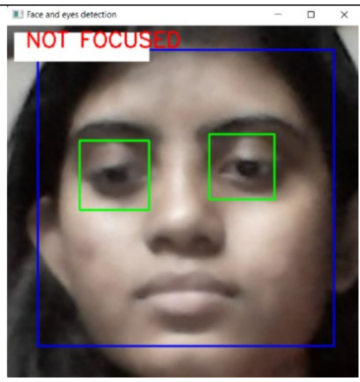
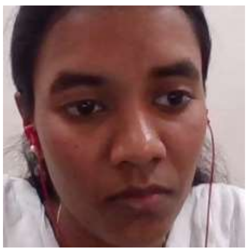
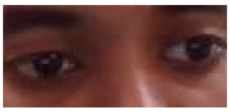
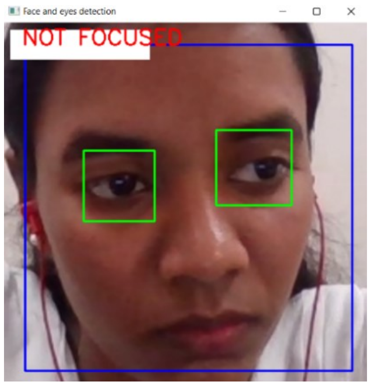
Sample no	Input frames	Detected faces	Detected eyes
Sample 1			
Sample 2			
Sample 3			

6.3 FOCUS DETECTION USING CNN

Using convolutional neural network, FER 2013 dataset images are trained and tested with Emodetect and Auto-Engage-Detect dataset images for focus detection. From this model, it will classify whether the student is focused or not-focused. Table 6.3 denotes the result of focus detection for both the datasets.

Table 6.3 Focus detection

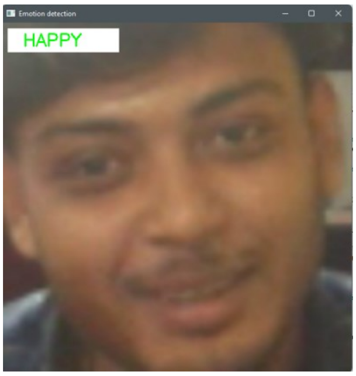
Sample no	**Detected faces**	**Detected eyes**	**Output**
Sample 1			
Sample 2			

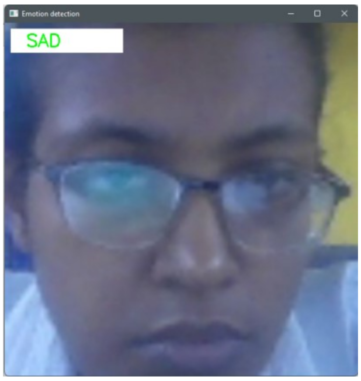
Sample 3			
Sample 4		-	
Sample 5			
Sample 6			

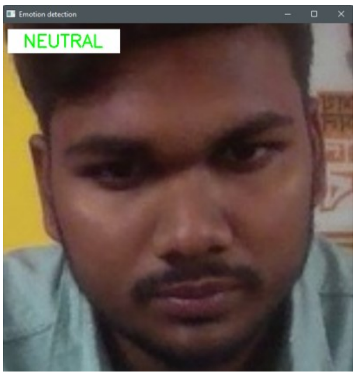
Based on their eye focus, their focus will be predicted. If they looking towards the screen over the time during the session then they are grouped under focused category. If their eyes are looking somewhere or below the screen or closed then they are categorized as not focused students.

6.4 EMOTION DETECTION USING MOBILENET

Video frames have been taken as input for Haar cascade classifier. It will detect the faces from the video frames. Those detected face images are fed into the mobilenet architecture as input in order to predict their emotions. The below images are sample outcomes of mobilenet architecture which shows the dominant as well as multiple emotions expressed by some students.







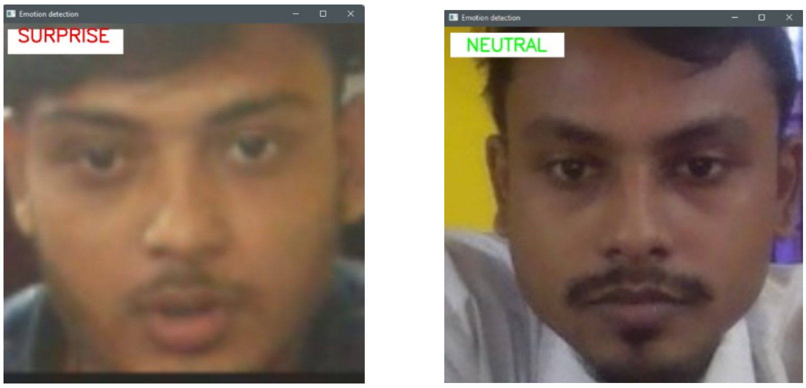


Figure 6.5 Student emotion detection

Images with various emotions such as angry, disgust, fear, happy, neutral, sad and surprise are used to train the model. A student can express multiple emotions as shown in figure 6.5. They cannot stay in same emotion throughout the session. Therefore, dominant emotion will be predicted that the student expressed for most of the time during the session.

6.5 EMOTION GROUP DISTRIBUTION AND WEIGHT CALCULATION

After the prediction of emotion for each student, they will be grouped according to their emotion which they have expressed the most of the time. Then mean score will be computed for each grouped student based on the score which they have achieved in the quiz.

The formula for calculating the mean score achieved in a quiz for each emotion group,

$$\text{Mean score, } EW_g = \text{Total score}_g \,/\, N_g$$

where,

EW_g -> Emotion weight belonging to a group

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

N_g -> The total number of students who took the quiz.

The below table 6.4.1 represents the dominant emotion and emotion weight for each emotion category :

Table 6.4.1. Emotion weight for Emodetect dataset

Dominant Emotion	Emotion Weight
Anger	0.44
Disgust	-
Fear	0.45
Happy	0.483
Neutral	0.425
Sad	0.942
Surprise	0.45

The above table represents the emotion weight for the grouped students in Emodetect dataset. In which the dominant emotion is the overall emotion that the student expressed for most of the time throughout the session. The emotion weight is calculated by determining their average score which is achieved in the quiz for each group.

Table 6.4.2. Emotion weight for Auto-Engage-Detect dataset

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

The above table 6.4.2 represents the emotion weight for the grouped students in our Auto-Engage-Detect dataset. In which the dominant emotion is the overall emotion that the student expressed for most of the time throughout the session. The emotion weight is calculated by determining their average score which is achieved in the quiz for each group.

6.6 CONCENTRATION INDEX CALCULATION

Based on the concentration index, engagement level of the student will be computed. It represents at what level the student has involved in the online class.

The resulting Concentration Index (CI) is determined by multiplying the Dominant Emotion Probability (DEP) value by the corresponding Emotion Weight (EW) for each grouped student, which is expressed by equation:

$$CI = DEP \times EW_g$$

The below tables 6.5.1 and 6.5.2 clearly represent the calculated CI value for both Emodetect and Auto-Engage-Detect dataset students.

Table.6.5.1.CI for Auto-Engage-Detect dataset

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

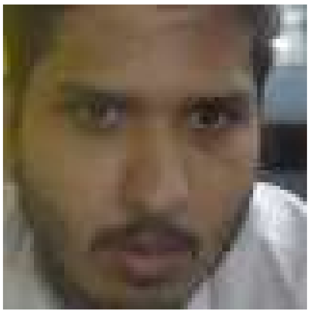
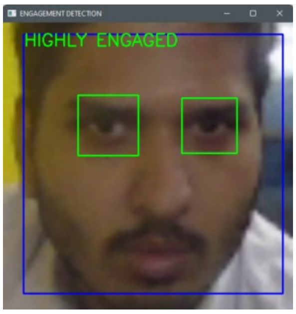
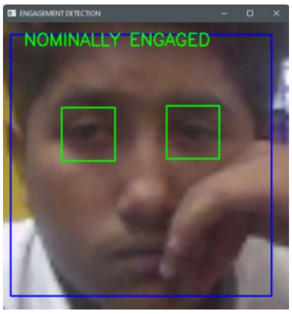
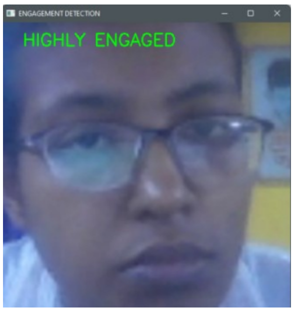
Table.6.5.2 CI for Emodetect dataset

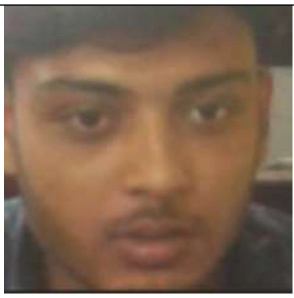
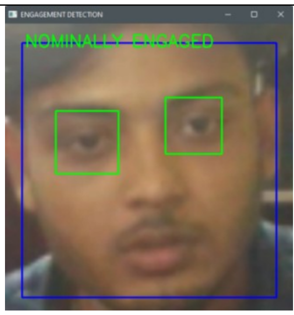
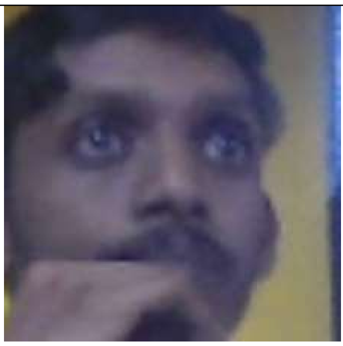
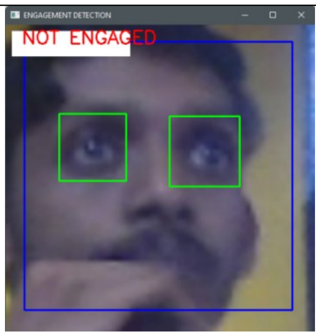
s_id	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	Tot_CI
S2	34.2	22.8	36.3	16	17.5	10	28.5	26.7	22	14.5	22.9
S3	16.2	16.2	16.2	16.2	16.2	16.2	16.2	26.9	16.2	162	17.2
S4	14.6	27.7	27.6	23.3	24.9	25.8	25.8	23.3	14.6	23.3	23.1
S5	20.5	23.8	23.8	12.7	23.8	23.8	20.2	13.7	25.2	20.5	20.8
S6	34.9	14.9	12.9	27.2	12.9	15.5	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

ENGAGEMENT CLASSIFICATION

Based on the calculated concentration index value, engagement level of the students are classified for Emodetect dataset as shown in Table 6.6.1.

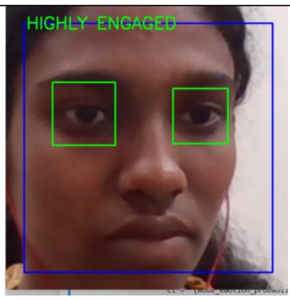
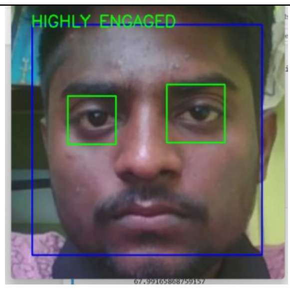
Table 6.6.1 Engagement classificationfor Emodetect dataset

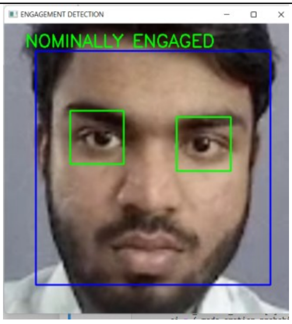
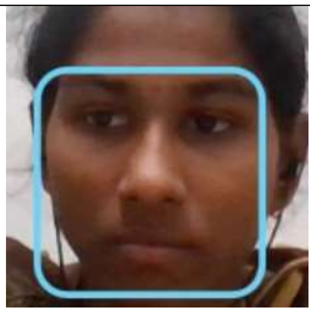
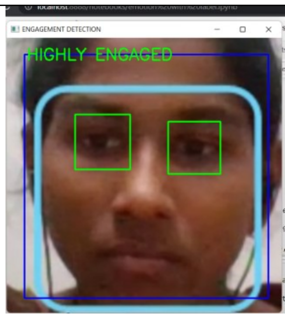
Sample no	**Input images**	**Predicted output**
Sample 1		
Sample 2		
Sample 3		

Sample 4		
Sample 5		

The Table 6.6.2 shows the level of engagement of the students based on the calculated concentration index value for Auto-Engage-Detect dataset.

Table 6.6.2 Engagement classification for Auto-Engage-Detect dataset

Sample no	Input images	Predicted output
Sample 1		
Sample 2		

Sample 3		
Sample 4		
Sample 5		

CHAPTER 7

CONCLUSION

7.1 CONCLUSION :

In this project, Convolution Neural Network (CNN) is used for focus detection. We have used Mobilenet architecture for Emotion detection. For these methods, face and eyes of the student are needed as input. We used Haar cascade classifier algorithm to detect their face and eyes. Then Emotion-wise students have been grouped and their emotion weight have computed using the emotion label taken from the mobilenet and the mean score achieved in the quiz. Using the determined emotion weight and Dominant Emotion Probability (DEP) which is generated from the mobilenet, Concentration Index (CI) have been calculated which is greater than or equal to zero. The engagement level of the student is obtained based upon the CI value.

7.2 FUTURE ENHANCEMENT :

In the future, it may be worthwhile to consider environmental factors that could affect a student's learning experience, such as their head position, lighting conditions, and health monitoring. Additionally, other variables like the learner's age, location, demographics, teaching methods, course design, and course content could also be explored. To better understand digital learning, it would be beneficial to investigate the reasons why students engage or disengage and how these factors are related to the learning process.

7.3 SOCIAL IMPACT :

There is a significant social impact in focus detection using eyes, emotion detection and engagement classification in various fields such as education, healthcare and business.

Education field :

Improved teacher-student relationships: The teachers can provide additional support and resources to students who are struggling to keep up with the course.

Increased student engagement: If teachers can detect when students are disengaged or distracted during online classes, they can take steps to re-engage them.

Early detection of mental health issues: It provides an opportunity to detect signs of mental health issues such as anxiety or depression. Teachers then support the student to overcome the issue.

7.4 APPLICABILITY OF THE PROJECT :

The applicability of this project work is in the field of online education, specifically in monitoring and improving student engagement in virtual learning environments. The use of techniques such as the Haar cascade algorithm and deep learning models for facial expression analysis enables the real-time detection of student engagement levels, which can be used to adapt the delivery of content and provide feedback to educators on the effectiveness of their teaching methods.

The project work can be applied in various online learning platforms, such as e-learning platforms, video conferencing tools, and online classrooms. It can be used by educators to monitor the engagement levels of their students during online classes, assess the effectiveness of their teaching methods, and provide personalized feedback and support to individual students.

The applicability of this project work is significant, as it addresses a critical issue in online education and provides a practical solution for improving student engagement and learning outcomes in virtual learning environments.

APPENDIX 1

SPECIFICATIONS

A.1.1. HARDWARE SPECIFICATION :

Processor	Intel core i7/i10
RAM	8.00 GB
System type	64- bit Operating system, x64-based processor
System model	DELL Laptop

A.1.2. SOFTWARE SPECIFICATION :

OS	Microsoft windows
Tool	Jupyter notebook, Google colab

APPENDIX II

SOURCE CODE

Sample source code:

```
# frame extraction and face detection
import cv2
import os
cap = cv2.VideoCapture("D:\\utube\\utube\\student19 ancy.mp4")
fps = cap.get(cv2.CAP_PROP_FPS)
frames_per_minute = int(fps * 2.4)
minute_count = 0
frame_count = 0
frames_folder = "D:\\learning_frames_faces_eyes\\own_dataset_frames\\ANCY\\"
os.makedirs(frames_folder, exist_ok=True)
output_folder_eyes =
"D:\\learning_frames_faces_eyes\\survey_eyes\\survey_eyes\\ANCY\\"
output_folder_faces="D:\learning_frames_faces_eyes\\survey_faces\\survey_faces\\ANCY
\\"
os.makedirs(output_folder_faces, exist_ok=True)
os.makedirs(output_folder_eyes, exist_ok=True)
face_cascade=cv2.CascadeClassifier(cv2.data.haarcascades+'haarcascade_frontalface_defa
ult.xml')
eye_cascade = cv2.CascadeClassifier(cv2.data.haarcascades + 'haarcascade_eye.xml')

#focus detection CNN architecture
from keras.models import load_model
model = load_model("C:\\Users\\DELL\\Downloads\\code\\f1.h5")
# Evaluate the model on the testing data
score = model.evaluate(test_generator)
print('Test loss:', score[0])
print('Test accuracy:', score[1])
#testing with sequence of frames
```

```
folder_path = "D:\\learning_frames_faces_eyes\\survey_eyes\\survey_eyes\\ANCY\\"
class_names = ['focused', 'not-focused']
# Loop over each image file in the folder
predicted_class = np.argmax(result)
print(f'Predicted class: {predicted_class}')
# Print the predicted class
if predicted_class == 0:
    print("focused")
    detect_face(0)
else:
    print("not-focused")
    detect_face(1)

#Dominant emotion and emotion probability using mobilenet
# Load the pre-trained model
model = load_model("C:\\Users\\DELL\\Downloads\\code\\new.h5")
folder_path = "D:\\quiz_faces\\fmadhu\\"
emotions = ["anger", "disgust", "fear", "happy", "neutral", "sad", "surprise"]
        # Find the mode emotion for the subfolder
        mode_emotions = [emotions[i] for i in y_preds]
        mode_emotion = mode(mode_emotions)
        mode_emotion_probability = max(emotion_probabilities[mode_emotion])
        mode_emotion_probabilities.append(mode_emotion_probability)
        mode_emotion_index = emotions.index(mode_emotion)
        mode_emotion_indices.append(mode_emotion_index)
        print(f"{subfolder}: Mode emotion: {mode_emotion}, Mode emotion probability: {mode_emotion_probability}")
        if mode_emotion=='angry':
            detect_faces('angry')
        elif mode_emotion=='disgust':
            detect_faces('disgust')
        elif mode_emotion=='fear':
            detect_faces('fear')
```

```
        elif mode_emotion=='happy':
            detect_faces('happy')
        elif mode_emotion=='neutral':
            detect_faces('neutral')
        elif mode_emotion=='sad':
            detect_faces('sad')
        else:
            detect_faces('surprise')

#Concentration Index Calculation
CI=[]
for i in range(len(mode_emotion_probabilities)):
    if mode_emotion_indices[i] == 0:
        ci = (mode_emotion_probabilities[i] * 0.5) * 100
        CI.append(ci)
    elif mode_emotion_indices[i] == 1:
        ci = ( mode_emotion_probabilities[i] * 0) * 100
        CI.append(ci)
    elif mode_emotion_indices[i] == 2:
        ci = ( mode_emotion_probabilities[i] * 0.12) * 100
        CI.append(ci)
    elif mode_emotion_indices[i]== 3:
        ci =  (mode_emotion_probabilities[i] * 0.68) * 100
        CI.append(ci)
    elif mode_emotion_indices[i]== 4:
        ci = ( mode_emotion_probabilities[i] * 0.75) * 100
        CI.append(ci)
    elif mode_emotion_indices[i]== 5:
        ci = ( mode_emotion_probabilities[i] * 0.628) * 100
        CI.append(ci)
    else:
        ci = ( mode_emotion_probabilities[i] * 0.628) * 100
        CI.append(ci)
```

```
#Engagement Classification
if avg_CI>50:
    print("HIGHLY ENGAGED...!!!")
    classification(avg_CI)
elif avg_CI>1 and avg_CI<=50:
    classification(avg_CI)
    print("NOMINALLY ENGAGED...!!!")
else:
    classification(avg_CI)
    print("NOT ENGAGED...!!!")
```

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]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/s12528-01

Ouriginal
by Turnitin

Document Information

Analyzed document	Class Engagement Monitoring Report.docx (D163810227)
Submitted	2023-04-13 09:12:00
Submitted by	
Submitter email	svanitha@mepcoeng.ac.in
Similarity	7%
Analysis address	svanitha.mepco@analysis.urkund.com

Sources included in the report

SA	**VTD640 Kujani 1st Page Abstract and all chapters.pdf** Document VTD640 Kujani 1st Page Abstract and all chapters.pdf (D159161165)	2
SA	**report on 6-4-23.pdf** Document report on 6-4-23.pdf (D163287784)	7
W	URL: https://iq.opengenus.org/mobilenet-architecture/ Fetched: 2021-06-17 09:32:23	1