Thesis for Bachelor of Science

Student Engagement from Facial Emotions using Deep Learning

Supervised By

Ms. Nusrat Jahan

Assistant Professor

Department of Software Engineering

Daffodil International University

Submitted By

Mariam Lima

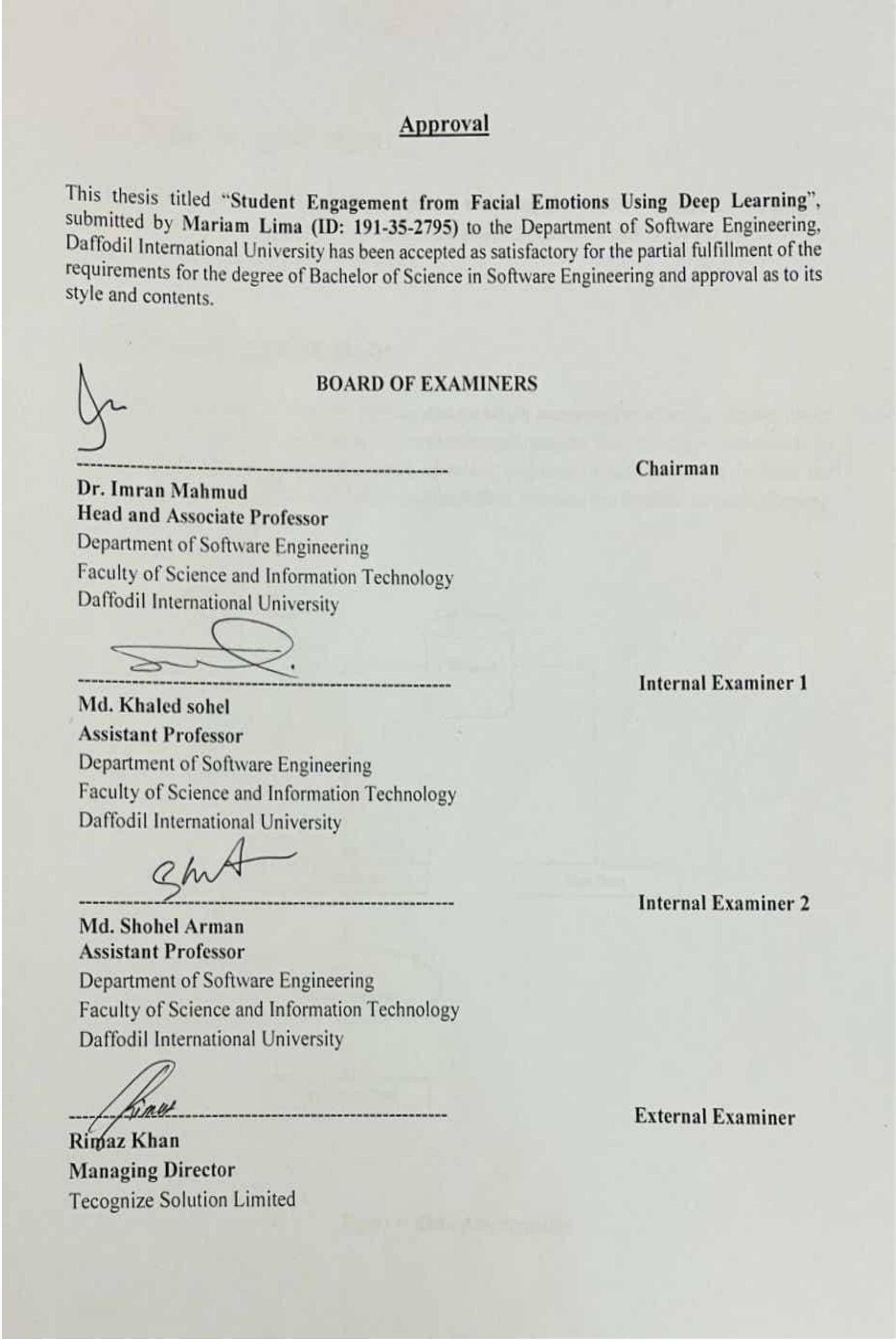
ID: 191-35-2795

Batch: 28th

Department of Software Engineering

This Project report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering.

Fall 2022



Approval

This thesis titled "**Student Engagement from Facial Emotions Using Deep Learning**", submitted by **Mariam Lima (ID: 191-35-2795)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- **Chairman**

Dr. Imran Mahmud
Head and Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 1**

Md. Khaled sohel
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Rimaz Khan
Managing Director
Tecognize Solution Limited

DECLARATION

I announce hereby that I am rendering this study document under Ms. Nusrat Jahan, Department of Software Engineering, Daffodil International University. I, therefore, state that this work or any portion of it was not proposed here therefore for Bachelor’s degree or any graduation.

Submitted by

mariam

……………….

Mariam Lima

191-35-2795

Department of Software Engineering

Daffodil International University

Certified by

…………………………………………

Ms. Nusrat Jahan

Assistant Professor

Department of Software Engineering

Daffodil International University

ACKNOWLEDGMENT

At this point, I want to express my gratitude to Allah, the Almighty, for enabling me to finish the final thesis. I also want to thank my family, who have always been encouraging and have always believed in me.

Last but not least, I would like to express my gratitude to my supervisor, Ms. Nusrat Jahan, Assistant Professor, Faculty of Science and Information Technology, Department of Software Engineering, Daffodil International University, for allowing me to work on this project and for his assistance with his direction and helpful advice on challenges I encountered while implementing this thesis.

Mariam Lima

TABLE OF CONTENT

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES FIGURES

LIST OF FIGURES

ABSTRACT

A fundamental idea in modern education, where it is considered as a goal in itself, is student engagement. In this study, we investigate methods for automatically identifying student engagement from their facial expressions. According to our opinion, the next generation of online learning environments ought to be able to monitor learners' involvement and offer tailored intervention. In this research paper, I used an existing model that is VGG16 to detecting online learner’s engagement through their facial expressions. Students facial expression will be recognized to categorize them into three-level engagement such as engaged, not engaged and neutral. The VGG16 model shows better accuracy that is 97.14%. This experiment conducted of FER2013 Dataset.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

A significant enhancement has happened in digital education or e-learning in the last few years, especially during the COVID-19 pandemic. And this e-learning is a growing industry that has grown by 900% and is expected to increase profit than before (Booth et al, 2017). The term e-learning was mentioned in 1999 during a seminar to evoke the use of computers to enroll in online degrees, to learn, and to advance their education. These web-based platforms provide facilitation to the learners with the help of the internet and web-based analytical tools (Whitehill et al, 2014). This type of education benefits institutions by allowing them to serve more students while spending less money overall, in addition to supporting individual learners in obtaining knowledge in a flexible manner (Cohen and Nachmias, 2006). The nature of e-learning courses has allured a good number of grown-up learners who were willing to get higher studies and work experience with their favorable course or work (Holder, 2007). This industry is so popular nowadays because of the flexibility to the learners as they can access any of the material like resources, videos, audio, etc. at anytime from anywhere. The most advantageous thing about e-learning is learners can interact with their educators through private chats or discussions from all over the world. It stimulates students' motivation. Some of the benefits of e-learning are given below:

- E-learning accommodates everyone's needs: E-learning courses can take any number of ages people at any time they are suitable to learn. It depends on their availability and flexibility.

- Accessible at any time: Learners can access the lecture or materials at any time and more than once as it is not like traditional classroom teaching.

- Scalability: E-learning helps in communicating and creating new concepts, ideas, and new training.

- Consistency: E-learning ensures that all the learners can receive the same learning mode with the same type of training.

- Reduce Cost: As learning through online mode is quicker and easier and the training time is also reduced that is the reason it became cost-effective compared to the traditional learning form.

1.2 BACKGROUND

In the 21st Century, students are moving to Digital education, the concentration is on teacher and student relations to acquire the target high-quality, dynamic, and meaningful education. In this era, one of the most significant problems in digital learning or in e-learning is learners' engagement in their educational activities. And this student engagement or disengagement could be detected through facial emotion using a real-time system. It is the best way for detecting student engagement in an e-learning environment through facial emotions like facial expressions, eye gaze, and head movement.

This research paper implemented an existing model with better accuracy which will work in a real-time video-based system for detecting student engagement circumstances. Facial engagement recognition is preferable because it does not require pricey technology or specialized knowledge to operate. The built-in camera or web camera of the laptop or PC acts well to capture the video of students' faces. The videos that capture using the built-in camera can be transferred into the input to the CNN model to decide whether the student is engaged or disengaged. When the students' facial expression is happy, surprised, neutral, or fearful then the student is engaged. If the students' facial expression is sad, angry, or disgusted then we'll consider them disengaged. This system will improve the quality of the model by detecting students' engagement through facial emotions in an e-learning environment.

This paper started by reviewing the area of the most recent work of different models and those models' accuracy in detecting student engagement through facial expressions using deep learning. Then I have described the methodology of this research. After that, it presents the result and discussions, and conclusion.

1.3 MOTIVATION

Statistics show that organizations that have extensive training programs have per employee 218% higher revenue and 24% higher profit margins. Many organizations offer e-learning services and many more organizations towards to offer new e-learning platforms as it is profitable. We know that the e-learning industry growing so fast day by day and became the most profitable business. So, it is important for every organization to know how to make students more engaged in their systems to be more profitable. Students always focus on quality over quantity so the organizer who was offering e-learning services should know on which content they need to focus or on which material students are more engaged or on which topic they are paying more attention. So, the e-learning service providers should the effectiveness of their courses by using real-time engagement data through students' facial expressions to make their organization more profitable by grabbing students' attention. Also, when educators teach any topic to the learners, it is a problem for them to understand whether students are understanding the topic, whether they are engaged with this course or not, or whether the student needs extra care from the educators. As the concentration is on teacher and student relations to acquire the target high-quality, dynamic, and meaningful education.

So, the motivation of this research paper is to help e-learning organizations and educators to know about their learners' engagement through their emotions and mostly to implement a better model than previous models' which will detect more accurately and give better accuracy to recognize engagement in real-time.

1.4 PROBLEM STATEMENT

The E-learning industry growing so fast with time but it is a problem to catch up on student concentration towards learning. It is not like the traditional teaching system in a classroom where a teacher can monitor their activities on live by taking some extracurricular activities or by taking attendance, quizzes, etc. In the traditional teaching system, educators can solve their student's problems, or can motivate them, or can talk with them personally but it is quite tough to do on the e-learning platform. In the e-learning system it is very much tough for an educator to look after each and every student personally whether they have done their work or not, are their materials are interesting enough to grab students' concentration. Also, students often register for a course but do not continue the course, or most of the time their focus on the course looks so poor that it is tough for the educator to ask or take care of each student that's why after starting a course the dropout rate increases. Students often feel very shy to ask any questions or to share their problems so it is another problem of the e-learning environment.

So, in this research, I tried to solve this issue by detecting student engagement through their facial expressions by that learner's image that was captured in a built-in camera for that I tried to use an existing model that will make the detection more correct by giving a more accurate result.

1.5 RESEARCH QUESTION

In this research paper, for detecting student engagement through their facial expressions I tried to implement a deep learning model to get a more accurate result. So, the research question is:

- Is this model performs better for classifying the FER2013 images?

1.6 RESEARCH OBJECTIVES

The main objective of my research paper is to get better accuracy with the model that has been used in our work and to find student engagement and help educators or e-learning service providers to know about their state through their facial expressions.

So, the objectives of the research are given below:

- Dataset collection
- Train dataset with VGG16
- Save the model's progress
- Classify emotions to understand their emotion
- Engagement prediction
- To get better accuracy

1.7 RESEARCH SCOPE

The main scopes of this research are as follows:

- It will help educators to understand the learner’s state.
- The e-learning service providers will be more conscious of their course quality.
- It will prevent dropout
- It can enhance the efficiency of an automated system

1.8 THESIS ORGANIZATION

In this first chapter, a certain part of the student engagement detection system through facial expressions and their usage, the background of this work, the motivation of this research, the problem statement, research questions, research objectives, and research scopes are discussed. The other parts that are related to our research are given below:

In the next chapter, I will discuss the literature review of some research studies that were already done before in the same field of engagement detection through facial emotions and about a better accuracy model. I tried to compare their dataset, methodology, and gaps on the basis of their work. Then I will discuss my research methodology which is about data collection, data preprocessing, and architecture. But the result of the methodology will be discussed in chapter four. Finally, the last chapter is about the conclusion of my work where I will explain briefly the total summary of my work including future work and the limitations of my work.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

A researcher reviews earlier work, research, conference papers, books, articles, etc. in a literature review. With it, one can learn what research has previously been done on the subject, give a general overview of it, and identify any gaps in the work. After analysis, they might focus on limits and find ways to get around them to improve results.

2.2 PREVIOUS LITERATURE

Automating the identification of learners' participation in traditional and online learning environments was the topic of a number of recent research papers. The methods described in the literature can be categorized into three primary groups: computer vision-based methods, physiological and neurological sensor readings, and engagement tracking.

In a study (Divjak and Bischof, 2009), they examined and evaluated three factors—eye tracking, head movement, and eye closure duration—in order to create an alert that would sound when they discovered that a user was suffering from "computer vision syndrome." The head and eye motions were localized using Open CV, and threshold values were set for each; if a movement exceeds the minimum threshold value, an alert will be issued to alert the user.

Using the Local Binary Point (LBP) algorithm, (Turabzadeh et al, 2018) focused on real-time facial emotion recognition. LBP features were extracted from the video footage and then utilized as input for a K-Nearest Neighbor (K-NN) regression using dimensional labels. Using MATLAB Simulink, the system's accuracy reached 51.28%, while it was just 47.44% in the Xilinx simulation.

Engagement detection techniques can be broadly categorized into three groups: manual, semi-automated, and automatic. Our earlier study included a thorough examination of the aforementioned classification (Dewan et al, 2019).

The manual category describes engagement detection strategies that necessitate learners' active participation. Self-reporting methods and observation checklists fall under this category. Instead of the learners, the observational checklist uses questions that are answered by outside observers (such as teachers). Students answer a series of questions in the self-reporting to describe their own levels of focus, distraction, enthusiasm, or boredom (D'Mello et al, 2006; Grafsgaard et al, 2012; O'Brien et al, 2010). Because it is simple to administer and can offer some helpful information about students' participation, self-reporting is of significant interest to many academics (Larson et al,1991; Shernoff et al, 2014). Their reliability, however, is contingent on a variety of variables, including the learners' candor, desire to disclose their feelings, and the correctness of their perception of their emotions (D'Mello et al, 2014).

Using an automated gazing system (Bidwell and Fuchs, 2011) assessed the level of student participation. Using recorded video from classrooms, they created a classifier for measuring student attention. To gather the students' focus, they employed a face-tracking device. For the purpose of training a hidden Markov model, the automated gaze pattern that resulted was compared to the pattern generated by the observations of a panel of experts (HMM). HMM produced a subpar classification, though; they intended to create 8 unique behavior categories, but could only determine if a student is "involved" or "not engaged."

Students' concentration is checked via eye and head movements by (Krithika et al, 2016) which generates an alarm for low concentration. The video was split up into frames before being subjected to analysis. The implementation was carried out in MATLAB utilizing several face detection and Violas-Jones feature detection routines. The technique is effective enough to identify students' unfavorable sentiments in e-learning settings.

The VilolaJones face identification algorithm is used (Kamath, 2016) to analyze the input photos, followed by a Histogram of Oriented Gradients (HOG) for facial representation for the patch to obtain the final vector of features. The instance-weighted Multiple Kernel Learning-Support-Vector Machine (MKL-SVM) was trained using those features to create a model, and the system's performance was then evaluated. Their best accuracy was 50.77%, while their average accuracy was 43.98%.

A non-intrusive continuous method of photographing people's faces while they use cell phones, computers, and even automobiles is provided by cameras. Numerous methods have been developed to automate this assessment process. This face information can be used to comprehend specific mental states of the user. In particular, geometric-based and appearance-based methods are employed to analyze facial expressions (D'Mello et al, 2014). The forms and placements of facial features, as well as fixed facial locations like the corners of the eyes and eyebrows, are used in geometric-based approaches (Littlewort et al, 2004; Valstar et al, 2007). By examining surface changes on the face in both static and dynamic space, appearance-based algorithms can identify facial expressions. There have been reports of facial expression recognition systems that use appearance-based elements in (Bartlett et al, 2006). The current appearance-based approaches employ a variety of features, including optical flow, active appearance models, and Gabor wavelet coefficients. Different techniques were examined (Bartlett et al, 2006). including explicit feature measurement, independent component analysis, and Gabor wavelets.

In the literature, a number of aspects of learners' engagement have been covered (Bosch 2016; Fredrick et al. 2004; Anderson et al. 2004). Engagement is categorized by Bosch (2016) into three categories: affective, behavioral, and cognitive. In their study investigations, Anderson et al. (2004) defined engagement as academic, behavioral, cognitive, and psychological, in contrast to Fredrick et al. (2004) who defined it as behavioral, cognitive, and emotional. Academic engagement refers to academic identification (e.g., getting along with teachers) and participation (e.g., time on tasks, not skipping classes) towards learning, whereas affective engagement refers to the emotional attitude, for instance, being interested in a topic and enjoying learning about it (Bosch 2016)). (Al-Hendawi 2012). The concept of behavioral engagement is based on participation, which includes taking part in class and extracurricular activities, maintaining focus, submitting projects on time, and listening to instructions from the instructor (Christenson et al. 2012). Cognitive engagement is the consideration and readiness to put up the effort required to grasp challenging concepts and develop challenging skills, such as focused attention, memory, and creative thinking (Anderson et al. 2004). Positive and negative responses to instructors, classmates, and academic performance are all included in emotional involvement (Fredrick et al. 2004). Relationships with professors, peers, and a sense of belonging are all examples of psychological involvement (Christenson and Anderson 2002).

Understanding various forms of involvement in the context of learning is helpful for designing personalized interventions that will enhance learners' experiences. But there must be a means to gauge learner involvement for research that concentrate on it (Harris 2008). One of the two categories of information outlined by engagement theorists—internal to the person (cognitive and affective) and externally observable elements (perceptible facial characteristics, postures, utterances, and actions)—can be used to accomplish this (Bosch 2016). Additionally, other studies underlined that combining observational data with personal information about an individual, such as self-reports, is necessary to measure engagement (Whitehill et al. 2014).

CHAPTER 3

RESEARCH METHODOLOGY

3.1 INTRODUCTION

Discovering the above topic, we can conclude that in online learning student concentration can be categorized into two categories, engaged students and non-engaged students. Our focus is to provide a model with better accuracy to detect that will give the more accurate result and predict student engagement from their facial expression by extracting features from the laptop computer's built-in web camera which works in real-time. In this study, we collected the publicly available dataset from Kaggle that works efficiently in real-time scenarios and utilized it in the VGG16 architecture.

3.2 RESEARCH SUBJECT AND INSTRUMENTS

In recent years the response using machine learning and deep learning architectures and algorithms is having a huge response for classifying, detecting or predicting something. This study is to have better accuracy given the model and detecting engagement through facial emotion. For that, I employed VGG16 architecture based on deep learning. The implementation of all of this work was performed in Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz, Installed RAM 4.00 GB (3.88 GB usable), system type 64-bit operating system, x64-based processor and running under Windows 10 Pro operating system. For the backend, I used the python programming language with several libraries such as Keras, TensorFlow, matplotlib, Pandand NumPy, etc. To implement this, I used Jupyter Note,book and to open this I used Anaconda Prompt.

3.3 DATA COLLECTION

The samples for training the model initially, we used one of the standard benchmarks and a publicly available dataset FER2013 (Facial Expression Recognition). The FER2013 dataset was launched at the Representation learning challenge of ICML which is the Kaggle facial expression recognition challenge in 2013. The images of this dataset were collected by searching API automatically from Google which makes the dataset dynamic and large with the environment variations and population. The dataset consists of 35887 samples that are 48 x 48 grayscale images and each of the images contains the following labels: happy, sad, surprise, anger, disgust, fear, and neutral. But for our architecture, we re-sized all of the images which are 32 x32. The mentioned samples were split into two sets that are training set and the testing set. The format of distributions of the dataset samples is given in Table which shows the overview of different emotions as delineated in the dataset.

Table 1: FER2013 Dataset's Summary

Name	Year	No. of Images	Image Size (Pixels)	Color	Face Variations
FER2013	2013	35887	48 x 48	Grey	Happy, Sad, Surprised, Anger, Disgusted, Fear, and Neutral

Table 2: Split Data from FER2013

Data	Number of Samples
Training	28709
Testing	7178

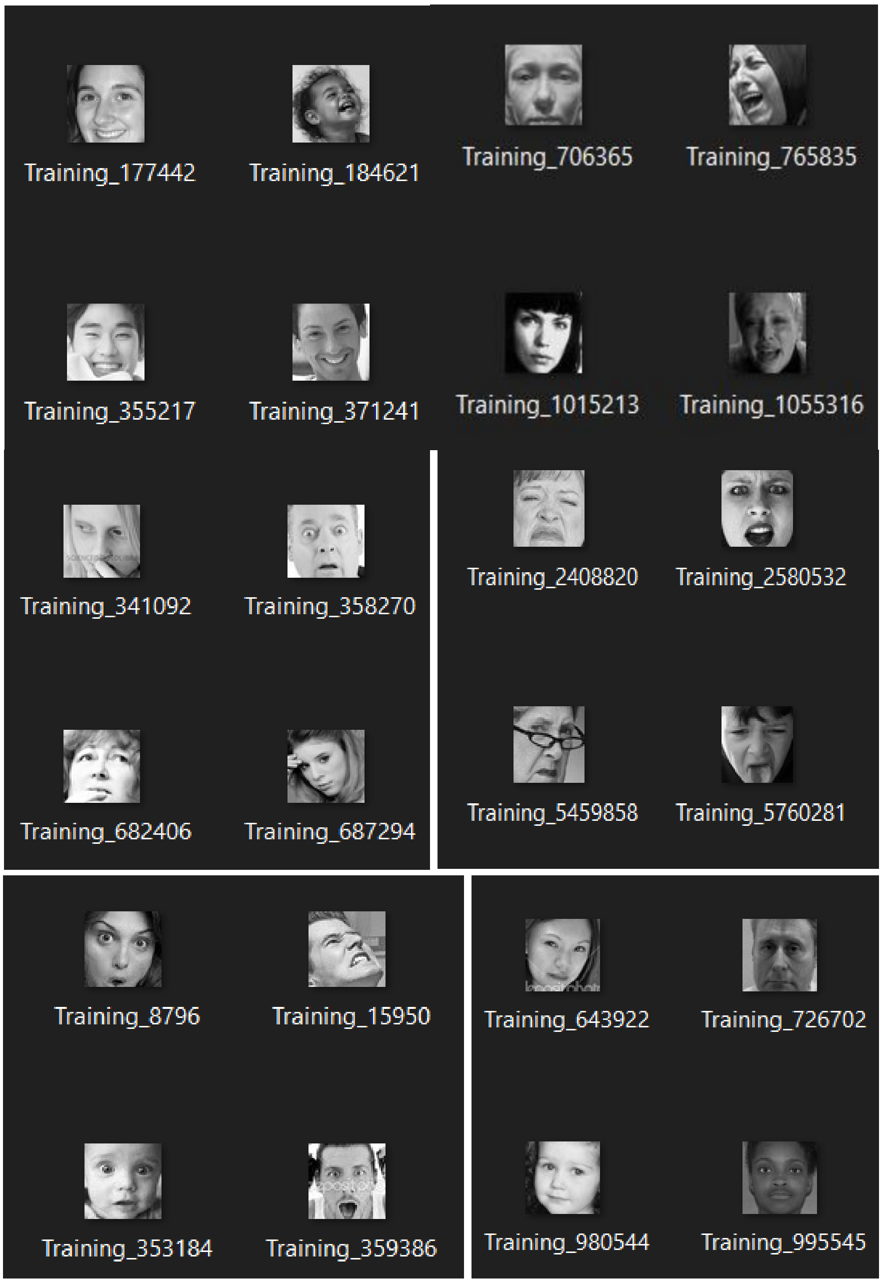


Figure 1: FER2013 Dataset’s Sample Images

3.4 DATA PREPROCESSING

Data preprocessing is a significant phase in the process as the data need to convert into a usable format. Data preprocessing is a stage of data mining. After collecting the dataset from Kaggle we preprocessed our dataset. This is the stage where we equip our data to train for the machine so that the machine can learn those data spontaneously. Creating a machine learning model is another significant part. If the data is raw or unprocessed then we must have to preprocess the dataset to make learning the machine and processed data increased the model accuracy. The FER2013 dataset is already classified into 7 types and they are happy, sad, surprise, anger, disgust, fearful, and neutral. This data was already divided into training and testing sets. In the training set, there have 80% of the data and in the testing set, there have 20% of the data, and both sets have seven folders of seven types of facial emotions data. In the training dataset, the happy dataset has 14430 data, neutral has 9930, surprise has 6342, fear has 8194, angry has 7990, disgust has 872, sad has 9660 and in the testing set, the happy dataset has 3548 data, neutral has 2466, surprise has 1662, fear has 2048, angry has 1916, disgust has 222, sad has 2494.

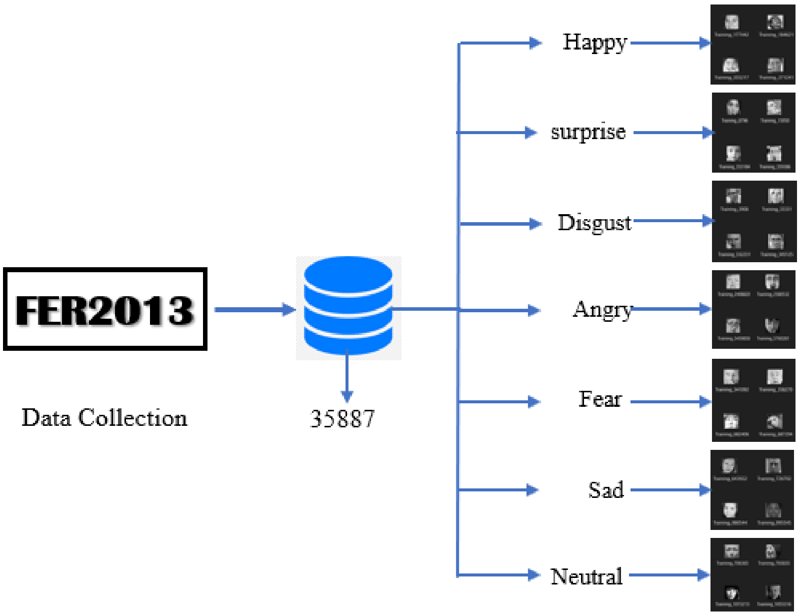


Figure 2: Classes of FER2013 Dataset

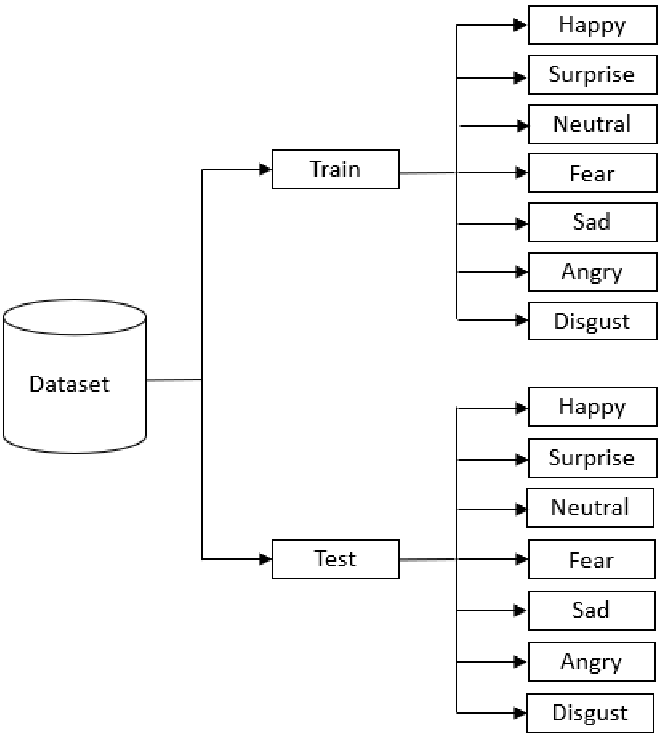


Figure 3: Labelled Data

3.4.1 DATA RESIZING

To get better results data resizing is another important task in image processing as using this function we can get another image with new dimensions without modifying the used image. This dataset was resized into (32,32) format.

3.4.2 DATA SHUFFLING

Here I have used the shuffle function to shuffle the sequence of the dataset. It will change the element’s position in the sequence so that we can get better accuracy.

3.4.3 DATA NORMALIZATION

In the dataset, features may have different units or scales for that dataset that need to normalize to improve the model's performance. So I have used astype() function which is used to cast an object towards a specified datatype that is float32.

3.4.4 DATA AUGMENTATION

Data augmentation is applied to the training dataset which increases the diversity and amount of training data by applying some realistic but random transformation. This technique is attained from original images with some different kinds of minor geometric transformation to increase the diversity of the training dataset and they are translation, rotation, the addition of noise, flipping, etc.

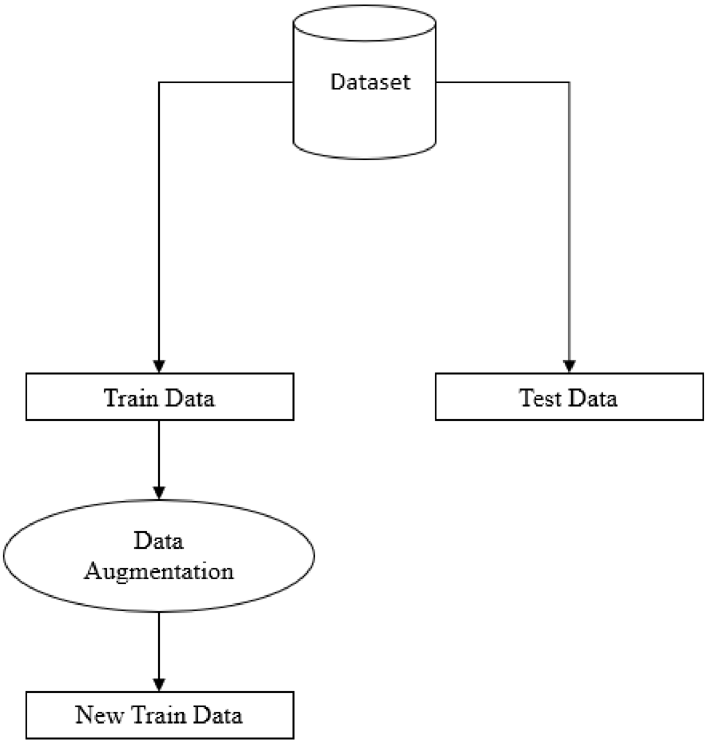


Figure 4: Data Augmentation

3.4.5 DATA PICKLE

This technique allows tracking of the objects that are already serialized so that the objects can be referenced without being serialized again. It consumes the execution time. Here, this technique is used to store the preprocessed data and to load that preprocessed data.

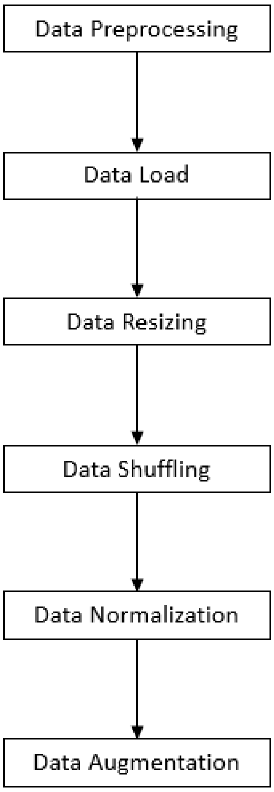


Figure 5: Data Preprocessing Diagram

3.5 CONVOLUTIONAL NEURAL NETWORK(CNN)

For our data classification, I am using the convolutional neural network architecture. So, the basic idea of CNN (convolutional neural network) is explained below:

Convolutional neural network (ConvNet or CNN) is a network architecture that is commonly used in deep learning to interpret visual data which is a class of ANN (Artificial Neural Network). It learns from data directly and manual feature extraction CNN eliminates the requirement. It is designed to detect automatically a specific part of an object and to learn hierarchies from low to high-level patterns of features. It uses mathematical operations named ‘convolution’ for the generic multiplication of the matrix in at least one of its layers.
A convolutional neural network structure is it has an input layer, some hidden layers, and an output layer. The input layer is the activation function and the output layer is the final convolution so any middle layers are called hidden layers and these hidden layers perform the convolution like this layer convolves the input and passes the result to its next layer.

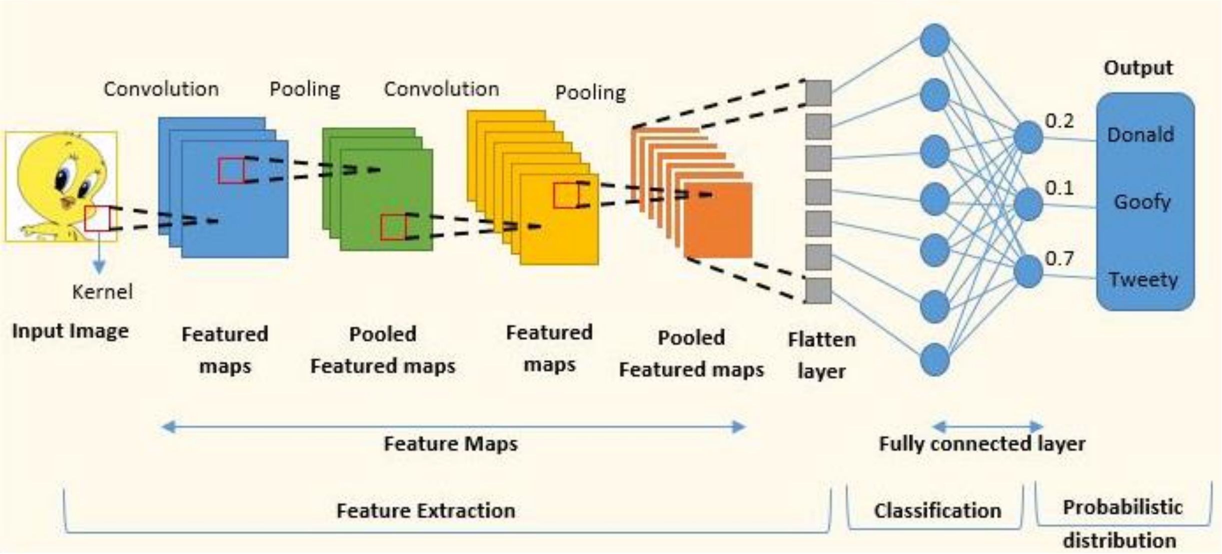


Figure 6: Basic Architecture of Convolutional Neural Network (color figure online)

The convolutional layer is the first layer that is used to extract features after taking input. The image is resized into (M*M) format in this stage. The next layer is the pooling layer and for decreasing the calculating cost the previous image size has been reduced. The largest element will be taken by the max-pooling layer from the feature map. Then the next stage which is the last layer afore the output layer named the fully connected layer. In this layer, the images are flattened which were taken from the previous layer. And the classification process begins here. In the fully connected layer, overfitting may occur after giving a connection. Dropout is used for overcoming this complication. And in this step, the decision will be made that which data needs to fire and which data will be taken for further steps. The activation function is one of the most important parameters and ReLu, tanHn Softmax are some usually utilized activation functions.

3.6 VGG16

CNN is a concept of the neural network whereas VGG is a specific convolutional neural network that is designed for localization and classification. Localization detects objects within an image and classification is classifying images with the label. Simonyan and Zisserman of the Visual Geometry Group Lab of Oxford proposed VGG16 in a paper in 2014 and won 1st and 2nd place in the ILSVRC challenge. Basically, VGG16 is developed for face recognition and is famous for accuracy with small convolutional filters. It is the 16-layer deep neural network which is an advancement over AlexNet where the kernel size filter has been replaced in VGG16. It focused on having a 3x3 filter of the convolutional layer with 1 stride and using the same padding 2x2 filter of the max pool layer of 2 strides. VGG16 follows the arrangement of convolution layers and max pool layer persistently throughout the entire architecture. It has 2 fully connected layers which are followed by a softmax output at the end. Vgg16's layers activation mechanism is in ReLu. So, it is a network that is commonly used for benchmarking other networks and the best pre-trained network for the recognition of images.

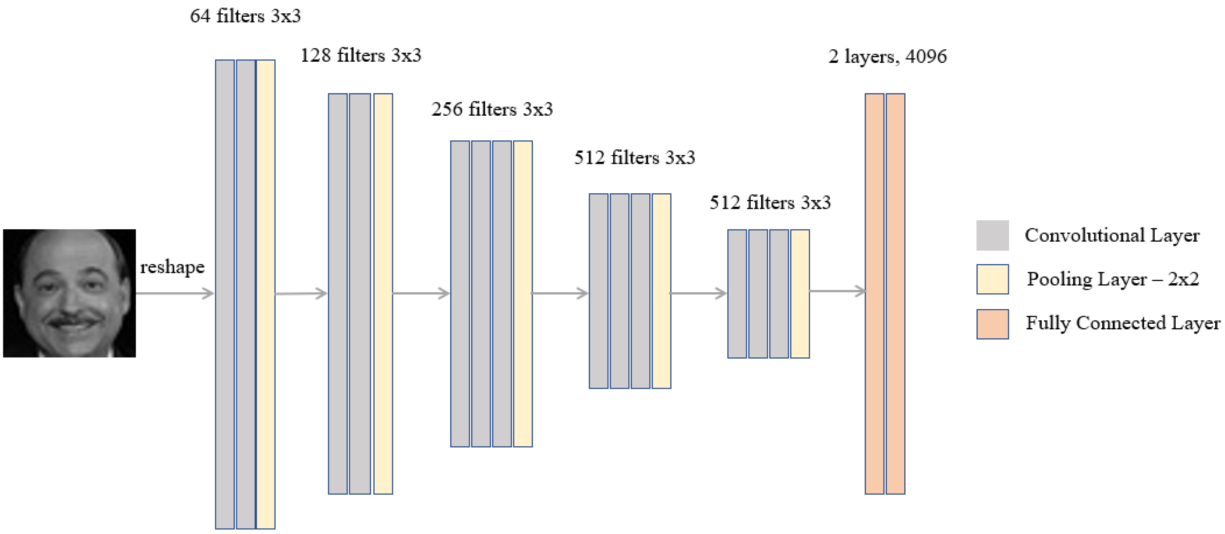


Figure 7: VGG16 Architecture Diagram

VGG16 is one of the network architectures of VGG architecture whose configuration is D.

VGG16 Configuration

Input Image (224×244 pixels, 3 channels Red, Green, Blue)
Convolution Layer: 64 3×3 filters, stride 1, padding 1 **Convolution Layer: 64 3×3 filters, stride 1, padding 1**
Max Pooling Layer: 2×2 with stride 2
Convolution Layer: 128 3×3 filters, stride 1, padding 1 **Convolution Layer: 128 3×3 filters, stride 1, padding 1**
Max Pooling Layer: 2×2 with stride 2
Convolution Layer: 256 3×3 filters, stride 1, padding 1 **Convolution Layer: 256 3×3 filters, stride 1, padding 1** **Convolution Layer: 256 3×3 filters, stride 1, padding 1**
Max Pooling Layer: 2×2 with stride 2
Convolution Layer: 512 3×3 filters, stride 1, padding 1 **Convolution Layer: 512 3×3 filters, stride 1, padding 1** **Convolution Layer: 512 3×3 filters, stride 1, padding 1**
Max Pooling Layer: 2×2 with stride 2
Convolution Layer: 512 3×3 filters, stride 1, padding 1 **Convolution Layer: 512 3×3 filters, stride 1, padding 1** **Convolution Layer: 512 3×3 filters, stride 1, padding 1**
Max Pooling Layer: 2×2 with stride 2
Fully Connected Layer: 4096 nodes
Fully Connected Layer: 4096 nodes
Fully Connected Layer: 1000 nodes (one for each category)
Softmax

Figure 8: The VGG Architecture for Configuration D (VGG16)

The problems we face during the implementation of VGG16:

- It works very slowly while training.
- It took a lot of disk space and bandwidth which makes it inadequate.
- It shows exploding parameters problem as it has 138 million parameters.

3.7 HAAR CASCADE ALGORITHM

An algorithm called Haar cascade can identify objects in photos. Utilizing this algorithm is simpler and easier. This can also operate in real time. A haar cascade detector can be programmed to pick up a variety of signals. like fruit, a car, face detection, etc. Faces may be picked out in a video in real-time or from an image using this object detection algorithm. In their 2001 research publication, "Rapid Object Detection using a Boosted Cascade of Simple Features," Viola and Jones made the initial recommendation for this technique. This program divides the data points into positive and negative categories.

- Positive data points are those that are a part of the object that we have detected which our classifier should recognize.
- Negative data points are those that do not include the thing we are attempting to identify.

This method is based on machine learning, and in it, the classifier is trained using a large number of both positive and negative images. In comparison to current object detection systems, Haar Cascade is not as accurate. However, it may operate in real-time and is considerably simpler to use.

So, I used this algorithm to detect facial emotions from the FER2013 dataset.

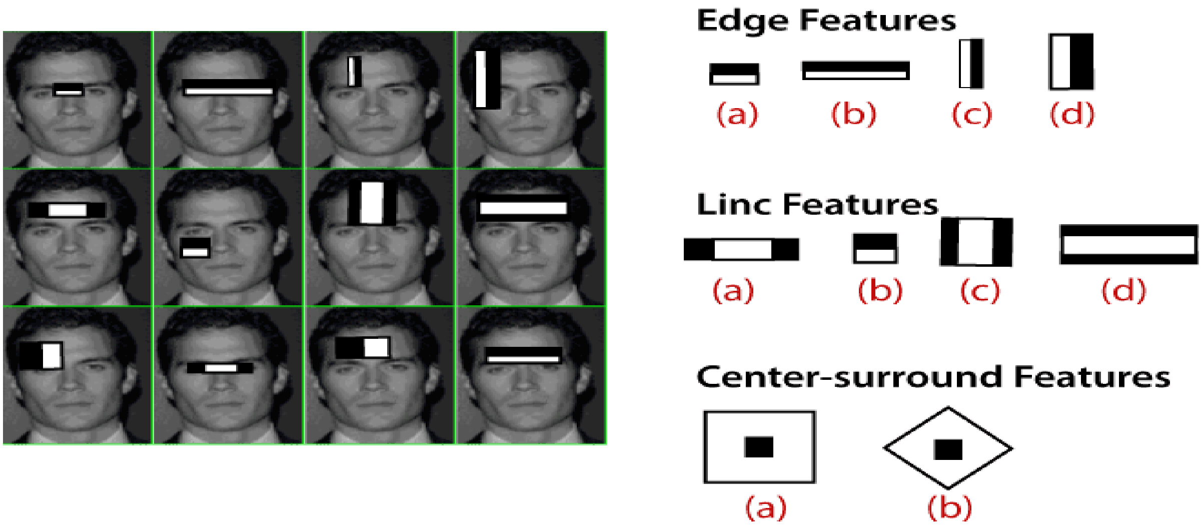


Figure 9: Face Recognition and Face Detection using HAAR Cascade Algorithm

3.8 CATEGORIES OF ENGAGEMENT

By analyzing the facial emotions, the engagement level will be divided into different three categories: engaged, not engaged, and neutral (Sharma et al, 2019). They are described below:

- Engaged: A student will be considered engaged when their facial emotion will be happy, surprised, or fearful.

- Not Engaged: A student will be considered not engaged when their facial emotion will be sad, angry, or disgusted.

- Neutral: A student will be considered neutral only if they have no emotions on their face.

Table 3: This research examines three different levels of engagement and their individual exemplary qualities in the context of learning.

Engagement Type	Level	Facial Emotions
Engaged	1	Happy, Surprise, Fear
Not Engaged	2	Sad, Angry, Disgust
Neutral	3	Neutral

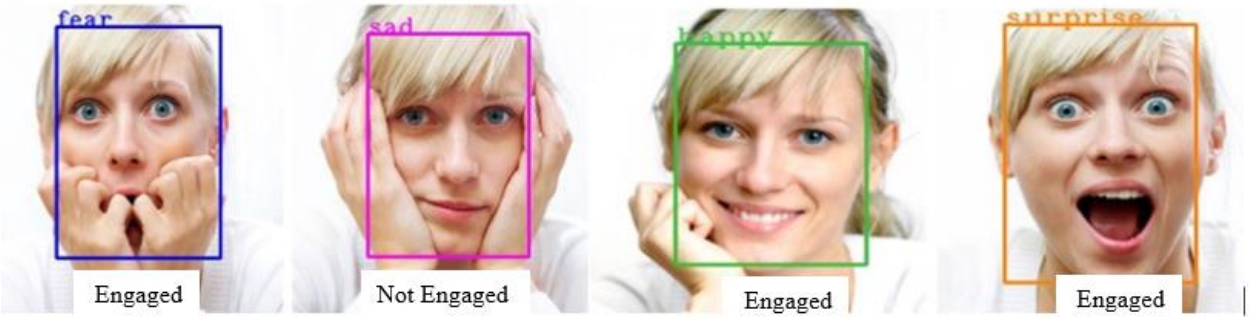


Figure 10: Engagement detection through facial expressions

3.7 IMPLEMENTATION REQUIREMENTS

3.7.1 MODEL CONSTRUCTION

The VGG16 architecture is used to construct a sequential model where Keras was boosted layer by layer in the erection of the model. For input images, Convolutional 2D layers were used to view the images as a matrix of 2 dimensions. For output images, the Dense layer was used. The kernel size of this matrix was 3x3 and its image size height and weight must be 32x32. The activation function is ReLU, the flattened layer was used to link between the 2D convolutional layer and the dense layer. Lastly, Softmax was used in this model as an activation function in the output layer which predicts a multinomial probability.

3.7.2 MODEL COMPILATION

To compile the model three parameters such as optimizer, loss function, and metrics were used. An optimizer utilizes Adaptive moment estimation and manages the learning rate and it reduces the loss and improves accuracy. For the loss function, we utilized sparse categorical cross-entropy so that the integer-based target doesn’t convert into other formats. And finally, the accuracy matrix was used to see the train or validation sets score when the model was training.

3.7.3 MODEL TRAINING

The fit () function was used to train the model with trained and validation data, epochs, and batch size parameters. The epoch was declared with a specific number and it needs to increase or decrease a certain number of times to improve the model.

3.7.4 MODEL COMPILED PARAMETERS SUMMARY

Now I will explain how I compile this version in this research using VGG16:

I run 24 epochs of my model architecture after defining the emotion labels, where at first an input Conv2D layer (64 filters) paired with a MaxPooling 2d layer changed into exceeded, then 4 pairs of Conv2D (128, 256, and 512 filters) and Max Pooling 2d layers were run, then 2 Dense layer with 4096 nodes, and finally as a dense output layer switched into a pop-out with 29 nodes.

The loss and the accuracy of the discussed model have given below:

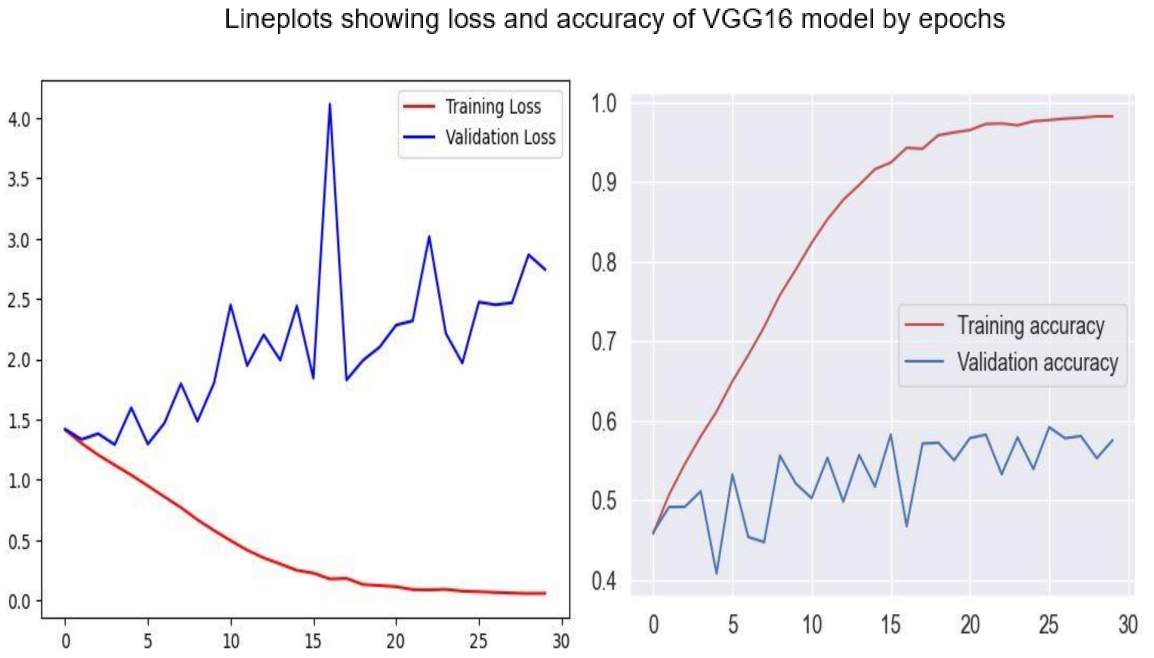


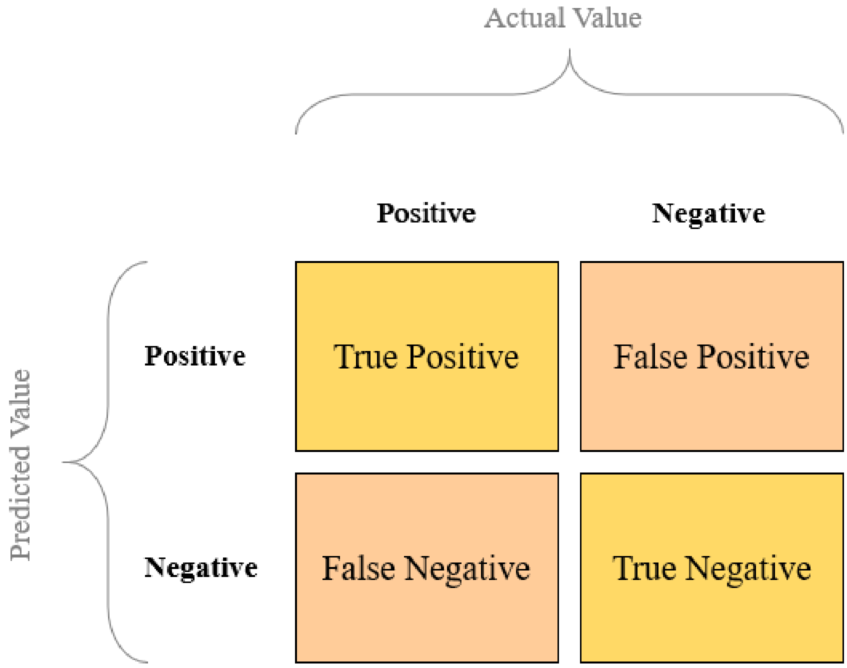
Figure 11: Loss and Accuracy of my model for emotion prediction

3.8 EVALUATION METRICS

I plotted the confusion matrix to evaluate the results. Evaluation requires both true positive and negative results as well as false positive and negative values. This is a real plus because the actual amount was as predicted. rightfully rejecting a true negative Although a positive number is predicted, the value is actually a false positive. False-negative is wrongfully disregarded.

3.8.1 CONFUSION MATRIX

To evaluate a classifier's performance, it is much better to look at the confusion matrix. The major objective is to ascertain the frequency with which examples of class A are classified as class B examples.
To create a confusion matrix, we require the following four characteristics:



	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

Figure 12: Confusion Matrix

- True Positives (TP): This occurs when the model correctly predicted both the label and the ground truth.

- True Negatives (TN): The label is not based on the truth and was not anticipated by the model.

- False Positives (FP): These are instances where a model predicts a label but the label is not actually present in the data (Type I Error).

- False Negatives (FN): Labels that the model does not predict but are true in reality (Type II error).

3.8.2 ACCURACY

The accuracy of the model determines how well a computer can predict outcomes. Something is significant when each class is equally important. For me, every course is equally important to my work. In order to assess the model's accuracy, accuracy is essential. For binary classification, accuracy can also be assessed in terms of positives and negatives, as shown below:

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

True Positives, True Negatives, False Positives, and True Negatives, respectively, are denoted by the letters TP, TN, FP, and FN.

3.8.2 LOSS FUNCTION

In order for a neural network to learn something, it must first undergo training. Then, by testing, it can be determined whether or not the training was effective enough to enable the model to perform its intended function. Additionally, this loss function evaluates how successfully the data was trained by comparing the target and projected output. And a model tries to reduce the amount of loss in the train and test data by finishing each epoch.

CHAPTER 4

RESULT AND DISCUSSION

4.1 INTRODUCTION

The detection of facial emotion is a state that shows the result if a student is engaged or not. This chapter discusses and presents the result of our research experiment. In this study, I used a sequential VGG16 architecture model. This model experiments with the test and train accuracy and at the end of the training, I find my expected outcome. In this section, I will discuss the outcome and will analyze it briefly.

4.2 EXPERIMENTAL SETUP

4.2.1 TRAINING PROCESS

The dataset was split into training data of 80% and validation data of 20% respectively. The validation data has also split into test data of 10%. And 24 epochs were used to measure the training progress.

4.2.1.1 TRAINING AND VALIDATION ACCURACY

In this graph, 0 and 1.0 can be considered the worst and best values for training, and 0 and 0.6 can be considered the worst or best values. In the first epoch, the accuracy of training and validation are 0.23 and 0.26 respectively. And after 24 epoch the accuracy of trained model was 0.96 and the accuracy of validation was 0.53.

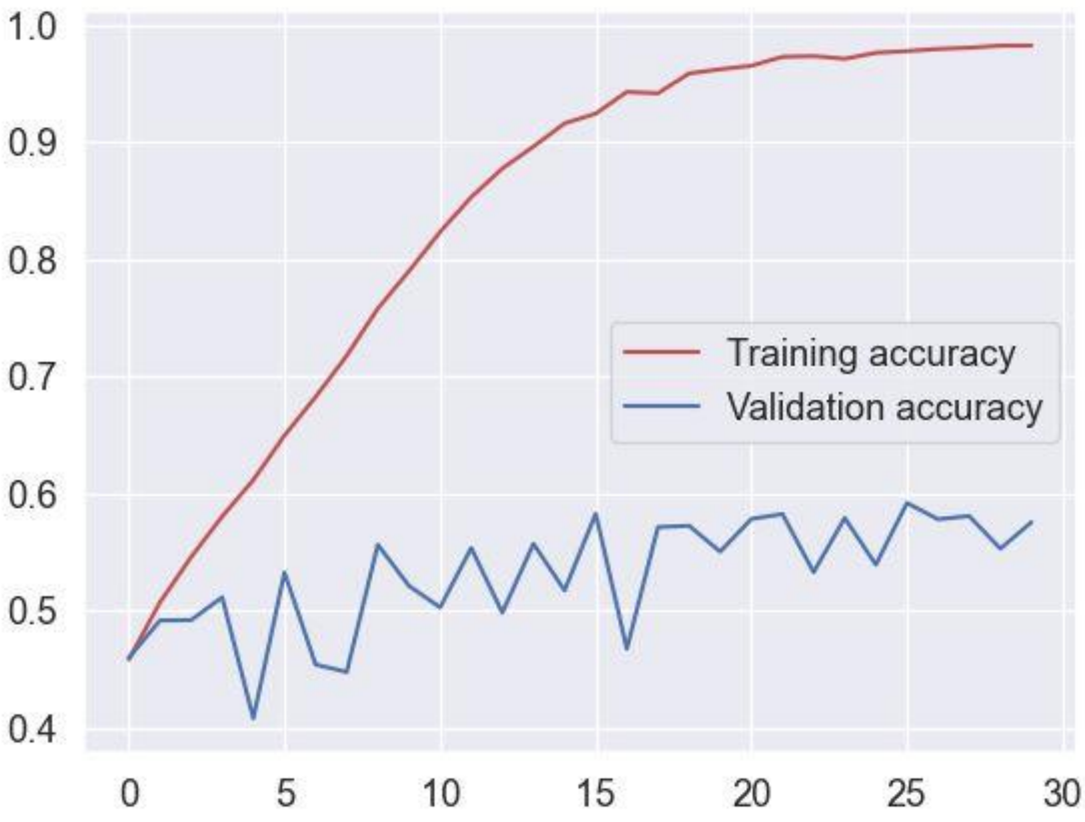


Figure 13: Accuracy of training and validation of the model

4.2.1.2 TRAINING AND VALIDATION LOSS

In this graph, after completing one epoch the loss is 1.90 and it dropped 0.11 after 24 epochs. And for validation loss, it is 70.51 after one epoch and after completing 24 epochs is 70.71.

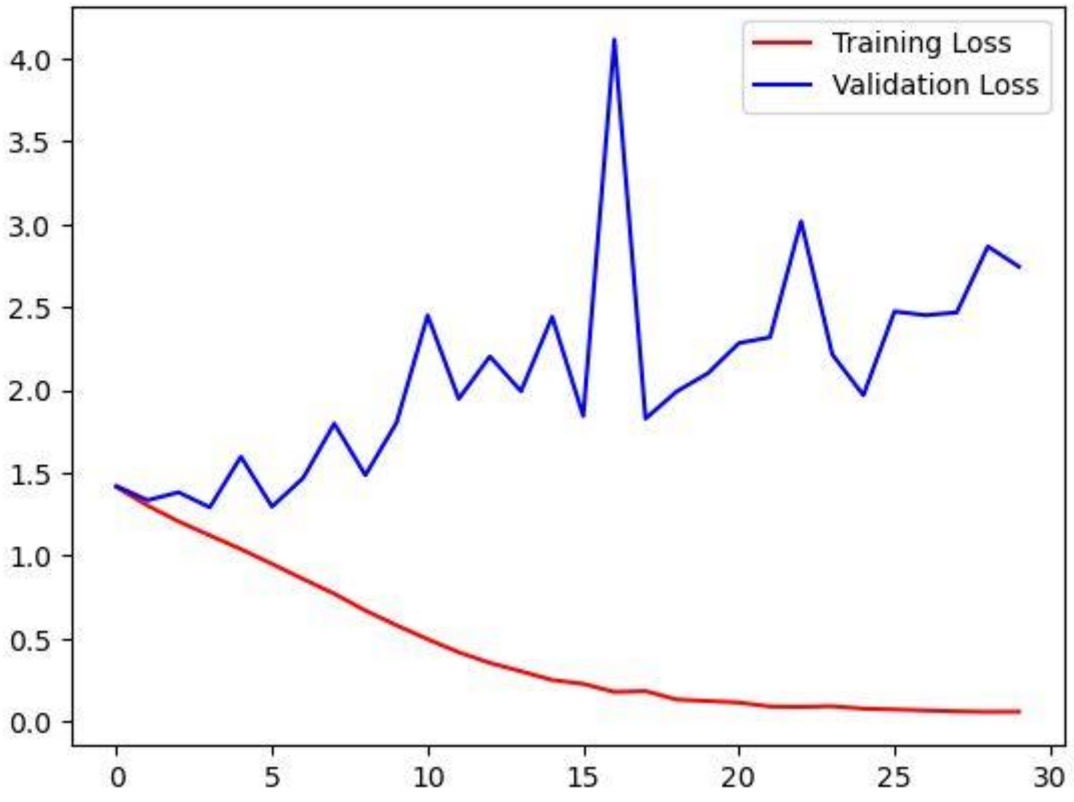
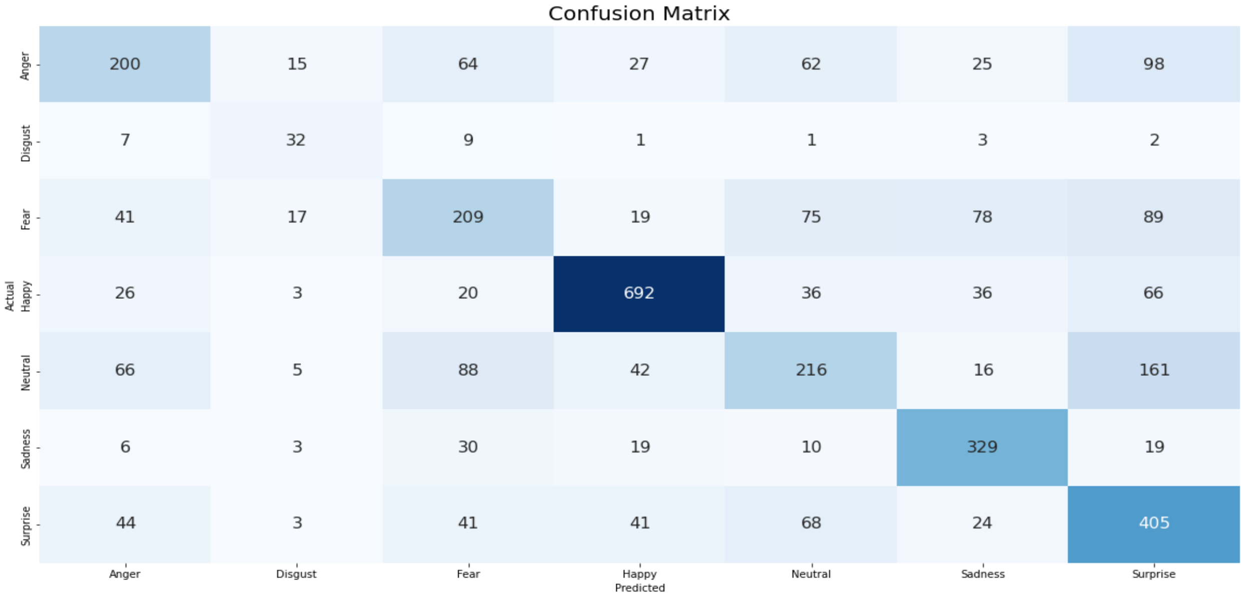


Figure 14: Loss of training and validation of the model

4.3 CONFUSION MATRIX

For evaluating the effectiveness of deep and machine learning for classification, the confusion matrix is crucial. We can better understand the kind of errors our system produces and the relevant responses by calculating the confusion matrix. Here, we can see that the accuracy is quite good in emotion happy. After that the accuracy for surprise and sadness is good. And the accuracy in anger, fear, and neutral is not bad. But the accuracy is not that much good for the emotion of disgust which is not expected.

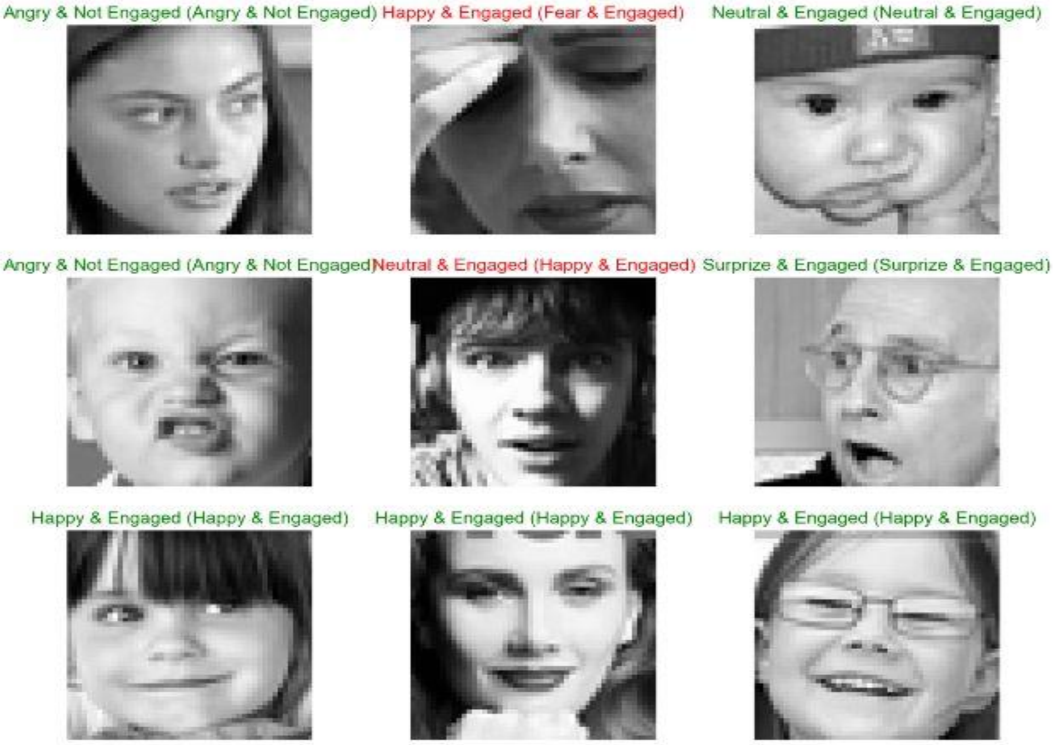


Confusion Matrix

Actual \ Predicted	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Anger	200	15	64	27	62	25	98
Disgust	7	32	9	1	1	3	2
Fear	41	17	209	19	75	78	89
Happy	26	3	20	692	36	36	66
Neutral	66	5	88	42	216	16	161
Sadness	6	3	30	19	10	329	19
Surprise	44	3	41	41	68	24	405

Figure 15: Confusion matrix of Facial Emotion recognition using VGG16 model

4.4 TEST RESULT



Angry & Not Engaged (Angry & Not Engaged)
Happy & Engaged (Fear & Engaged)
Neutral & Engaged (Neutral & Engaged)
Angry & Not Engaged (Angry & Not Engaged
Neutral & Engaged (Happy & Engaged)
Surprize & Engaged (Surprize & Engaged)
Happy & Engaged (Happy & Engaged)
Happy & Engaged (Happy & Engaged)
Happy & Engaged (Happy & Engaged)

Figure 16: Actual vs Predicted level

4.5 OUTPUT

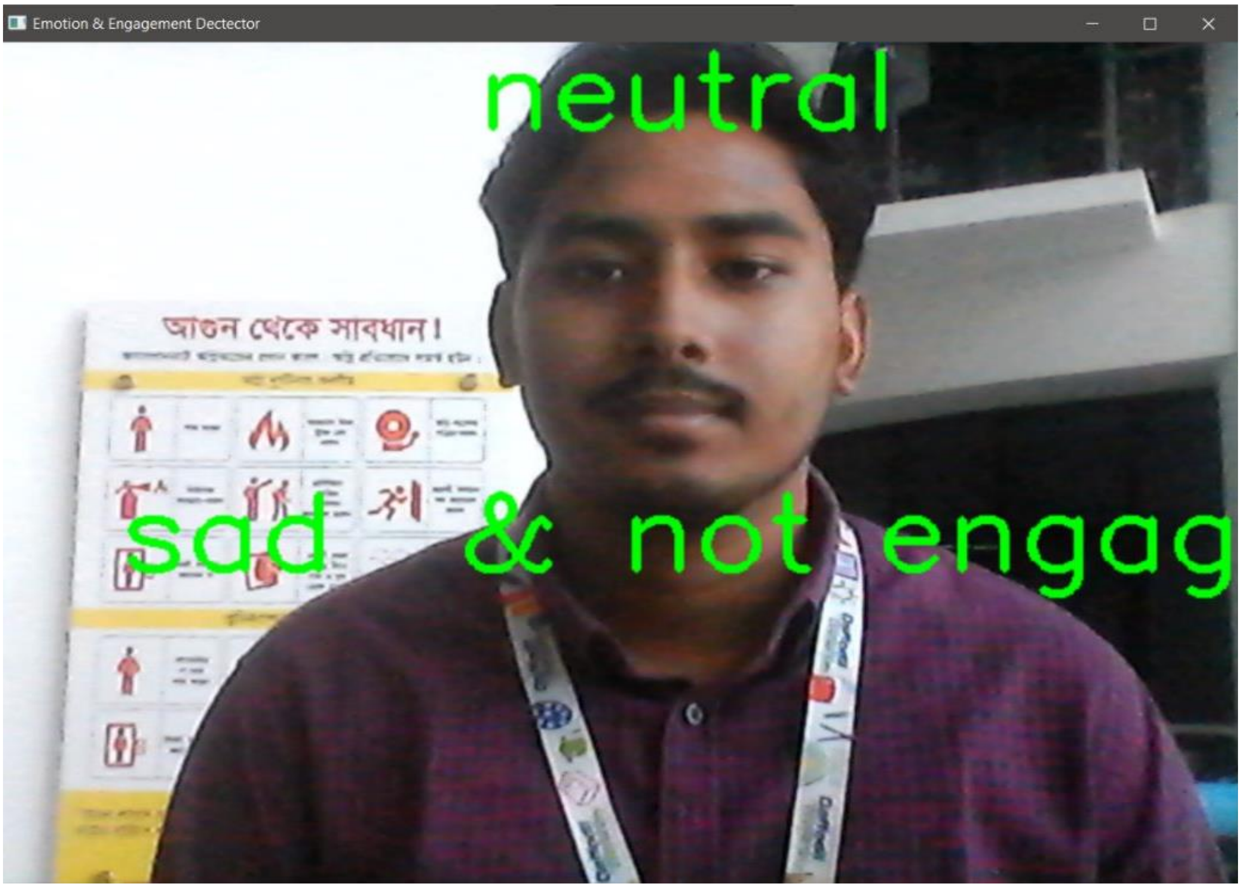


Figure 17: Output from Webcam using OpenCV

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

The ability of e-learning systems to intelligently adapt to the learner in online learning depends on reliable models that can identify learners' engagement during a learning session, especially in situations where there is no instructor present. In this research, we investigated one of the popular models VGG16 for facial emotion recognition, and through their facial expression, the student’s engagement will be classified. In this study, three-level (engaged, not-engaged, and neutral) decisions will be given through their facial emotion recognition, where the model shows the better accuracy which is 97.14%. I hope that this model will perform efficiently and will predict facial emotion for engagement detection more accurately.

5.2 FUTURE SCOPE

In the future, I would like to use some other model to get a more efficient result of facial emotions and analyze some other features like eye gaze, and head movement to get a more accurate result for engagement detection.

CHAPTER 6

REFERENCES

1. A Crowdsourced Approach to Student Engagement Recognition in E -learning Environments

Kamath, A., Biswas, A., & Balasubramanian, V. (2016, March). A crowdsourced approach to student engagement recognition in e-learning environments. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-9). IEEE.

2. A phenomenographic investigation of teacher conceptions of student engagement in learning

Harris, L. R. (2008). A phenomenographic investigation of teacher conceptions of student engagement in learning. *The Australian Educational Researcher*, *35*(1), 57-79.

3. Boredom in the middle school years: blaming schools versus blaming students

Larson, R. W., & Richards, M. H. (1991). Boredom in the middle school years: Blaming schools versus blaming students. *American journal of education*, *99*(4), 418-443.

4. Check and connect: The importance of relationships for promoting engagement with school.

Anderson, A. R., Christenson, S. L., Sinclair, M. F., & Lehr, C. A. (2004). Check & Connect: The importance of relationships for promoting engagement with school. *Journal of school psychology*, *42*(2), 95-113.

5. Classroom analytics: Measuring student engagement with automated gaze tracking

Bidwell, J., & Fuchs, H. (2011). Classroom analytics: Measuring student engagement with automated gaze tracking. *Behav Res Methods*, *49*(113).

6. Combined support vector machines and hidden Markov models for modeling facial action temporal dynamics,"

Valstar, M. F., & Pantic, M. (2007, October). Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. In *International workshop on human-computer interaction* (pp. 118-127). Springer, Berlin, Heidelberg.

7. Confusion can be beneficial for learning

D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153-170.

8. Detecting Student Engagement: Human Versus Machine (Conference on User Modeling Adaptation and Personalization

Bosch, N. (2016, July). Detecting student engagement: human versus machine. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (pp. 317-320).

9. Dynamics of facial expression extracted automatically from video," Image and Vision Computing,

Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J., & Movellan, J. (2004, June). Dynamics of facial expression extracted automatically from video. In *2004 Conference on Computer Vision and Pattern Recognition Workshop* (pp. 80-80). IEEE.

10. Engagement detection in online learning: a review," Smart Learning Environment

Dewan, M., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: a review. *Smart Learning Environments*, *6*(1), 1-20.

11. Eye blink based fatigue detection for prevention of computer vision syndrome

Divjak, M., & Bischof, H. (2009, May). Eye Blink Based Fatigue Detection for Prevention of Computer Vision Syndrome. In *MVA* (pp. 350-353).

12. Facial expression emotion detection for real-time embedded systems

Turabzadeh, S., Meng, H., Swash, R. M., Pleva, M., & Juhar, J. (2018). Facial expression emotion detection for real-time embedded systems. *Technologies*, *6*(1), 17.

13. Fully automatic facial action recognition in spontaneous behavior

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006, April). Fully automatic facial action recognition in spontaneous behavior. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)* (pp. 223-230). IEEE.

14. Handbook of Research on Student Engagement

Christenson, S., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement* (Vol. 840). New York: Springer.

15. Multimodal analysis of the implicit affective channel in computermediated textual communication,

Grafsgaard, J. F., Fulton, R. M., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2012, October). Multimodal analysis of the implicit affective channel in computer-mediated textual communication. In *Proceedings of the 14th ACM international conference on Multimodal interaction* (pp. 145-152).

16. Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue

D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, *16*(1), 3-28.

17. Student emotion recognition system (SERS) for e -learning improvement based on learner concentration metric

Krithika, L. B., & GG, L. P. (2016). Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. *Procedia Computer Science*, *85*, 767-776.

18. Student engagement in high school classrooms from the perspective of flow theory

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2014). Student engagement in high school classrooms from the perspective of flow theory. In *Applications of flow in human development and education* (pp. 475-494). Springer, Dordrecht.

19. The centrality of the learning context for students' academic enabler skills. School Psychological Review

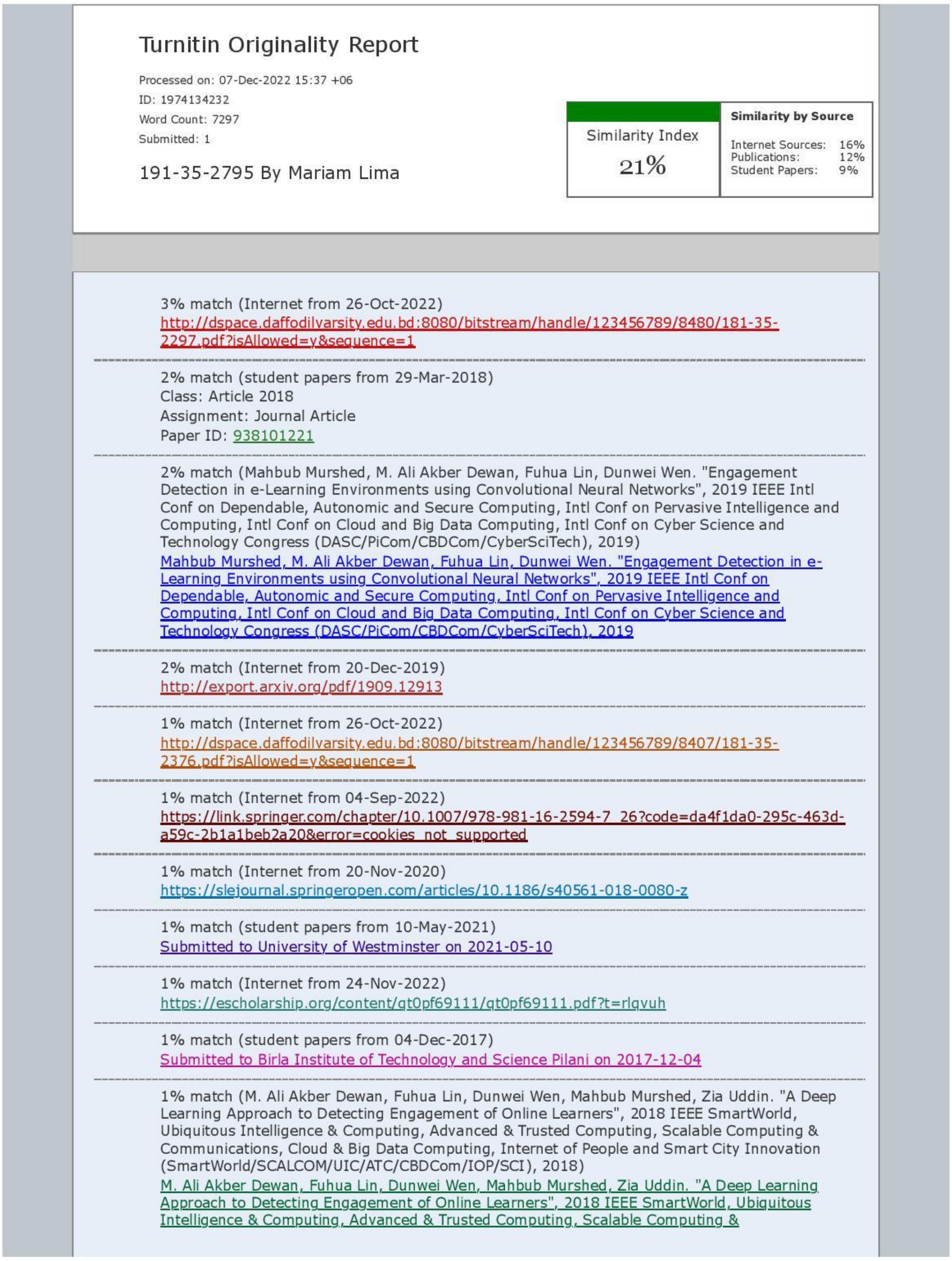
Christenson, S. L., & Anderson, A. R. (2002). Commentary: The centrality of the learning context for students' academic enabler skills. *School Psychology Review*, *31*(3), 378-393.

20. The development and evaluation of a survey to measure user engagement

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, *61*(1), 50-69.

21. The faces of engagement: Automatic recognition of student engagement from facial expressions

Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Transactions on Affective Computing*, *5*(1), 86-98.



Turnitin Originality Report

Processed on: 07-Dec-2022 15:37 +06
ID: 1974134232
Word Count: 7297
Submitted: 1

191-35-2795 By Mariam Lima

Similarity Index	Similarity by Source	
21%	Internet Sources:	16%
	Publications:	12%
	Student Papers:	9%

3% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8480/181-35-2297.pdf?isAllowed=y&sequence=1

2% match (student papers from 29-Mar-2018)
Class: Article 2018
Assignment: Journal Article
Paper ID: 938101221

2% match (Mahbub Murshed, M. Ali Akber Dewan, Fuhua Lin, Dunwei Wen. "Engagement Detection in e-Learning Environments using Convolutional Neural Networks", 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), 2019)
Mahbub Murshed, M. Ali Akber Dewan, Fuhua Lin, Dunwei Wen. "Engagement Detection in e-Learning Environments using Convolutional Neural Networks", 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), 2019

2% match (Internet from 20-Dec-2019)
http://export.arxiv.org/pdf/1909.12913

1% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8407/181-35-2376.pdf?isAllowed=y&sequence=1

1% match (Internet from 04-Sep-2022)
https://link.springer.com/chapter/10.1007/978-981-16-2594-7_26?code=da4f1da0-295c-463d-a59c-2b1a1beb2a20&error=cookies_not_supported

1% match (Internet from 20-Nov-2020)
https://slejournal.springeropen.com/articles/10.1186/s40561-018-0080-z

1% match (student papers from 10-May-2021)
Submitted to University of Westminster on 2021-05-10

1% match (Internet from 24-Nov-2022)
https://escholarship.org/content/qt0pf69111/qt0pf69111.pdf?t=rlqvuh

1% match (student papers from 04-Dec-2017)
Submitted to Birla Institute of Technology and Science Pilani on 2017-12-04

1% match (M. Ali Akber Dewan, Fuhua Lin, Dunwei Wen, Mahbub Murshed, Zia 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 (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2018)
M. Ali Akber Dewan, Fuhua Lin, Dunwei Wen, Mahbub Murshed, Zia Uddin. "A Deep Learning Approach to Detecting Engagement of Online Learners", 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing &