JACK N. AVERITT COLLEGE of GRADUATE STUDIES
GEORGIA SOUTHERN UNIVERSITY

Georgia Southern University
Digital Commons@Georgia Southern

Electronic Theses and Dissertations

Graduate Studies, Jack N. Averitt College of

Spring 2021

Unobtrusive Assessment Of Student Engagement Levels In Online Classroom Environment Using Emotion Analysis

Sasirekha Anbusegaran

Follow this and additional works at: https://digitalcommons.georgiasouthern.edu/etd

 Part of the Robotics Commons

Recommended Citation

Anbusegaran, Sasirekha, "Unobtrusive Assessment Of Student Engagement Levels In Online Classroom Environment Using Emotion Analysis" (2021). *Electronic Theses and Dissertations*. 2265.
https://digitalcommons.georgiasouthern.edu/etd/2265

This thesis (open access) is brought to you for free and open access by the Graduate Studies, Jack N. Averitt College of at Digital Commons@Georgia Southern. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact digitalcommons@georgiasouthern.edu.

UNOBTRUSIVE ASSESSMENT OF STUDENT ENGAGEMENT LEVELS IN ONLINE CLASSROOM ENVIRONMENT USING EMOTION ANALYSIS

by

SASIREKHA ANBUSEGARAN

(Under the Direction of Andrew Allen)

ABSTRACT

Measuring student engagement has emerged as a significant factor in the process of learning and a good indicator of the knowledge retention capacity of the student. As synchronous online classes have become more prevalent in recent years, gauging a student's attention level is more critical in validating the progress of every student in an online classroom environment. This paper details the study on profiling the student attentiveness to different gradients of engagement level using multiple machine learning models. Results from the high accuracy model and the confidence score obtained from the cloud-based computer vision platform - Amazon Rekognition were then used to statistically validate any correlation between student attentiveness and emotions. This statistical analysis helps to identify the significant emotions that are essential in gauging various engagement levels. This study identified emotions like calm, happy, surprised, and fear are critical in gauging the student's attention level. These findings help in the earlier detection of students with lower attention levels, consequently helping the instructors focus their support and guidance on the students in need, leading to a better online learning environment.

INDEX WORDS: Online learning, Convolutional neural network, Extreme gradient boost, Amazon rekognition, Machine learning, Emotion affect

UNOBTRUSIVE ASSESSMENT OF STUDENT ENGAGEMENT LEVELS IN ONLINE CLASSROOM ENVIRONMENT USING EMOTION ANALYSIS

by

SASIREKHA ANBUSEGARAN

B.Tech., Pondicherry University, India, 2014

M.B.A, Pondicherry University, India, 2017

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

©2021

SASIREKHA ANBUSEGARAN

All Rights Reserved

UNOBTRUSIVE ASSESSMENT OF STUDENT ENGAGEMENT LEVELS IN ONLINE CLASSROOM ENVIRONMENT USING EMOTION ANALYSIS

by

SASIREKHA ANBUSEGARAN

Major Professor:	Andrew Allen
Committee:	Gursimran Singh Walia
	Lixin Li

Electronic Version Approved:
May 2021

DEDICATION

I dedicate my work to my husband, my parents, my sister, and my family for supporting me throughout my grad school, whose words of encouragement and push for tenacity helped me a lot during hard times. A special feeling of gratitude to my husband, Mr. Giridharan Munuswamy for his endless support and cheering. Without him, this work would not have come to fruition.

ACKNOWLEDGMENTS

I would like to sincerely thank Dr. Andrew Allen for his guidance throughout these years and for being such a wonderful mentor who patiently guided me through every step of my thesis work. I would also like to thank Dr. Gursimran Walia and Dr. Lixin Li for accepting to be part of my committee and for their support throughout the project. Special thanks to Dr. Muralidhar Medidi for his support and encouragement during difficult times in the course of learning.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

This chapter will cover the purpose of the research conducted for this thesis. It will concisely describe the research history and significance of the field and explain the growing importance of this study. It will also familiarize the focus of the problem and the contributions of this research.

1.1 MOTIVATION

Emotions are of great significance in education in all aspects of human life. It is universally acknowledged that emotions do exist and are evaluated. In the modern education system, student engagement is an important concept, and how much information the student receives is equally important in learning.

The development of advanced teaching techniques coupled with increased computing power has explored and resolved many research problems related to student engagement in the traditional classroom environment and obtained positive results. A typical in-person classroom model helps students expand their focus, refine their critical thinking, and reinforce their significant learning experience; despite these benefits, current world events have forced the students to adapt to the online classroom model.

Consequently, the research dimension has also progressed towards the problems and challenges faced by the students during the synchronous online classes. Online learning gradually gained its relevance in recent years and has become a mandatory method of uninterrupted learning during any crisis. Knowing the students' attentiveness level in the online classroom environment is crucial for designing an adaptive learning system. Emotions and facial expressions are substantial cues used by the instructors to identify a student's attention level, but this is not possible when the learning happens in a digital environment.

As online learning and synchronous online classes have become a way of education in

recent days due to the COVID-19 pandemic, recognizing students' attention level with the system they are interacting with can alter how any instructor interacts with their students. Identifying student attention levels will lead to a better understanding of their engagement with the system and pave the way for adjusting teaching strategies. Also, it helps in identifying and categorizing the students based on their attention level. The success of online classes hinges on the outcomes of student's knowledge and results related to their engagement.

Other research in this area focuses on detecting the student's different emotions (happy, sad, angry, confused, disgusted, surprised, calm, neutral) during lectures, labs, and research conducted in classes. Most of the recent research in this domain was primarily focused on measuring just the student's emotional state. Such studies are limited in their usefulness for the instructors due to the absence of any correlation model between student's engagement level in class and their emotional state.

Hence, in order to keep it less complicated for the instructors, research was carried out to learn whether a student is attentive or not during class (binary classification on attentiveness). As it's always helpful to know about students being either attentive or inattentive, but most often, students aren't at these disparate ranges. Practically, a student can be partially attentive too during lectures. Hence a student's attentiveness level may not possibly be constrained as either 0 or 1 at all the time.

As there are various levels of attentiveness, the instructor may support the students in different ways based on each level. For example, a highly attentive student may not require any additional resources, a less attentive student may require minimal support from the instructor for learning. In contrast, the disengaged set of students may require more help and guidance to improve their learning process.

1.2 CONTRIBUTION

In this thesis, we expanded the research to determine if there are multiple categories to classify student engagement, thereby using a multi-level classification of student attentiveness level (attentive, partially attentive, inattentive) in an online classroom environment. The advantage of this method is that it helps instructors identify the inattentive students and students who are partially attentive at an early stage and provide the required guidance leading to a better online learning environment.

We proposed a system architecture using a combination of multiple machine learning techniques and leveraging a cloud-based computer vision service. Machine learning techniques are used to establish the prediction model for a different level of student's attentiveness. Cloud-based computer vision service is used to establish different emotional states of the students. A statistical model is then developed to correlate the emotional states with the student's attentiveness level.

The first outcome is the results obtained from the two popular machine learning models – Convolutional Neural Network (CNN) and Extreme Gradient Boost (XGB). These models were used for recognizing student engagement based on their facial expressions. Of both models, the highest average accuracy of 91.4% was achieved by CNN, indicating that it is certainly possible to generate a predictive model for different student engagement levels through information obtained from a recorded video.

The second outcome is from an experimental procedure that includes a composite model that performs equally well compared to the performance of the XGB technique. This method was used to observe how well machine learning techniques performed with well-optimized features rather than regular inputs.

The final outcome discusses the significance between emotion analysis and the predictive model of student attention levels in an online class setting by performing regression analysis.

The rest of the thesis is organized as follows. Chapter 2 discusses the literature review performed. Chapter 3 details the methodology of the thesis work. Chapter 4 focuses on the results obtained. Chapter 5 summarizes and concludes the thesis. Chapter 6 highlights the limitations and future research directions.

CHAPTER 2

RELATED WORKS

Most of the research associated with measuring student attentiveness uses numerous possibilities at various environmental setups. Generally, the experimental study of gauging various engagement levels was conducted from a traditional classroom environment to an independent learning setting. Here we will discuss some of the researches performed in multiple domains.

In classroom setup, monitoring the student learning process and providing feedback to the teachers is the recent advancement in automated learning analytics. This concept of real-time feedback is made possible by using kinetic data obtained from Kinect One sensor device to build the feature set. This study compared seven different classifiers to predict student attention over time and their average attention levels (Janez Zaletelj et al. 2017).

A model was presented to detect student emotions from student interaction with a cognitive tutor for mathematics. Cognitive tutors were designed to operate based on the student's action within the user interface. Log data was collected from the software, and observations were conducted in the school's computer laboratory. The classification algorithms like decision tree, step regression, naïve bayes were used to analyze the data collected. The detectors validated on re-sampled data achieved an accuracy of 19% more than the established base rate (Ryan SJ et al. 2012).

A research study was conducted to improve student's involvement in E-learning platforms by using their facial features to extract mood patterns. The study helps to assess and identify lapses in sustained attention by a student in an E-learning session. Analyzing the moods based on the emotional states of a student during an online lecture provided results that could readily be used to improve the efficacy of the content delivery mechanism within the E-learning platform. The study investigates whether facial expressions are the most critical means of nonverbal expression and list the most common facial features that

describe a student's involvement in a lecture. A neural network approach was used to train the models like the radial-based Neural Network (NN) model, Hidden Markov Model, and Support Vector Machine (SVM). The result shows a high correlation with the feedback and a success rate of over 70% in assessing the student's mood (Abdulkareem Al-Alwani et al. 2016).

In the early years, researchers organized a relationship between visual attention and saccadic eye movement (Deubel and Schneider 1996), where they used the Viola-Jones algorithm to detect facial images (Dingus, Hardee, and Wierwille 1987). Support Vector Machine (SVM) was availed to classify the actions of eye movements. These classic concepts were utilized as base ideas in building various machine learning techniques.

Video Tutoring System with automated Facial Expression Recognition (VTSFER) was developed to determine the performance level of the students. This research studied the effect of facial expression algorithms by testing two groups of students taking the same computer programing course. One group of students were subjected to video tutoring with VTSFER software, and the rest were subjected to traditional video tutoring without any facial expression recognition software. VTSFER software analyzed the students' facial patterns and suggested students with negative emotions to re-watch the video. The approach using VTSFER software achieved an increased average performance rating of 72.2% compared to the traditional video tutoring method (Christopher John R. Llanda 2019).

The research from the above study was expanded from the learning domain to the general public domain. A deep learning-based system was presented for monitoring customer behavior, specifically for the detection of their interest. For those customers whose heads were directed towards the advertisement or the product of their interest, the system further evaluates their facial expressions and reports the customer's interest. A webcam was used for head pose estimation and facial expression recognition. A multi-task 3-cascade CNN

model was used for the system. This system achieved 99.90% accuracy for head pose estimation and 94.61% accuracy for facial expression recognition (Gozde Yolcu 2019).

When analyzing emotions, it is not always essential to use machine learning concepts for evaluation. We can also experiment with various associated software related to it. A study was conducted to evaluate the synchronization of three emotional evaluation methods (automatic facial expression recognition, self-report, electrodermal activity) and their convergence regarding learner's emotions. FaceReader 5.0 was used to collect the learner's physiological arousal data using Affectiva's Q-Sensor 2.0 electrodermal activity measurement bracelet. The outcomes defined a high-level agreement between the self-report and facial recognition modalities by up to 75.6% and a low level of concurrence between electrodermal activation and other modalities (Jason M. Harley 2015).

The acceptance and cultural differences in facial expression of emotions fall into seven categories: happiness, sadness, anger, fear, surprise, and disgust, according to Ekman et al. (1987). The authors proposed the Facial Action Coding System (FACS), which encrypts facial expressions in terms of atomic facial actions called Action Units (AUs). AUs can be measured by identifying the affective states and emotions. Ekman's work encouraged many researchers to develop automated multi-facial emotion recognition approaches. In a study conducted by Grafsgaard et al. (2013), it was recognized that AU2 was negatively correlated with the learning gain factor, whereas AU4 was positively correlated with frustration emotion. AU14 was positively correlated with both frustration and learning gain. Computer Expression Recognition Toolbox (CERT) has been used in numerous research studies for engagement recognition in the learning environment. CERT gives confidence values for facial AUs from a wide range of FACS which authorizes fine-tuned analyses for gathering affective states of learners using facial expression analysis. The authors further expanded their research by studying facial movements consisting of raising eyebrows, lowering eyebrows, tightening the eyelid, and dimples in the mouth during video tutoring using

CERT. In this study, upper facial movements were used for predicting the facial expression for engagement, frustration, and learning. Dimples in the mouth were found to be a positive predictor of learning and self-reported performance. The authors confirmed the usage of intensity and frequency of facial expressions to identify the engagement outcomes from a tutoring software system.

A hybrid information system was proposed by Uğur Ayvaz et al. in 2017 for visual and interactive learning systems. It detects the emotional state of learners and gives feedback to an educator based on facial expression. It helps an educator be aware of the general emotional state of the students in the virtual classroom system. It used classification algorithms, including random forest and regression trees, which were applied to learn the emotional states of the learners. Skype was used as the preferred test platform to evaluate the learner's emotion within the virtual classroom. The best accuracy rates were obtained by K-Nearest Neighbor (KNN) algorithm with 96.38% and the SVM algorithm with 97.15%.

A similar study conducted by Mohammed Megahed and Ammar Mohammed in 2020 presented a composite intelligent technique that integrates a deep learning network and fuzzy technique. A loosely coupled integration was adopted to build the CNN framework. CNN was used to detect the learner's facial expression, and the fuzzy technique was used to detect the advanced learning level based on extracted facial expression states from CNN. Online video was streamed during the test and exam session. CNN was used to model, analyze and classify learner's facial expressions that reflect their emotional states, whereas the fuzzy part is responsible for handling uncertainty from the learning environment. Based on the CNN findings, the outcome of this analysis provides answers to the validity ratio, elapsed test time, and current learning level. The results from this system helped the decision-makers (for example, teachers/lecturers) with graphs explaining the learning flows and emotional states of participants in the learning activity.

Developing an intelligent tutoring system is also another avenue that is commonly

pursued. Turgay Celik et al. in 2017, developed the Witwatersrand Intelligent Teaching System (WITS) that aims to assist lecturers with real-time feedback regarding student affect. Student engagement is labeled based on their behavior and postures in the classroom. A database was constructed using a Histogram of Oriented Gradient features in conjunction with Support Vector Machine (HOG SVM). A classifier was built to recognize these proxies over the dataset. AlexNet algorithm was developed using CNN architecture to predict student interest levels based on data collected in the classroom. This algorithm performed exceptionally well on benchmark databases. Also, a Compute Unified Device Architecture (CUDA) - enabled GPU hardware was used to outperform the parallel CPU implementation of HOG feature extraction and SVM classification with the library for Support Vector Machines (libSVM). Cross-validation on a random subset of frames and cross-subject validation was performed to evaluate the outcomes. Both experiments showed CNN architecture significantly outperforms an SVM trained on HOG SVM features in terms of actual accuracy on the data set and generalization capabilities. Similarly, Khelfallah et al. in 2015, proposed an intelligent tutoring system called Remote Laboratory, which allowed learners from anywhere to access the internet and conduct computational experiments with real laboratory equipment. The attention levels of the learners were examined with respect to their emotions like frustration and serenity using 70 small classifiers.

Features extracted from the facial recognition system with two-dimensional and three-dimensional data were combined with different observational cues to detect students' attention levels. Frank et al. in 2016 proposed a framework for attention recognition which includes facial expression, speech, body postures, and motion using different dimensional sensors. An SVM classifier was used to detect various engagement levels such as disengagement, relaxed engagement, involved engagement, intention to act, action, and involved action. This methodology was applied during public gatherings like meetings to detect the participants' engagement levels.

Research studies incorporating spatial and temporal information into different deep networks have come to the fore in recent years. A multi-channel deep spatial-temporal feature fusion neural network (MDSTFN) was proposed by Huan.Y Du et al. in 2017. The temporal information explains the optical flow from the peak expression facial image and the neutral facial image. In contrast, spatial information describes the gray-level image of an emotional face.

Similar to previous research, Zhang et al. in 2020 developed a framework that combines double-channel Weighted Mixture Deep Convolution Neural Networks (WMDCNN) and double-channel Weighted Mixture Deep Convolution Neural Networks with Long Short-Term Memory Network (WMCNN-LSTM) on image sequences. WMDCNN network identifies facial emotion and provides static image features for the WMCNN-LSTM network. Later, those static image features are used to gain the image sequences' temporal features to precisely recognize the facial expression. The authors also proposed a deep spatial-temporal network as their expanded study. It is a methodology that merges a temporal network for modeling dynamic evolution called Part-based Hierarchical Bidirectional Recurrent Neural Network (PHRNN) and a spatial network for global static features called multi-signal CNN.

CHAPTER 3

METHODOLOGY

3.1 DATASET

The dataset[1] contains 9068 video snippets captured "in the wild" from 112 users using an HD webcam setup for recognizing user affective states, which are raw crowd annotated and associated with a standard annotation built using an expert team of psychologists. Based on the research made by Jacob Whitehill et al. in 2014, each video was 10 seconds long as this duration provided adequate information for the labeling action. Each of the subjects was presented with two different 20 minutes length videos to simulate an E-learning environment. One of the videos was educational, and the other was recreational to capture a focused and relaxed setting. It allows capturing the natural transitions in user attention levels. The students enrolled in this study were between the age group of 18 to 30 years.

As it was structured as an E-learning environment, the videos were captured at various locations like dorm rooms, crowded lab space, and library with various illumination levels (light, dark and neutral). The video dataset was labeled with different affective states like boredom, confusion, engagement, and frustration. Each affect was further categorized into four labels: very low, low, high, and very high.

3.1.1 DATA PRE-PROCESSING

Our research experiment's initial step was to build the dataset with student images captured in their E-learning environment. Image frames were extracted from the video files. As the videos were captured at different locations with varying illumination levels,

1. A. Gupta et al., "DAISEE: Dataset for Affective States in E-Learning Environments," *ArXiv* abs/1609.01885 (2016).

there were challenges setting up the image dataset. The challenges included dark image frames, the student is not within critical proximity of the webcam, and students not within the image frame due to external distractions. Hence, we focused on gathering the data by centralizing and cropping the facial parts at equal pixel size for each frame.

We obtained the facial images using an object detection approach called Haar Cascade Classifier proposed by Paul Viola and Michael Jones (2001) in their research paper. It is a cascade function built using OpenCV as .xml files and trained with many positive images (with face) and negative images (without face). This pre-trained function was then used to detect the object in new images. This function also contained the ability to identify full-body posture, lower-body posture, eye movement, and frontal face.

For our thesis, the file "haarcascade_frontalface_default" was used to detect an individual's face and "haarcascade_profileface" was used to detect image frames with the side profile of the individual. The images were resized to 200*200 pixels, later converted to grayscale, and resized when needed to increase the possibility of prediction accuracy. These normalized images were then converted to flatten array for training different machine learning models. The image dataset was labeled using different engagement levels. This dataset is later shuffled and split to be used for training and validating different machine learning models to be discussed later.

3.1.2 Feature Extraction and Labeling

In order to label the images correctly based on their attentiveness, the extracted facial features for each engagement level should be significant and carefully considered for labeling. There are two scenarios to be considered for labeling the images.

As discussed in Section 3.1, for the first scenario, the video files were recognized based on their affects and given a range of 0 to 3 (very low to very high). Among those, the images from video files with engagement affect at level-3 with other effects at level-0

are labeled as “Attentive.” Similarly, the images from video files with engagement affect at level-2 with other effects at level-0 are labeled “Partially attentive.” Finally, the images from video files with engagement affect at level-0 and level-1 are labeled “Inattentive.”

Highly Attentive	➢ Gazing at the screen ➢ Supporting head ➢ Curious emotion ➢ Smiling/delighted emotion ➢ Facial expressions ➢ Lean forward ➢ Gazing with hand near the mouth
Partially Attentive	➢ Eyes barely open ➢ Eyebrow lowering ➢ Lean forward with eyes barely open ➢ Being neutral ➢ Hand covering the mouth with eyes barely open ➢ Smirky lips ➢ Twitching one eye
Inattentive	➢ Eyes closed ➢ Looking up or above the screen cam ➢ Looking down ➢ Using phone ➢ Yawning ➢ Talking to others ➢ Turning extreme left /right side (not looking at any quadrant of the screen) ➢ Unclear Image ➢ Frame contains no person at all

Table 3.1: Indicators used for multiple engagement levels

As given in Table 3.1, the video files with diverse affective states are considered for the second scenario. Here, the images are labeled based on the carefully considered indicators from facial and behavioral attributes proposed by Lane and Harris in 2015. The authors’ guidelines designed for engagement level classification were modified by adding features of the facial expression, hand gestures, and body postures based on the visual

cues observed from the video files for the above-stated scenario. Still, the attention levels definition remained the same.

The labels are mutually exclusive for our classification problem, where each sample belongs to only one class. Therefore, sparse categorical cross-entropy was used to label the images in the neural network. It reduced the execution time and saved memory space while training the machine learning model. For example, the images will be labeled with numeric values [1] or [2] or [3] instead of [1,0,0] or [0,1,0] or [0,0,1] as in one-hot encoding respectively.

3.1.3 SPLITTING THE DATASET

The dataset consists of 2800 preprocessed images with a dimension of 200*200 pixels. The dataset was shuffled and split into 3 phases: training, testing, and validation sets with a ratio of 70:20:10, respectively. Here the training dataset is used to fit different models using the weights determined by the accuracy and loss function of the prediction algorithm.

A validation set was used to avoid overfitting the network and fine-tune the model's hyperparameters. The model occasionally encounters the trained data values, but it does not adjust its weights. Instead, it helps to condition a stopping point for the back-propagation algorithm. A test dataset was used to evaluate the efficiency of the trained model. The accuracy of each model on the test data provides an unbiased estimate of the model's performance on unlabeled images and confirms the network's predictive power.

The images were distributed into training, testing, and validation set as per the data split ratio shown in Table 3.2. It is vital to create a balanced dataset and distribute it equally; otherwise, the imbalanced dataset is quite challenging for the machine learning models. We may have to use algorithmic-level approaches to overcome data imbalances to fix the misclassification caused between multiple classes.

	Total	**Train set**	**Test set**	**Validation set**
Inattentive	924	627	185	91
Partially attentive	943	668	188	95
Attentive	937	665	187	94
Total	2800	1960	560	280

Table 3.2: Dataset distribution

3.2 Flow Chart of the Architecture

This chapter covers the fundamental architecture of the research flow and the details related to the experiments carried out for this research. Figure 3.1 shows the complete flow diagram of the proposed methodology for the student engagement analysis, including dataset creation, the proposal of models, and statistical analysis. The details of each phase are discussed below.

Each phase occurs sequentially as the subsequent step requires the outcome from the previous step. The preprocessed and labeled data gathered from the data collection phase is passed to the machine learning module and cloud-based computer vision platform called Amazon Rekognition. The machine learning module consists of three models — a deep NN "Convolutional Neural Network (CNN)", XGBoost classifier, and a hybrid model, which is a fusion of CNN and XGB model. The machine learning models' outcome gives the test accuracy and predicts the engagement level of the test data. Meanwhile, the confidence score for eight standard emotions was obtained for every test image using the Amazon Rekognition.

The results from the above two phases were used to evaluate the relationship between the emotion states and the student's attentiveness. A statistical analysis called Multinomial Logistic Regression Analysis (MLR) was used to answer this research question. MLR

provides insight into the statistical correlation between various emotion states and different engagement levels of the student.

3.3 Machine Learning Models Used

This section details the machine learning models used. This phase uses the previously built image dataset from the data collection phase as an input to make predictions on the data. The work intends to tweak the machine learning models and the supporting tools to improve the effectiveness of the teaching/learning process in an online classroom environment by using non-verbal cues. The experiment makes use of three machine learning models, which also include a deep neural network. They are CNN, XGB, and a hybrid model. The algorithms and techniques used to generate the models are explained in detail.

3.3.1 Convolutional Neural Network (CNN)

CNN is a popular model in the domain of image classification and predictions. It is a prevalent technique for identifying patterns and objects from images. Due to this capability, various studies were held on research problems like finding patterns from various unlabeled data sources like images, log data, and sensor signals. CNN is a deep learning technology combined with the concept of Artificial Neural Networks (ANNs). A generic CNN architecture consists of an input layer, multiple hidden layers, activation functions, pooling layers, normalization layers, fully connected layers, and an output layer. In our study, the role of CNN is to predict the images and provide a multi-level classification of engagement levels as accurately as possible for the test data.

Figure 3.2 gives a standard CNN architecture presented by Murshed et al.[2] in 2019 where input, convolutional, pooling, fully connected, and output layers are demonstrated.

2. M.Murshed et al., "Engagement Detection in e-Learning Environments using Convolutional Neural Networks," *IEEE*, 2019, 80–86.

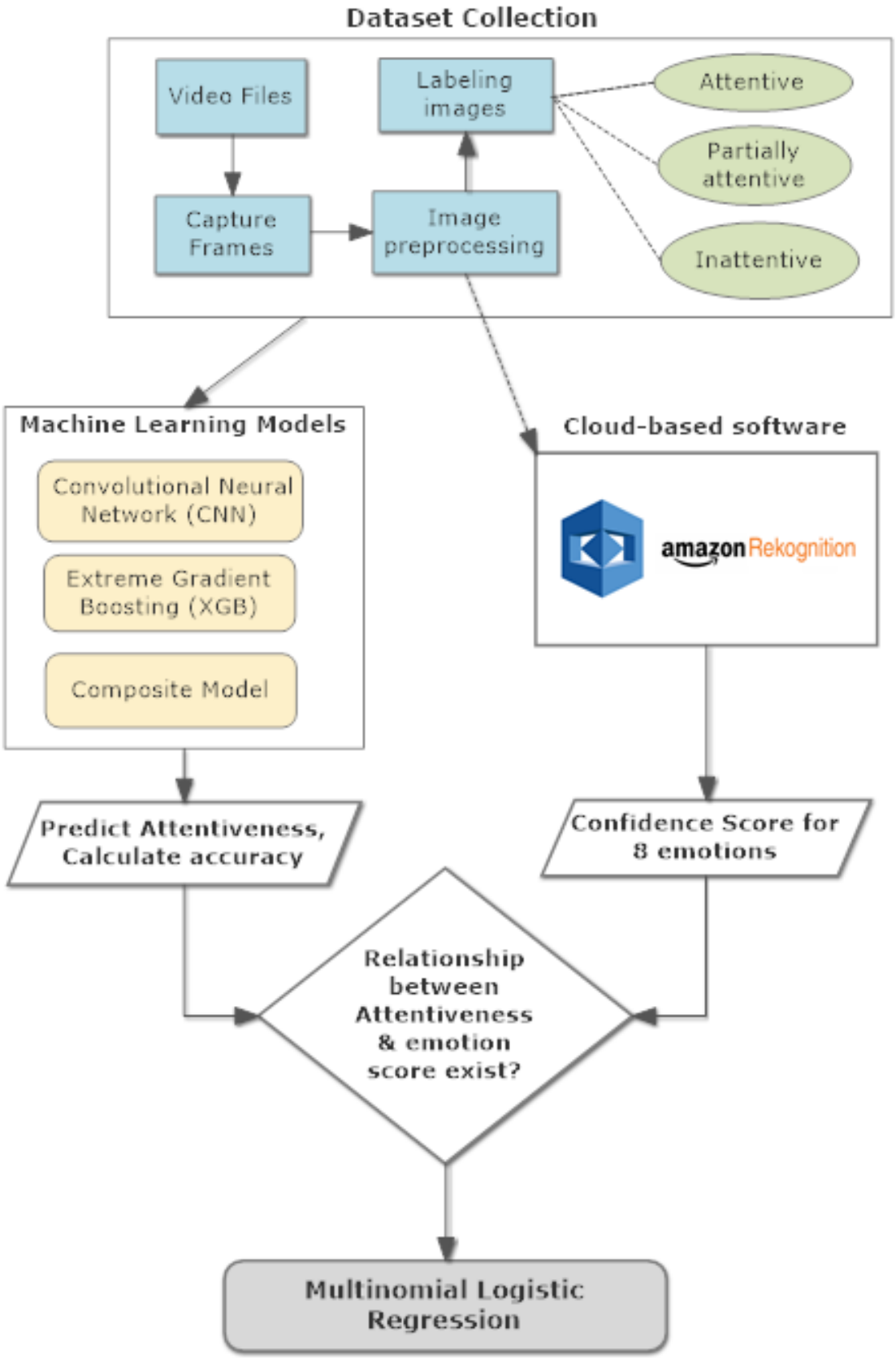

Figure 3.1: Flowchart of research methodology

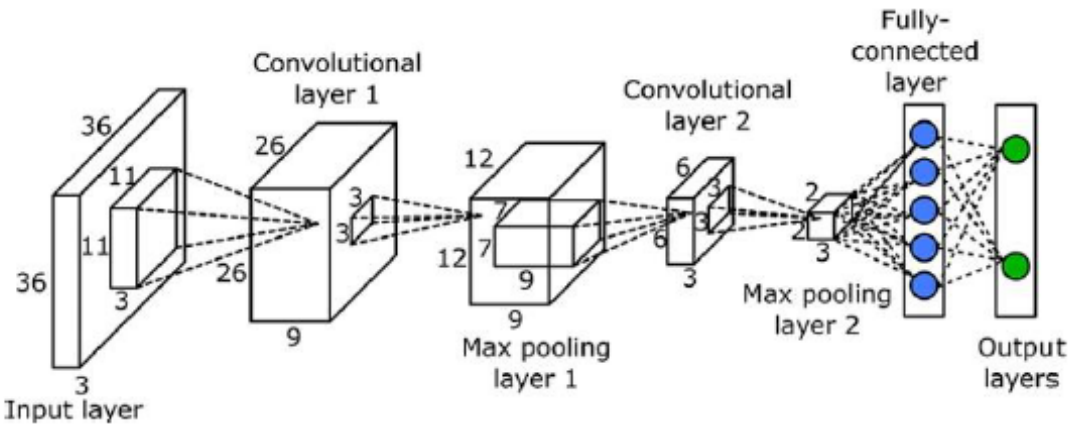

Figure 3.2: Sample architecture of CNN

In a CNN architecture, the convolutional layer convolutes an input image with a set of kernels or filters and produces feature maps. Here the kernel is a sliding window that convolutes across the input data to detect the features. Different feature maps can distinguish the presence of unique features at all possible locations. The formula for calculating the spatial size of the feature map is K*((W-F+2P)/S+1), where W – the size of the input volume, F - the receptive field size of the kernel, S – stride in use, P - the number of zero paddings used on the border, K - the depth of the convolutional layer.

Each output of the convolutional layer is then passed through an activation layer that uses an activation function to choose a neuron's final value. The activation function transforms the linear combination of features into non-linear features so that the neural network can learn faster with high accuracy. The commonly used activation functions are - sigmoid, tanh, Rectified Linear Unit (ReLU), etc. ReLU was used as the activation function in CNN except for the output layer. It can be presented as f(x) = max(0,x) and applied elementwise. Softmax activation function was used in the output layer as output classes were discrete from each other.

The pooling layer reduces the feature map's dimension while retaining important information. It partitions the images into overlapping or non-overlapping regions where each

region's spatial resolution is reduced by non-linear functions such as max pooling and average pooling. Max pooling gives maximum value for each region as output by down-sampling. In contrast, average pooling gives each region's average value as output by averaging the parameters in the pool.

The normalization layer is used to update the data values to a standard distribution scale without deforming the data range differences. The normalized data helps to connect the network efficiently while training the model. The dropout layer is often used as a regularizing function for data redistribution. It helps to reduce overfitting the neural network by assigning a probability to drop off at each unit in the layer during the training phase of the model.

The loss function was used to optimize our multi-level classification problem during the training phase. A lower loss value significantly improves the prediction results. Sparse categorical cross-entropy is used as our loss function where the truth labels are encoded in integers. The model aims at reducing the loss value at every epoch, and the "adam" optimizer was used as an optimization function.

The last convolutional layer's feature map was flattened and passed on to one or more fully connected layers that are then passed on to the output layer (softmax function) for multi-level classification. Most CNN architectures use the fully connected layer before the output layer.

The proposed CNN model from Figure 3.3 consists of seven convolutional layers with 3×3 kernel filters. In this model, we have increased the number of filters to decrease the spatial volume of the output. The modified CNN is an inspiration from the VGG16 architecture.

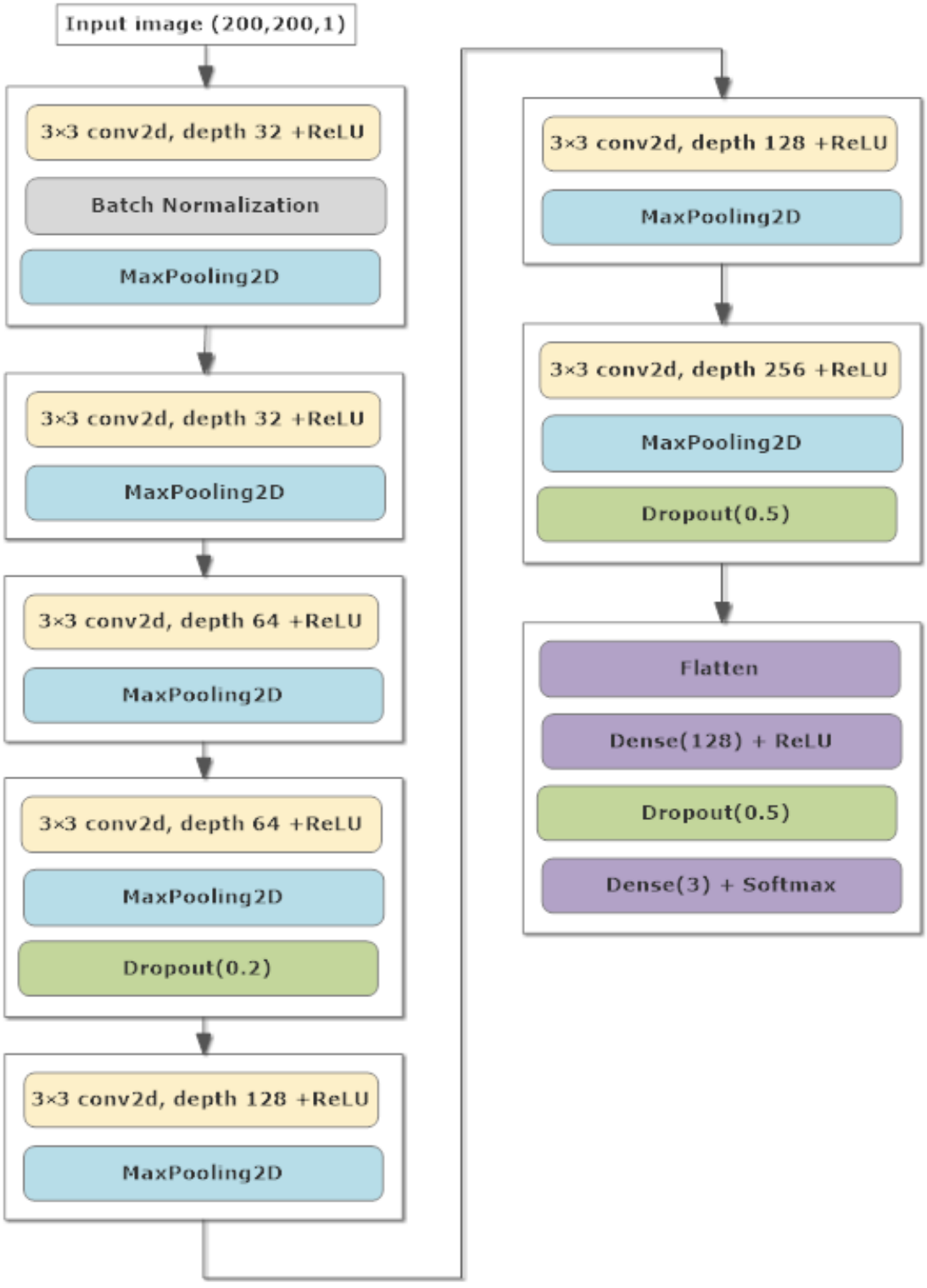

Figure 3.3: Proposed CNN architecture

3.3.2 Extreme Gradient Boosting Algorithm

Extreme Gradient Boosting or XGBoost is a decision tree-based ensemble machine learning algorithm that uses a gradient boosting framework. Ensemble-based models are considered to be more advanced methods until the recent development of neural network models. Adaboost and XGBoost are the bagging subsets of ensemble methods to reduce the models' fundamental biases.

Figure 3.4: XGBoost process flow diagram

XGBoost algorithm was chosen due to its simpler construct and the ability to handle complex datasets with significant accuracy, and the feasibility to add missing values within the dataset. It can penalize complex models using L1 or L2 regularization, which will eventually help reduce overfitting. It will improve the performance until the loss function is as small as possible. Also, XGBoost optimizes the available disk space and makes the best out of memory when dealing with a large dataset. Hence, this algorithm efficiently utilizes hardware and software resources to yield desired outcomes within a shorter time. Decision tree-based models are best in classification and prediction problems when the dataset sizes are between small to medium. Figure 3.4 depicts the process flow of the XGBoost algorithm, which uses k-fold cross validation where k is 3.

3.3.3 Composite Model

Among the image classification problems, CNN and XGBoost algorithms performed well on image processing and became the focus of research due to the diversity and complexity of image data. Sometimes, these models fail to capture all the information from the images. Hence, a hybrid approach was needed to use the CNN-trained features as input to the XGBoost classifier. As we know that feature extraction is the crucial step in automating the image classification process, the quality of extracted features can efficiently utilize the algorithm's performance, which was time-intensive in the traditional prediction algorithms. The overall idea is to obtain the features yielded from the dense layer of the trained CNN model and use those features to train the XGBoost classifier. The number of convolutional layers used for CNN is the same as in the previous model discussed in Section 3.3.1. XGBoost classifier uses softmax function for the objective parameter.

3.4 Statistical Analysis

A statistical model was used on data from the machine learning model with best predictive performance. This model provides statistical correlation between the student's attentiveness level and multiple emotional states. Furthermore, this analysis was expanded to understand the strength of predictive markers between different emotion states with respect to the attentiveness level.

Amazon Rekognition

This research uses Amazon Rekognition, a cloud-based computer vision platform to detect faces from images and videos. This web service returns an emotion confidence score using a commercially available facial emotion detection API on the images by detecting facial landmarks like eye, pupil, nose, mouth, and jawline positions. Apart from emotion

analysis, Amazon Rekognition can perform various other image operations such as label detection, celebrity recognition, text detection, PPE (Personal Protective Equipment) detection, sunglasses detection, and facial attributes based on gender. The emotions supported by Amazon Rekognition API are: *happy, sad, anger, confused, disgusted, fear, surprised, calm*.

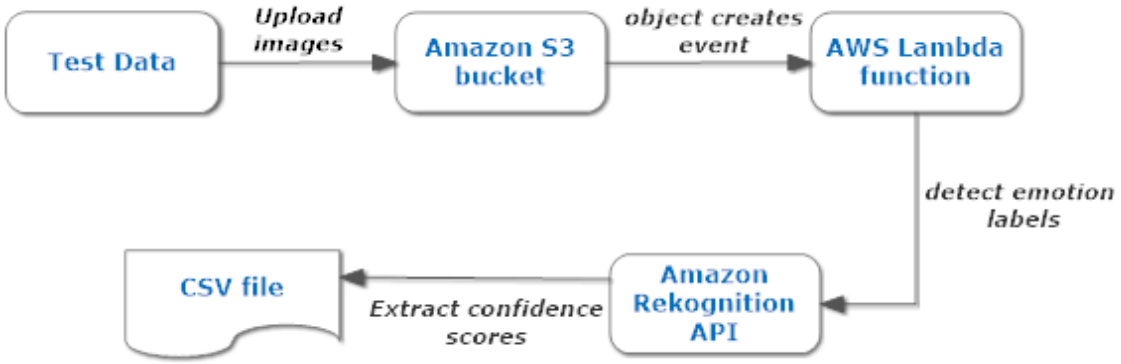

Figure 3.5: Workflow of Amazon Rekognition API

As shown in Figure 3.5, the test data images were fed into Amazon Rekognition API with the help of the AWS lambda function by uploading images in the Amazon S3 bucket. The object in the lambda function triggers an event to detect emotions in the image. Each image uploaded to the API returns a confidence score based on the emotions detected in the uploaded image. The confidence value ranges from 0 to 100 for all the emotions. The results obtained from the machine learning model and the confidence score from the Amazon Rekognition were combined to perform the statistical analysis.

Regression Analysis

Regression analysis is used to understand the relationship between two or more variables and calculate correlation coefficients between variables. The correlation coefficients are critical in evaluating the statistical validity of our research's null hypothesis.

For our thesis, we performed *"Multinomial Logistic Regression (MLR) analysis"* using Statistical Product and Service Solutions (SPSS) tool to explain the relationship between one nominal dependent variable and one or more independent variables. This regression type is chosen as it allows for more than two categories of the dependent variables to be tested. In our case, we have three categorical dependent variables – Attentive, Partially attentive, Inattentive, and the eight independent variables (the eight emotions supported by Amazon Rekognition API).

In this study, Attentive students are used as a control group to measure the statistical variation in an emotional response from students with lower levels of attentiveness. Among all the emotion states, anger is considered as a non-normal response from a student in a typical E-learning environment. One of our research goals is to statistically test this null hypothesis by explicitly looking for the response of the independent variable "anger." Two iterations of MLR analyses were performed to validate this hypothesis by including and excluding the independent variable "anger."

CHAPTER 4

RESULTS

4.1 Experimental Setup

The development environment used for the experimental setup has the following configuration:

Operating System: Windows 10 Pro

Processor: Intel(R) Core(TM) i5-8350U CPU @ 1.70GHz

System Memory: 16GB DDR4

GPU: Intel(R) UHD Graphics 620

4.2 Metrics Used

For evaluating the performance of the models, we used various metrics like classification report, confusion matrix, accuracy graph, loss graph and ROC-AUC graph. The common terminologies that were used in these metrics are: *TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.* The metrics used for model evaluation are discussed below:

- *Confusion Matrix:* This matrix is a table visualizing the model prediction performance. Each entry denotes the number of predictions made by the model either accurately or inaccurately. We used a multi-class confusion matrix for our research problem with three predicted classes and three actual classes. Figure 4.1 illustrates the generation of metrics TP, TN, FP, and FN in a confusion matrix for a positive case of "Class A."

- *Accuracy & Error rate:* Accuracy is the fraction of correct predictions in a sample to the total sample size. Being the most common metric used in classification problems,

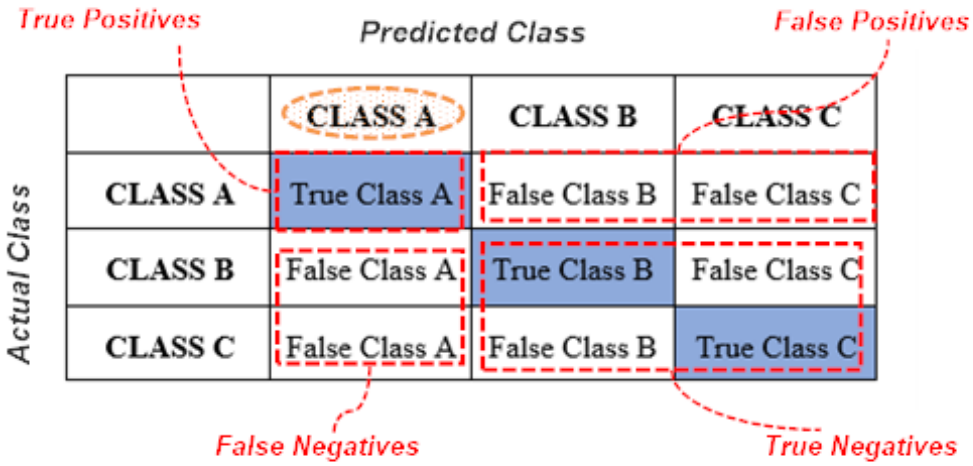

Figure 4.1: Confusion Matrix for Multi-class classification

this metric is very sensitive to imbalanced datasets. The formula to calculate accuracy is:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

Error rate is the fraction of incorrect predictions in a sample to the total sample size. The formula to calculate error rate is:

$$Error\ Rate = \frac{FP+FN}{TP+TN+FP+FN}$$

- *Precision, Recall & F1-Score:* Precision gives the proportion of positive predicted values and calculated using the following formula:

$$Precision = \frac{TP}{TP+FP}$$

Recall gives the proportion of actual positive values that are predicted positive and calculated using the following formula:

$$Recall = \frac{TN}{TN+FP}$$

F1-score is a harmonic mean between Precision and Recall. Accuracy is used when TP and TN are significant, and F1-score is used when FN and FP are identified as crucial metrics. Unlike Accuracy, F1-score provides a stable measure in models with an imbalanced dataset. It is calculated using the formula shown below:

$$F1 - score = \frac{2TP}{2TP + FP + FN}$$

- *Micro F1, Macro F1 & Weighted average F1:* Micro F1-score is the fraction of correctly predicted samples to all samples considered for prediction. The Micro F1 metric is preferred in multi-level classification problems due to the possibility of an imbalanced dataset in any model.

  The Macro F1-score will independently compute the metrics of each class and aggregate their average. It takes the unweighted mean of the measure. In contrast, the weighted average F1-score is considered as the weighted mean of measure. F1-scores are calculated for the number of actual instances for each class. There is a possibility for the F1-score to not fall between Precision and Recall metrics.

4.3 Machine Learning Module Results

In this research, the performance of all machine learning models was tested based on metrics defined in Section 4.2. To acquire a fair outcome from all the models, we used the same dataset. The model evaluation of each of the machine learning classifiers is discussed below.

4.3.1 Model evaluation outcome for CNN classifier

We used 40 epochs with the default batch size 32 by setting the learning rate at 0.001. In the early stopping callback function, the patience parameter was set to 15 to monitor any validation loss within the model. ReducedLROnPlateau was used as one of the callback

functions to lower the learning rate by a factor of 0.3 when the validation loss metric has stopped improving. It monitors the loss quantity, and when it notices zero improvements until patience is 5 units, the learning rate was reduced to 0.3. These numerical factors were identified based on the empirical trial and error method. For our model, the loss function reached saturation point on approaching 0.35 before completing the 40 epochs, and the overall accuracy reached a maximum value of 91.4%. The metrics used to assess the CNN model were confusion matrix, classification report, ROC-AUC graph, accuracy graph, and loss graph. Evaluating the model efficiency based only on accuracy and loss value obtained from the validation set could be misleading when the dataset is imbalanced between the classes.

	Precision	**Recall**	**F1-score**	**Support**
Inattentive	0.95	0.95	0.95	175
Partially attentive	0.92	0.87	0.89	181
Attentive	0.84	0.90	0.87	180
micro avg	0.91	0.91	0.91	536
macro avg	0.91	0.91	0.91	536
weighted avg	0.91	0.90	0.91	536

Table 4.1: Classification report for CNN model

A classification report is a crucial metric that measures the quality of predictions for a classification problem. It gives information about Precision, Recall, F1-scores, and Support value for each class in the model. From Table 4.1, Precision, Recall, and F1-score for inattentive students were observed to be marginally higher than other classes.

A confusion matrix was generated to summarize the classifications. It provides a clear visualization of correct and incorrect predictions for each class. Figure 4.2 shows the

confusion matrix for the CNN model, providing insight into the model's performance with the percentage of positive and negative predictions made by the model. By observing the TP measure of all classes, we can confirm that approximately 91% of the validation images were classified correctly. In contrast, the rest of all incorrect predictions stays lower than 20 students.

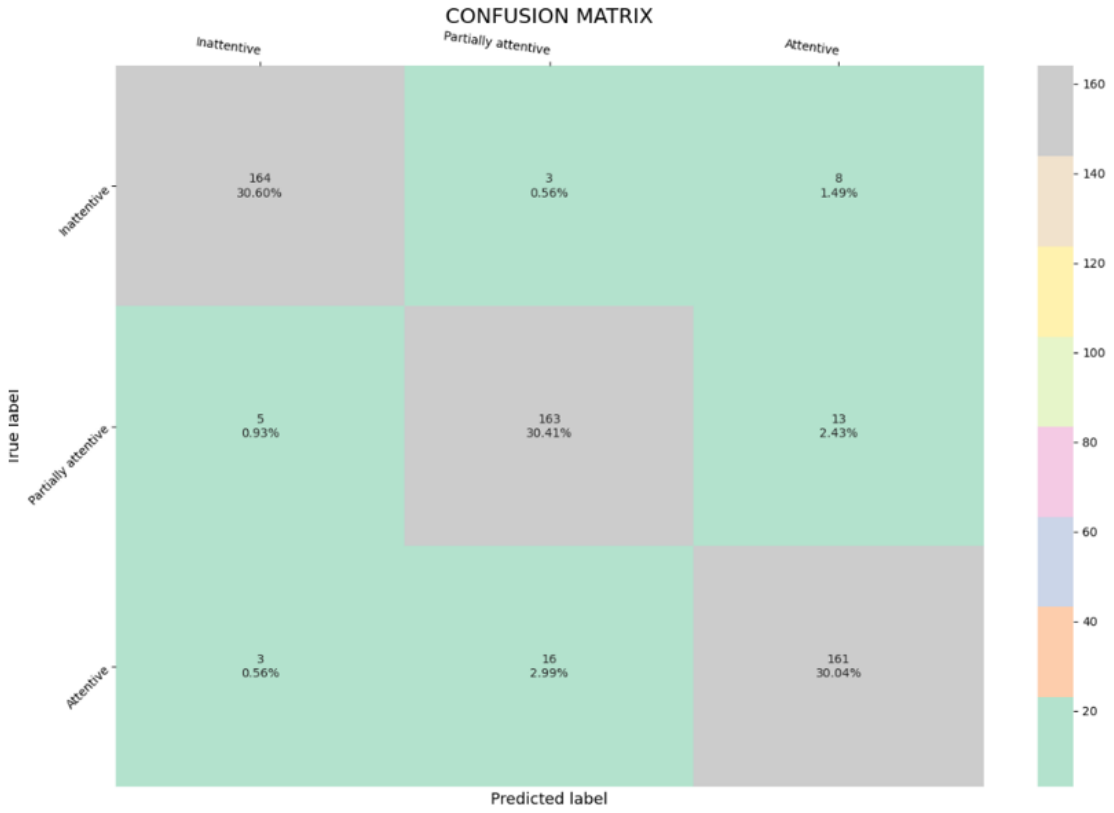

Figure 4.2: Confusion matrix for CNN model

The relationship between the accuracy of the training set and the validation set for each epoch is shown in Figure 4.3. The graph demonstrates that the accuracy is increased positively with each epoch for both training and validation sets. It is not always required to consider the validation learning curve's last data point with the model's highest accuracy. In our research at epoch 25, the highest accuracy of the model was reached.

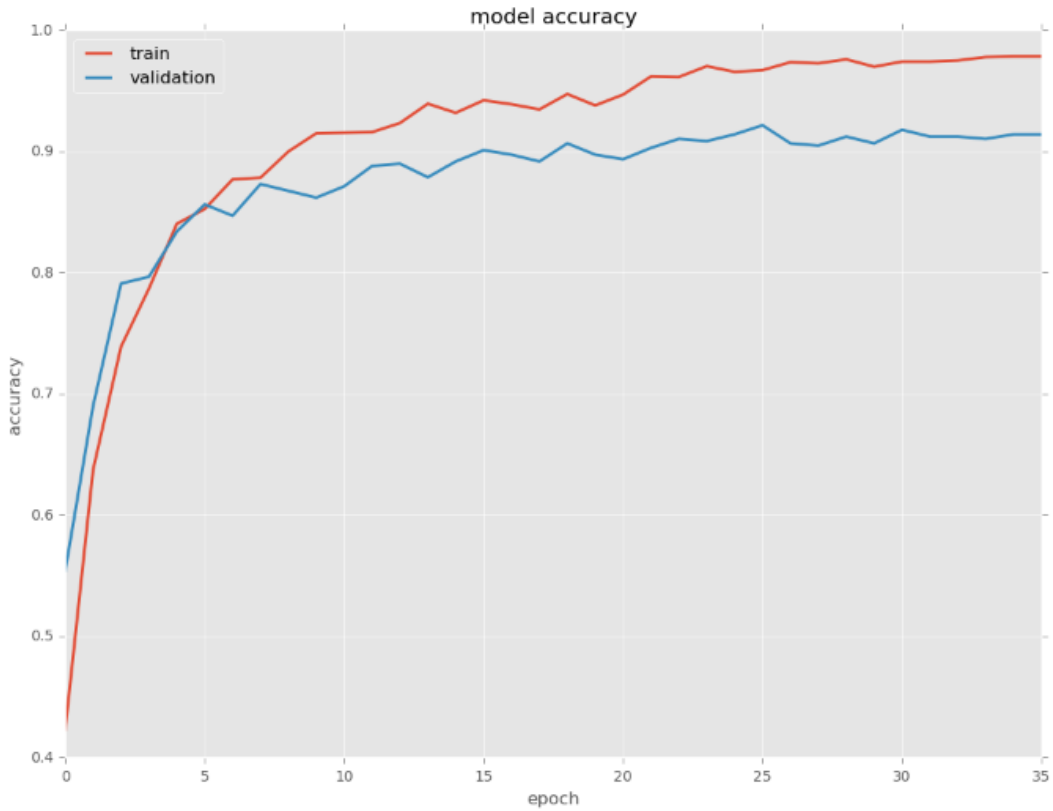

Figure 4.3: Accuracy graph: Training vs. Validation

The loss function relationship between training and validation sets concerning each epoch is shown in Figure 4.4. Per the callbacks tuned in the CNN model, the graph stops at epoch 35 with the patience parameter at 15 as the validation loss function observed no improvement.

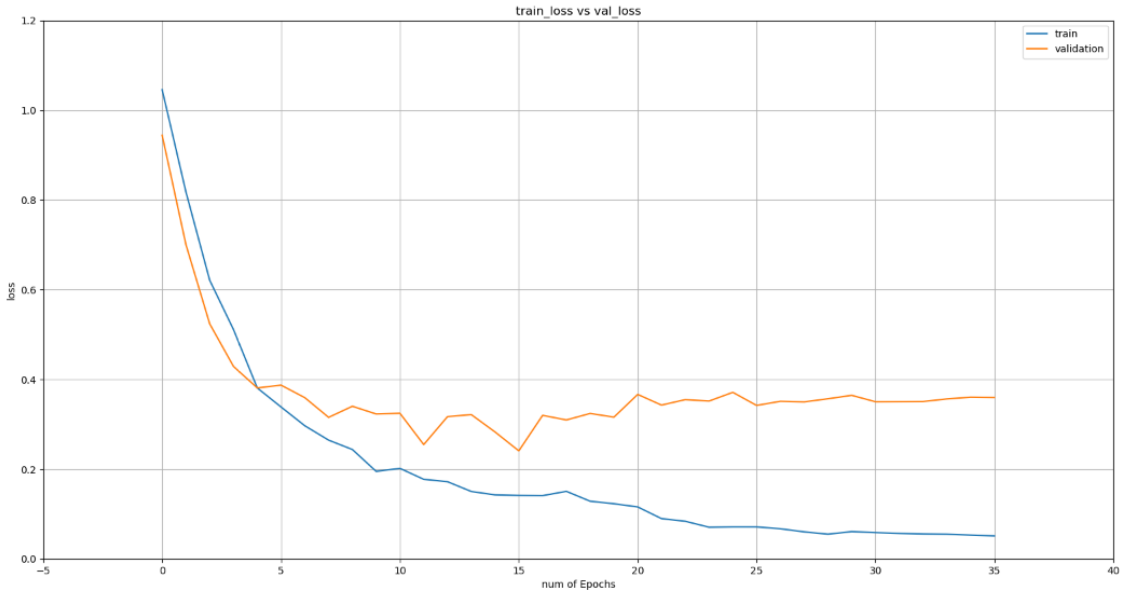

Figure 4.4: Loss graph: Training vs. Validation

ROC-AUC curve in Figure 4.5 demonstrates the probability score for all three classes, micro average and macro average. When the curves are closer to the top-left corner, it indicates better performance, and the AUC scores give a general measure of predictive accuracy. A closer look to the top-left corner of the ROC-AUC graph is shown in Figure 4.6.

4.3.2 Model Evaluation Outcomes for XGBoost Classifier

A three-fold cross-validation approach was used to evaluate the performance of the XGBoost classifier. This method helps to estimate the performance of any ML classifier with less variance. The result from this approach is a reliable estimate of the model's

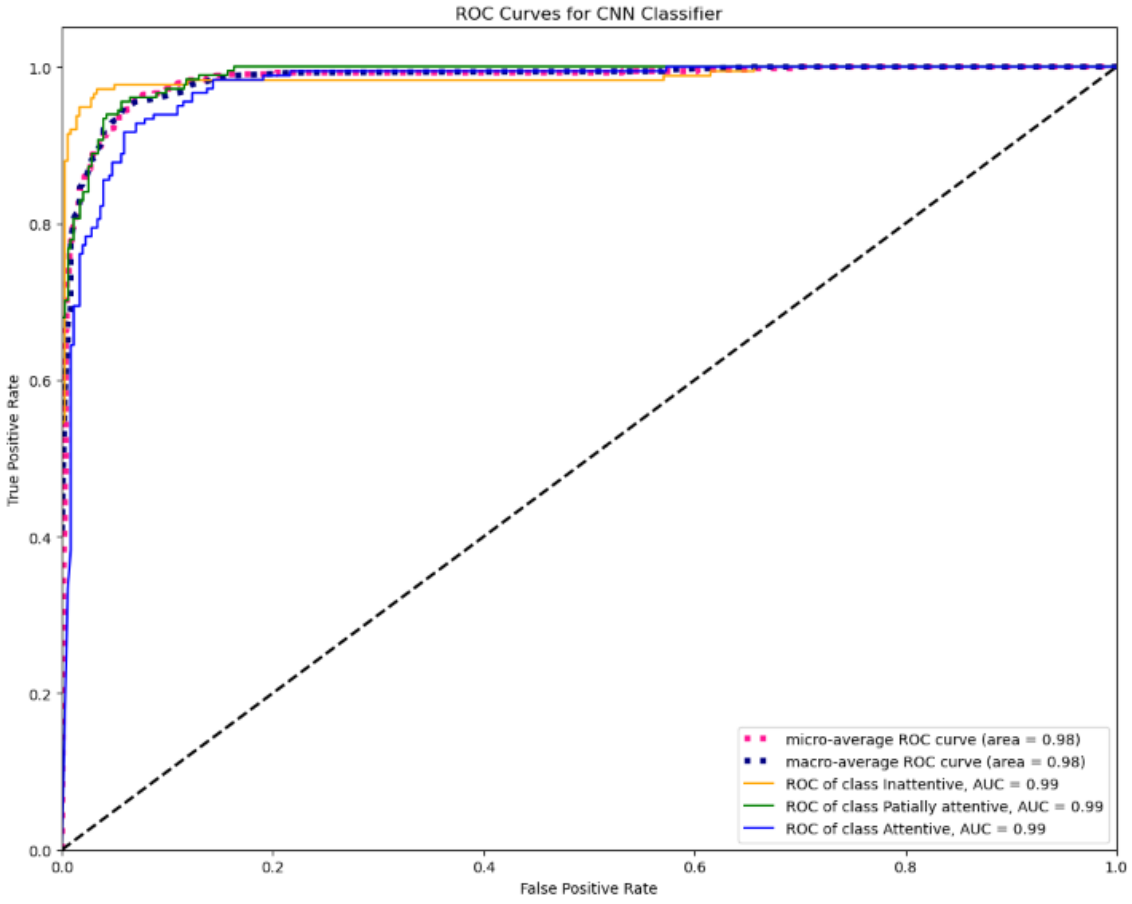

Figure 4.5: ROC-AUC curve for CNN model

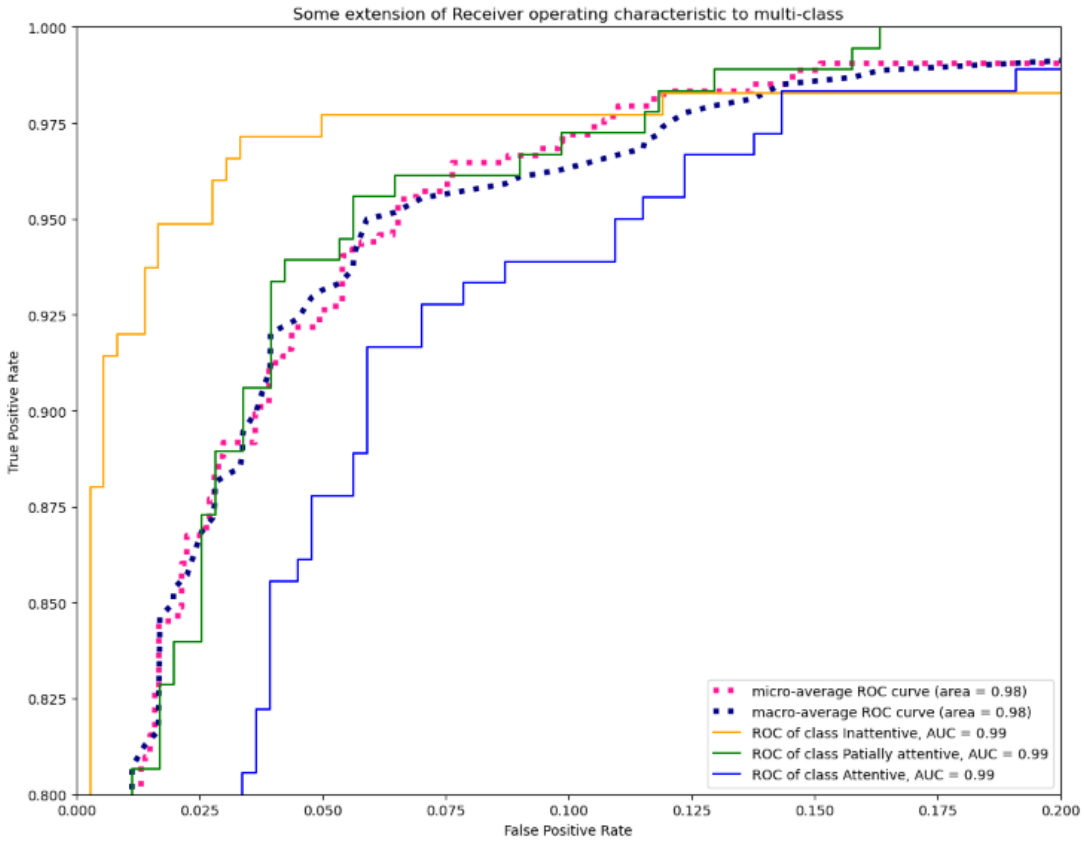

Figure 4.6: Extended ROC-AUC curve for CNN model

performance as the algorithm is trained and evaluated multiple times on a different data split. The classification accuracy of the XGBoost classifier was observed at 88.06%.

Table 4.2 shows the classification report for the XGBoost classifier. This table observed that Precision, Recall, and F1-scores for inattentive students were marginally higher than other classes. Here the micro average, macro average, and weight average remained the same.

ROC-AUC graph for XGBoost classifier is shown in Figure 4.7. It is observed that the ROC curve for all the classes was located in the top-left corner, which proves that the XGBoost algorithm performs well on classification.

	Precision	**Recall**	**F1-score**	**Support**
Inattentive	0.94	0.94	0.94	175
Partially attentive	0.88	0.84	0.86	181
Attentive	0.83	0.87	0.85	180
micro avg	0.88	0.88	0.88	536
macro avg	0.88	0.88	0.88	536
weighted avg	0.88	0.88	0.88	536

Table 4.2: Classification report for XGBoost classifier

The confusion matrix for the visualization summary of classification is shown in Figure 4.8. It's observed that the TP measure for the model is approximately 88%, and all incorrect predictions were found to be less than 24 students.

4.3.3 Composite Model And its Model Evaluation Outcomes

The composite model performed relatively better than the XGBoost model but less than the CNN model. In the hybrid model, the trained CNN features were used for building

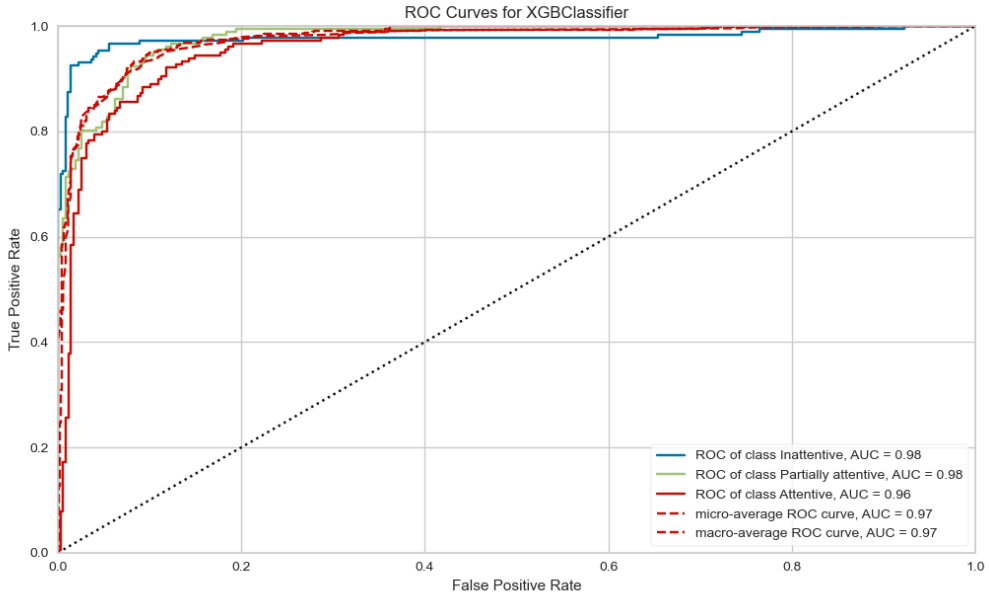

Figure 4.7: ROC-AUC curve for XGBoost classifier

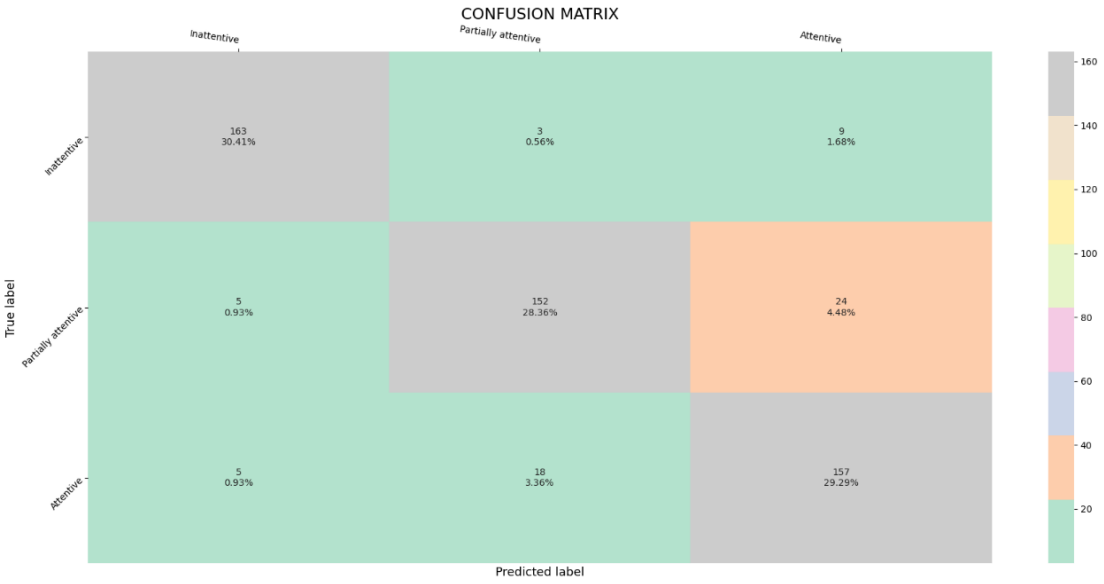

Figure 4.8: Confusion matrix for XGBoost model

the XGBoost classifier. The composite model's accuracy was observed at 90%, but the model's training time does not improve and is similar to the time taken to train both CNN and XGB models. Hence, it can be concluded that no improvement was observed in the training time of the hybrid model, but it gives better classification performance than a regular XGBoost classifier. There is an opportunity to improve the performance of the composite model by adding additional layers to the CNN segment of the composite model.

The classification report summarizes Precision, Recall, and F1-scores for all the classes in Table 4.3. The confusion matrix visually represents the correct and incorrect predictions with proportion for all the classes given in Figure 4.9. We can observe that the inaccurate predictions of all categories are less than 20 students indicating better performance than the XGBoost classifier.

	Precision	**Recall**	**F1-score**	**Support**
Inattentive	0.94	0.92	0.93	175
Partially attentive	0.92	0.88	0.90	181
Attentive	0.84	0.90	0.87	180
micro avg	0.90	0.90	0.90	536
macro avg	0.90	0.90	0.90	536
weighted avg	0.90	0.90	0.90	536

Table 4.3: Classification report for hybrid model

4.4 Multinomial Logistic Regression Analysis Results

The results obtained from the CNN model were compared with the confidence score obtained from the Amazon Rekognition model. This statistical comparison was made using MLR analysis. As it is uncommon for a student to exhibit the emotion state anger within

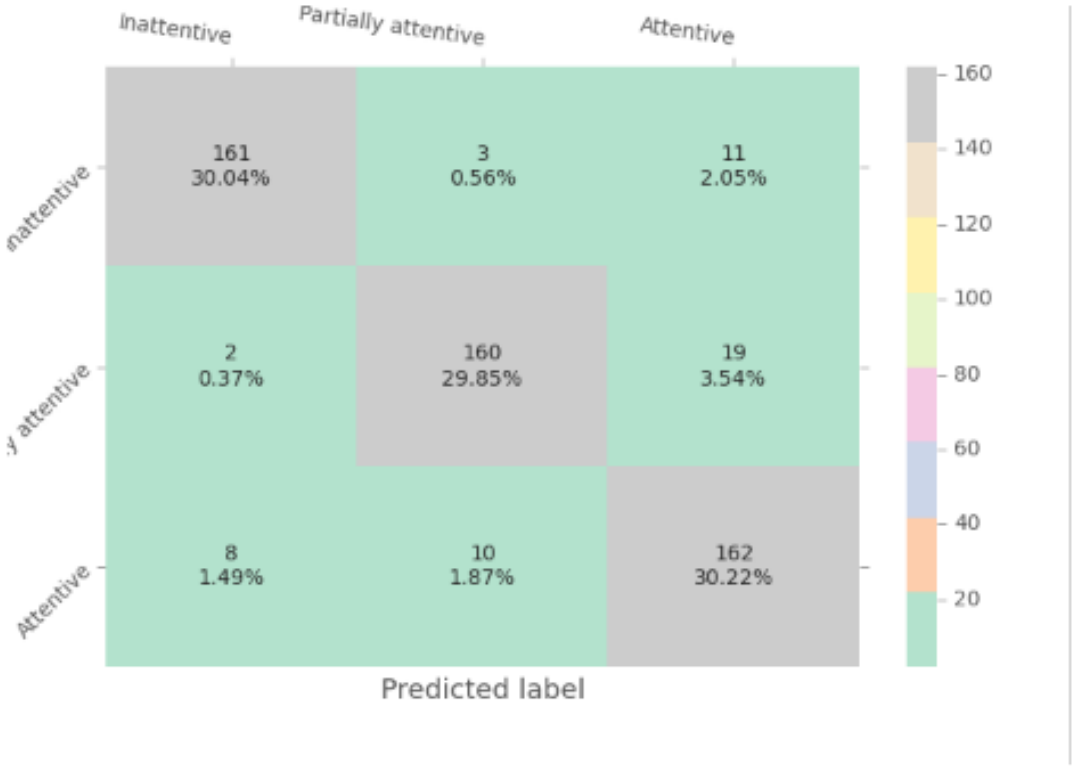

Figure 4.9: Confusion matrix for hybrid model

a classroom environment, this research tries to understand the impact of anger over other emotional states while establishing the correlation model between emotion state and attentiveness level. Therefore, two MLR analyses were performed - One analysis included all emotion states, and the other omitted the emotion state anger.

In MLR analysis, the Likelihood ratio test examines each emotion's overall contribution (except anger) to the correlation model. When the conventional α is at 0.05 threshold, it is observed that emotions like calm, happy, surprised, and fear exhibit a statistically significant response, as shown in Table 4.4.

Effect	**Significance**
Intercept	0.051
Calm	0.019*
Confused	0.727
Disgust	0.381
Fear	0.050*
Happy	0.014*
Sad	0.735
Surprised	0.045*

Table 4.4: Likelihood ratio tests

Note: * indicates significance <= 0.05

The Parameter estimation technique provides a deeper understanding of attentiveness levels by indexing the base reference category as a highly attentive group and comparing the response against inattentive and partially attentive groups. Parameter estimation including the emotion state anger is shown in Table 4.5. The set of emotions for inattentive students were compared against the parameters used for predicting the highly attentive students.

As illustrated in Table 4.5, none of the emotion states exhibited a statistically significant response in predicting inattentive students.

On comparing partially attentive students with highly attentive students using Parameter estimates, it was observed that the emotion states anger and sad showed a statistically significant response. Their respective correlation coefficients were negative, indicating an indirect correlation between partially attentive students and emotion states anger and sad. The emotion state confused showed a near-significant statistical response with a p-value of 0.051 closer to the statistical acceptance criteria.

Precision		**Intercept**	**Anger**	**Calm**	**Confused**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprised**
Inattentive	Coefficient	.456	-.012	-.006	-.011	-.012	.022	-.051	-.009	0
	significance	.862	.664	.814	.709	.829	.496	.197	.722	-
Partially attentive	Coefficient	4.897	-.092	-.043	-.072	-.029	-.161	-.050	-.073	0
	significance	.160	.033*	.217	.051*	.586	.129	.170	.039*	-

Table 4.5: Parameter estimation with emotion state anger

Note: The reference category is *'Attentive'* and * indicates significance <= 0.05

Table 4.6 shows the parameter estimation results excluding the dataset with emotion state anger. Based on the results from Table 4.6, it can be concluded that inattentive students showed a statistically significant response with direct correlation to the emotion states of fear and sad.

Also, partially attentive students showed a statistically significant response with a direct correlation to the emotion states of calm, confused, and surprised.

From Table 4.6, it is evident that a statistically significant response can be achieved with different emotion states for inattentive and partially attentive students by controlling the emotion variable at anger. Also, the emotion state surprised showed a statistically significant response in the second MLR analysis which was confounded in the first MLR analysis by the emotion state anger.

Precision		Intercept	Calm	Confused	Disgust	Fear	Happy	Sad	Surprised
Inattentive	Coefficient	-.786	.006	.002	.000	.035	-.039	.003	.012
	significance	.456	.575	.893	1.000	.051*	.201	.048*	.664
Partially attentive	Coefficient	-4.316	.049	.020	.063	-.069	.042	.019	.092
	significance	.061	.037*	.045*	.204	.411	.0.88	.461	.033*

Table 4.6: Parameter estimation without emotion state anger

Note: The reference category is *'Attentive'* and * indicates significance <= 0.05

4.5 Result Discussion

In this research, different machine learning models were tested for efficiency and accuracy in predicting student's engagement levels. A balanced dataset was used in evaluating all the machine learning models. Based on the performance metrics, it can be concluded that the deep learning CNN model outperforms both the XGB model and the hybrid model. Comparing the latter two models, the hybrid model edges over the traditional XGB model as it utilizes the input features from the trained CNN model. Despite exhibiting better accuracy, the hybrid model experiences a higher training time. The XGB model generated the highest proportion of incorrect classification compared to the other models. It can be further summarized that the CNN model yielded the best performance in all metrics with the highest accuracy of 91.4%.

The results from the CNN model were then combined with the emotion confidence parameters generated from the Amazon Rekognition tool to develop the statistical correlation model between engagement levels and emotional states. Results from the statistical analysis showed the confounding effect of emotion state anger onto emotion state surprised. This confounding effect could be due to the shared facial feature responses like raised eyebrows, mouth lines, and wrinkled nose exhibited during the emotion states anger, surprised and neutral face as substantiated by Artemisa et al. (2020). Including the emotion state

anger in the MLR analysis yields a predominantly negative correlation between different emotional states and student engagement levels. In contrast, the MLR analysis without emotion state anger exhibits a combination of both positive and negative correlation. This statistically confirms the negative confounding effect of the emotion state anger on other emotion states while predicting the student's engagement levels. It is concluded from the MLR analysis that the emotion states calm, happy, surprised, and fear were significant prediction markers to identify any category of student engagement levels. The emotion states fear and sad were directly correlated to predict inattentive students, whereas emotion states calm, confused, and surprised are directly correlated to predict partially attentive students.

CHAPTER 5

CONCLUSION

The findings of this research contribute to a better online learning environment by helping instructors accurately identify inattentive and partially attentive students. It enables the instructors to guide the students in need, consequently leading to a better learning experience. Our work evaluated three machine learning models to measure student's engagement levels based on their emotions. The research framework used in this study identified the CNN model as the suitable machine learning model to gauge a student's attentiveness based on their emotion state with the prediction accuracy of 91.4%. This research also tested the impact of emotion state anger on the relationship between emotion states and student's engagement levels. Understanding the confounding effect of anger on other emotion states helped us statistically identify critical emotions exhibited by inattentive and partially attentive students. With the results from this study, we can conclude that the deep learning CNN model provides a reliable and accurate platform to gauge multiple gradients of student engagement based on their facial emotions.

CHAPTER 6

FUTURE WORK

This research work can further be expanded in many possible avenues. For future research, the CNN model can be updated to utilize computing resource-intensive architectures like VGG16, VGG19, and ResNet, which would further improve the prediction accuracy of the machine learning model.

This research can be developed by adopting a broader spectrum of engagement levels to gain a granular understanding of student's attention levels with facial emotions. Additionally, the research framework can be improved by including a web-based application to convert livestream video files into images, which would provide a real time data-feed into the prediction model. A student survey can be added to the end of every online session to generate a user-driven feedback data point to improve and validate the prediction metrics of the machine learning models.

Another direction of the project is to perform auto-labeling of the images instead of manual labeling. Once the correlation and the significance between the emotions and the engagement levels is meticulously identified, the cloud-based software can act as an AI expert in the labeling process. This method is helpful in handling large-scale datasets.

REFERENCES

Ayvaz, Uğur, Huseyin Guruler, and Mehmet Devrim. "USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS." *Information Technologies and Learning Tools* 60 (September 2017): 2076–8184. https://doi.org/10.33407/itlt.v60i4.1743.

Frank, Maria, Ghassem Tofighi, Haisong Gu, and Renate Fruchter. "Engagement Detection in Meetings." July 2016.

Grafsgaard, J. F., J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester. *Automatically Recognizing Facial Expression: Predicting Engagement and Frustration (International Conference on Educational Data Mining*. Memphis, 2013.

Grafsgaard, Joseph, Joseph Wiggins, Kristy Boyer, Eric Wiebe, and James Lester. "Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis," 159–165. September 2013. https://doi.org/10.1109/ACII.2013.33.

Gupta, A., Richik Jaiswal, Sagar Adhikari, and Vineeth Balasubramanian. "DAISEE: Dataset for Affective States in E-Learning Environments." *ArXiv* abs/1609.01885 (2016).

Gupta, A. D'Cunha, K. Awasthi, V. Balasubramanian, and DAiSEE: Towards. *user engagement recognition in the wild. Computer Vision and Image Processing. arXiv*. Preprint. arXiv 1609, 01885, 2018.

Harley, Jason, François Bouchet, Sazzad Hussain, Roger Azevedo, and Rafael Calvo. "A Multi-Componential Analysis of Emotions during Complex Learning with an Intelligent Multi-Agent System." April 2014.

Johnson, Andrew P. "It's Time for Madeline Hunter to Go: A New Look at Lesson Plan Design." *Action in Teacher Education* 22, no. 1 (2000): 72–78. https://doi.org/10.1080/01626620.2000.10462994.

Klein, Richard, and Turgay Celik. "The Wits Intelligent Teaching System: Detecting student engagement during lectures using convolutional neural networks," 2856–2860. September 2017. https://doi.org/10.1109/ICIP.2017.8296804.

Llanda, Christopher John R. "Video Tutoring System with Automatic Facial Expression Recognition: An Enhancing Approach to E-Learning Environment." In *Proceedings of the 2019 4th International Conference on Intelligent Information Technology*, 5–9. 2019.

M.Murshed, M.A.A.Dewan, F.Lin, and D.Wen. "Engagement Detection in e-Learning Environments using Convolutional Neural Networks." *IEEE*, 2019, 80–86.

Megahed, M., and A. Mohammed. "Modeling adaptive E-Learning environment using facial expressions and fuzzy logic." *Expert Syst. Appl.* 157 (2020): 113460.

"Model specification in regression analysis." In *Understanding Regression Analysis*, 166–170. Boston, MA: Springer US, 1997. https://doi.org/10.1007/978-0-585-25657-3_35. https://doi.org/10.1007/978-0-585-25657-3_35.

Murshed, M., M. A. A. Dewan, F. Lin, and D. Wen. "Engagement Detection in e-Learning Environments using Convolutional Neural Networks." In *2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 80–86. 2019. https://doi.org/10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00028.

Murshed, Mahbub, M. Dewan, Fuhua Lin, and Dunwei Wen. "Engagement Detection in e-Learning Environments using Convolutional Neural Networks," 80–86. August 2019. https://doi.org/10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00028.

Paquette, Luc, Ryan S. J. D. Baker, Michael A. Sao Pedro, Janice D. Gobert, Lisa Rossi, Adam Nakama, and Zakkai Kauffman-Rogoff. "Sensor-Free Affect Detection for a Simulation-Based Science Inquiry Learning Environment." In *Intelligent Tutoring Systems*, edited by Stefan Trausan-Matu, Kristy Elizabeth Boyer, Martha Crosby, and Kitty Panourgia, 1–10. Cham: Springer International Publishing, 2014.

"Facial Expression Recognition for Intelligent Tutoring Systems in Remote Laboratories Platform." International Conference on Advanced Wireless Information and Communication Technologies (AWICT 2015), *Procedia Computer Science* 73 (2015): 274–281. https://doi.org/https://doi.org/10.1016/j.procs.2015.12.030.

Ren, X., H. Guo, S. Li, S. Wang, and J. Li. "A Novel Image Classification Method with CNN-XGBoost Model." In *Kraetzer C*, edited by Shi YQ., J. Dittmann, and H. Kim. vol 10431. Springer, Cham: Digital Forensics / Watermarking. IWDW 2017. Lecture Notes in Computer Science, 2017.

Spezialetti, Matteo, Giuseppe Placidi, and Silvia Rossi. "Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives." *Frontiers in Robotics and AI* 7 (2020): 145. https://doi.org/10.3389/frobt.2020.532279. https://www.frontiersin.org/article/10.3389/frobt.2020.532279.

Sun, Ning, Qi Li, Ruizhi Huan, J. Liu, and G. Han. "Deep spatial-temporal feature fusion for facial expression recognition in static images." *Pattern Recognit. Lett.* 119 (2019): 49–61.

T S, Ashwin, and Rammohana Reddy Guddeti. "Unobtrusive Behavioral Analysis of Students in Classroom Environment Using Non-Verbal Cues." *IEEE Access* 7 (October 2019): 1–1. https://doi.org/10.1109/ACCESS.2019.2947519.

Tabassum, Tasnia, Andrew A. Allen, and Pradipta De. "Non-Intrusive Identification of Student Attentiveness and Finding Their Correlation with Detectable Facial Emotions," 127–134. New York, NY, USA: Association for Computing Machinery, 2020. https://doi.org/10.1145/3374135.3385263. https://doi.org/10.1145/3374135.3385263.

Tekle, Halefom, Han Liu, Anwar Ul Haq, Emmanuel Bugingo, and Defu Zhang. "A New Hybrid Convolutional Neural Network and eXtreme Gradient Boosting Classifier for Recognizing Handwritten Ethiopian Characters." *IEEE Access* 8 (December 2019): 17804–17818. https://doi.org/10.1109/ACCESS.2019.2960161.

Veliyath, Narayanan. "iFocus: A Framework for Non-intrusive Assessment of Student Attention Level in Classrooms," 2019. https://digitalcommons.georgiasouthern.edu/etd/1939.

Whitehill, J., Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan. "The faces of engagement: Automatic recognition of student engagement from facial expressions." *Affective Computing, IEEE Transactions on* 5, no. 1 (2014): 86–98.

Zaletelj, Janez, and Andrej Košir. "Predicting students' attention in the classroom from Kinect facial and body features." *EURASIP Journal on Image and Video Processing*, 2017, 80.

Zhang, K., Y. Huang, Y. Du, and L. Wang. "Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks. IEEE Trans Image Process. 2017 Sep;26(9):4193-4203." *EpubMar* 30 (2017). https://doi.org/10.1109/TIP.2017.2689999.