Ministry of Higher Education and Scientific Research
University of Babylon
College of Science for Women
Department of Computer Science

Students Engagement Detection During Computer Interaction Using Deep Learning Techniques

A Project

Submitted to the Council of the College of Science for Women, University of Babylon in Partial Fulfillment of the Requirements for the Degree of Diploma in Science / Computer Science

Submitted by

Zahraa Ahmed Jasim

Supervised by

Asst.Prof. Dr. Amina Atiya Dawood

2022 A.D 1444A.H.

قال تعالى:

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

(يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ
أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ
خَبِيرٌ)

صدق الله العلي العظيم

سورة المجادلة: الآية١١

Supervisors Certification

I certify that this research entitled “**Students Engagement Detection during Computer Interaction using Deep Learning Techniques**” was done by (**Zahraa Ahmed Jasim**) under my supervision.

Signature:

Name: Asst.Prof. Dr. Amina Atiya Dawood

Date: / / 2022

Address: University of Babylon- College of Science

The Head of the Department Certification

In view of the available recommendations, I forward the research entitled "**Students Engagement Detection during Computer Interaction using Deep Learning Techniques**" for debate by the examination committee.

Signature:

Name: Asst. prof. Dr Saif Mahmood Alalak

Date: / / 2022

Address: University of Babylon/College of Science for Women

Dedications

To my family especially my great parents…

To my teachers and friends,

I dedicate this work.

Zahraa Ahmed Jasim

Acknowledgments

All thanks and praise to Allah, the Lord of the world, who gave me courage and enabled me to achieve this work.

I would like first to express my sincere appreciation to my supervisor **Asst.Prof. Dr.Amina Atiya** for her advice, guidance, continues encouragement, and her generous dedication of precious time through the work of this project.

Thanks and Gratitude to all my tutors and all the staff of the Department of Computer Sciences \ College of Sciences for Women \University of Babylon for their help.

All Thanks and Gratitude to all my family for their support and encouragement.

I would like to thank my friends for helping me.

Zahraa Ahmed Jasim

Abstract

With the growing development in applications of human computer interaction, especially in education domain by using computer-based-learning. Scientist try to develop an intelligent interface that is more interactive, detective and responsive into the users. During pandemic especially in covid-19 and social restrictions, online learning approved it's a great role during this situation. Therefore, it requires to a model has an ability to detect and response to students' spontaneous behaviours automatically from their expressions without interrupt learning process. So, this project considers these requirements and design an affective model to infer students' engagement in computer-based-learning process. The model adapted deep learning techniques due to their powerful and its best engineering to abstract high level of features from raw data against other shallow machine learning techniques. In addition to, these methods need to another algorithm to be as a classifier to get the desire output.

The model based on facial expressions of students which are captured from webcam during interacting with a computer. The high results and a good generality of the model toward unseen data comes from the complexity of the layers' connectivity of deep learning.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Term	Meaning
AI	**Artificial Intelligence**
ANN	**Artificial Neural Network**
CNN	**Convolution Neural Network**
DL	**Deep Learning**
DNN	**Deep Neural Network**
LSTM	**Long Short Term Memory**
ML	**Machine Learning**
RNN	**Recurrent Neural Network**
SVM	**Support Vector Machine**

Chapter one

General Introduction

Chapter One

General Introduction

1.1 Introduction

This chapter will present a short description about the proposed model, and its related information in computer-based learning environment. In addition, the chapter will present the aim of this study and its objectives.

Online learning has significant impact in education domain. AS, online or E-learning is assistance tool that support in a pandemic situation or any emergency situations which happened. The best example of the importance of e-learning can be seen during Covid-19 pandemic. as the pandemic have forced all educational institutions in the world to adapt online learning methods and its applications instead of traditional teaching methods(face-to-face). In online learning, the lectures delivered to students through using an internet, and students interact with their teachers through computers or phone devices. To estimate outcome of learning and students interacting require affective tools that match the virtual learning environment (Sarwar et al., 2020).

In spite of positive advantages for online learning such as replace schools' closedown during disaster and pandemic, and help students to learn at anytime and anywhere. However, it considered as a difficult task and have more challenges than face-to-face learning methods because in traditional methods the teachers can control and monitor their students during the class by real observing. But, in online platforms it is difficult to observe students interaction or engagement during learning session(M. Murshed et al.,2019).

It is known that engagement can increase the learning outcome, also, pedagogical intervention and optimal responses can improve learner engagement (Poon et al.,2022).

In educational domain it is necessary to measure learners' engagement especially in online learning environment. To understand how can take the correct decision which connect with engagement in online teaching. Generally, there are two methods to measure students' engagement: internal and external measurements (S. L. Christenson et al.,2012). Internal methods mean inner feeling of learning during online course, and to measure these feelings require sensitive tools to capture them. Using sensitive tools need specific requirements and environment. Other methods to measure engagement, based on external measurements which depend on learner face, body, and gestures.

Face considers important mean communication during interaction with others. A lot of studies consider human face is an important channel to express emotions(Altuwairqi, K et al.,2021),(M. Murshed et al.,2019).

Researchers in modeling human emotions and psychologists have taken more attention to human face, as it the observable part that contain emotions(Zeng et al., 2009). A common approach that used to recognize face based on still or static images. Static images do not convey a lot of information. Therefore, the demand of videos has widely increased for processing and recognizing emotions. For detecting engagement during online learning require different communication channels to simulate learner emotion, such as facial expressions, head movement, eye gaze, body postures.So, this work used facial expressions to detect engagement.

Previously, more techniques are developed to detect human face and extract facial expressions (features points) such as Fisher faces[6], Bayesian face[38], Eigen faces[59], support vector machine (SVM)[15], Boosting[14], Metaphase[66], and etc. In addition to, these features will be used by classifier to detect emotions different emotion (Woźniak, M et al., 2019).

Recently, deep learning techniques were developed to overcome the manual features extraction in machine learning.

In this project, the optimal model was developed based on deep learning techniques like convolution neural network (CNN) and long short term memory (LSTM) to detect students' engagement during online learning. Using CNN in this study due to its efficiency in abstract special features from raw data, and LSTM ability for time series.

1.2 Problem Statement

To date, there is a lack in an affective model that has an ability to extract and detect a complex emotion like engagement in real time during interacting with computer or using e-learning tool. Most studies focused on engagement as a cognitive state (engaged or disengaged) and ignore the relative elements (bored, interest, focused, happy, and sad) that effect engagement as an affective state.

1.3 Aims and Objectives of Project

This work aimed to develop an affective model that infer engagement affective states for students in e-learning environment or during students' interaction with computer. The model will become assistance tool for teachers to control students during learning process. It will track students' facial expression during learning process dynamically in real time without any intervention and sensitive equipment.

The project objectives are:

- To understand and abstract emotions that are relevant to engagement affect during learning process.
- To investigate and review state-of-art techniques in modelling engagement in learning context.
- To collect related natural affective data in computer- based-learning platforms.
- To prepare and analyze data to understand the nature of these data and how can be used for detecting purposes.
- To build an efficient model needs to training and testing different parameters.
- To evaluate a model performance toward unseen data and compare it efficiency with other previous models.

1.4 Related Works

In the educational domain, generally different studies were developed to track students' emotions which are occurring during the learning process. Verity of studies and works were developed to infer students' emotions during learning process. For example, researchers have tracked emotions that occur during interactions with Auto Tutor that is used to teach students physics, mathematics, and science .The set of learning-centered emotions identified from interactions with Auto Tutor; are flow, frustration, boredom, and, confusion, in addition to little frequent of surprise and delight studies considered affective states as appraisal situation. Advanced technology like Auto-Tutor is an intelligent tutoring system (ITS), and educational games were used by the researchers to track students' affective states. Different sources of information were used to predict students' emotions in the computer-based learning environment like facial expressions, posture, gestures, dialog conversation, and physiological data. Another path of research related with using sensitive instruments (i.e. sensors) , which tracked physiological changes during interacting with ITS. They used sensors like a posture analysis chair, pressure mouse, skin conductance, and facial expressions to collect the behavioral data.

Machine learning is an important technique in computer vision and its applications,due to the importance CNN techniques in abstract features map from raw data, and importance LSTM in extracting temporal features. This work focused on both CNN and LSTM to abstract and detect engagement from students in natural environment without using any specific equipment.

To detect engagement,it should be known specific issues related with it, like learning environment (e-learning or traditional), the methods or channels that used to capture engagement like (facial expressions, posture, gesture, head movements, and eye gaze), and the tools that should use to gain information from those channels like (physiological instruments or non- physiological). Below, are some previous works for engagement detection based on mentioned above:

Shen et al. (2022) depended on facial expression to detect students' engagement. They used CNN to asses' students' engagement during MOOC session. The researchers based on static images to build a model.

M. Murshed et al.,(2019) was conducted to classify three types of engagement (not-engaged, normally engaged, and highly-engaged). The model achieved different experiments by using different structure of CNN. The model accuracy for disengagement, normal, and high engagement were (91.74%, 89.55%, and 95.69%) respectively. They used DAiSEE dataset.

Turabzadeh et al.,(2018) , Local binary point(LBP) was used to extract facial features which are then fed as input to K-nearest neighbor to detect engagement. The model accuracy was reached 51.28%.

Kamath et al.,(2016) used histogram of oriented gradients(HOG) to extract the features of facial images,and Support-Vector Machine (SVM) for classification engagement. The model accuracy was 50.77% .

Hernandez, J., et al.,(2013) , Engagement was detection on TV viewer program by them. They used Support Vector machine(SVM) to detect engagement from facial expressions for people while they watched TV program. The best model accuracy was 77.92%.

Bidwell et al.,(2011) used eye gaze to detect student engagement. The features were extracted by expert human observations, then they used Hidden Markov Model (HMM) for training purposes .

1.5 Project Organization

The remainder of the project is structured into four chapters, the contents of which are listed below :

- **Chapter Two: (Theoretical Background)**
  This chapter presents a theoretical background of facial expression.
- **Chapter Three: (Proposed System Design and Implementation)**
  This chapter presents a proposed method for design and Implementation.
- **Chapter Four: (Results and Discussion)**
  This chapter presents the description of the different experiments and discusses the results and evaluations obtained from the implementation of the proposed system.
- **Chapter Five: " Conclusions and Future Works"**, The conclusions and future works will be presented in this chapter.

Chapter Two

Theoretical Background

Chapter Two

Theoretical Background

2. Introduction

Artificial intelligence (AI) is an important field in the computer science, it is a term coined by John McCarthy (Amit.K ,2018) .It can be considered as a machine that has an ability to think and take a decision intelligently. AI as a field can deal with other different fields like computing, mathematics, psychology, cognitive science, linguistic, philosophy, and engineering (Amit.K ,2018) .AI has significant effects on different applications such as image classification, object detection, intelligent games, and search engine. Machine learning (ML) is on of AI field which have large attention in the forecasting, prediction, and classification techniques.

2.1 Machine Learning

AI tries to make machines have abilities to learn, predict, and behave intelligently like human. These abilities have collected in a term called machine learning.

ML is a sub-area in AI and that have computational methods to give machines human-like learning ability. Machines ability to forecast, predict came from previous experience of knowledge(data). ML aims to develop algorithms that allow a computer to perform tasks without user intervention. The learning in ML focused on solving some problems like medical diagnoses, languages processing, weather forecasting and so on. Building artificial block of neural graphs in ML is called neural network(NN). These neurons can form groups of layers through weights and functions connections. ANN definition and its structure can be find in next section.

2.1.1 Artificial Neural Network

There are different definitions of ANN, the simplest one is a type of computational algorithms that formed in graphical way to solve special problems (Patterson & Gibson, 2017). Another definition consider ANN as a net consist of a connection of artificial neurons that mimic biological neural systems in human. While, defined ANN based on its structure, as a network consists of three layers; input layer, hidden layer, and output layer. These layers partially or fully connected layers. Below the figure (2.1) shows the general ANN.

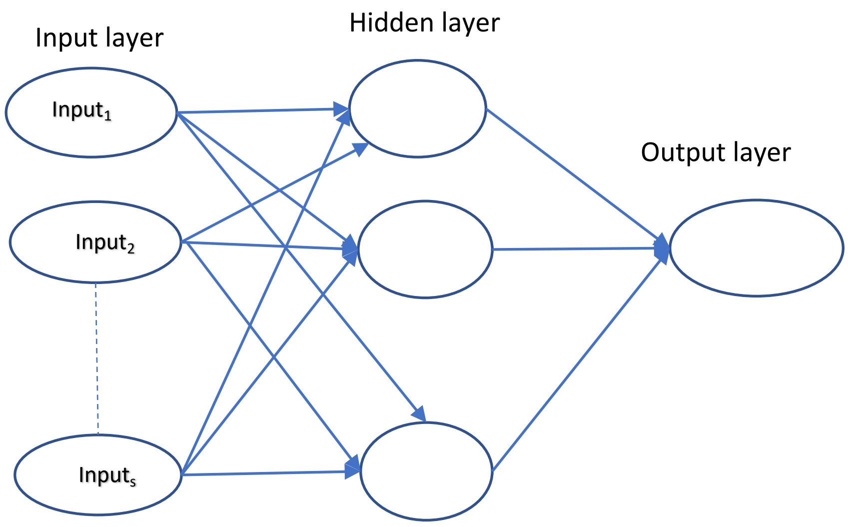

Figure (2.1): Typical diagram of ANN

Generally, any neural networks can consist of

1. Input layer: as is the first layer which feed the inputs to other layers and connected with the first of hidden layer.
2. Hidden layer(s): is one or more than one hidden layer based on the type of using algorithm. Is a bridge to connect input with output

layer. This connection passes the weights between layers to gain optimal solution.

3. Output layer: The last layer in the model which produces the final production which either prediction, or classification, or detection. This layer used activation function like Sigmoid or Soft Max. Mathematically can be expressed the activation functions work with other neurons of hidden layers as:

$$Wi = [w_i^1, w_i^2, \ldots, w_i^n] \tag{2.1}$$

$$X = [x_1, x_2, \ldots, x_n] \tag{2.2}$$

Where W is the vector of weights of neuron i, n is the number of features, and X is the input vector.

The value of output to the neuron y_i as in the equation below:

$$y_i = \sum_{t=1}^{n} w_t x_t + b \tag{2.3}$$

Where b is a bias.

After applying the activation function, then the output will be:

$$O_i = f(y_i) \tag{2.4}$$

O is then feed to the next layer as an input value to processed with their bias and own weights.

The structure will be used to learn machines with unknown data to abstract desired knowledge. The last of ten years ANN approved it efficiency in solving complex problems. Nowadays, there is growing attention towards another subnetwork of NN called Deep Learning (DL).

2.1.2 Deep Neural Network (DNN)

Previously, AI discussed as a main field that mimic or simulate human ability and experiences. To build a model in AI needs to abstract different representation of data, these are called by features extraction. This process can be done by machine learning. As can be define features extraction is a process to extract salient data and find the best meaning behind these data for specific task. Construction a model in machine learning, depend on algorithm choosing, as there are different algorithms available. Such as Support Vector Machine(SVM), Hidden Markov Model(HMM), K-nearest neighbors' models, Decision Trees (DT), and Dynamic Bayesian Models (DBM). The input features should be related to the task predictor and model purpose. One method used to generate features is to manually (or semi-automated) select the features that will be used to represent the model dimensionality. This method may be suitable for a model with low dimensionality of data (number of features). Using this method (hand-crafted features) in building models with high number of features and from different domains can be considered unreasonable. For example, in study achieved by Dawood et al., (2017) to extract different head motion required to use computational geometric knowledge and dimensionality reduction to build feature vectors. After that, these vectors will be the input of HMM. It can easily extract basic features in machine learning but it is difficult to extract high features and micro details from raw data (Wiley, 2016; LeCun et al., 2015). Therefore, deep learning widely spread recently to extract high dimensionality of features.

Deep learning is a neural network with deep structure, can depict its deep structure that consists of more than two hidden layers. This depth given DL dynamicity to extract a high level of features from raw data verses other shallow neural networks (Wiley, 2016). Complex connectivity of help layers to abstract more details of features. DNN can reduce dimensionality and handle nonlinear complex representation from basic of raw data against other methods. Based on these characteristics DNN becomes state-of-the-art machines to solve complex problems like classification, prediction, detection, and recognition.

2.1.3 Components of DNN

The common and basic components of any deep learning model include: -

1- Layers
1. Parameters
2. Activation functions
3. Loss functions
4. Optimization methods
5. Hyper parameters.

1. Layers: are the basic units in DNN, they are different in their numbers, connections, hyper parameters and usage based on DL structure and type.
2. Parameters: are input vector to the network which consists of number of features along with their associated weights.
3. Activation Function: means the function $f(w_x)$ is called activation function. It receives input data and processes it to decide the output, which is fired into other neurons.

Generally, there are different types of activation such as:

1. Linear function
2. Rectified Linear Function
3. Sigmoid Function
4. Tanh Function

5. 5oftMax Function

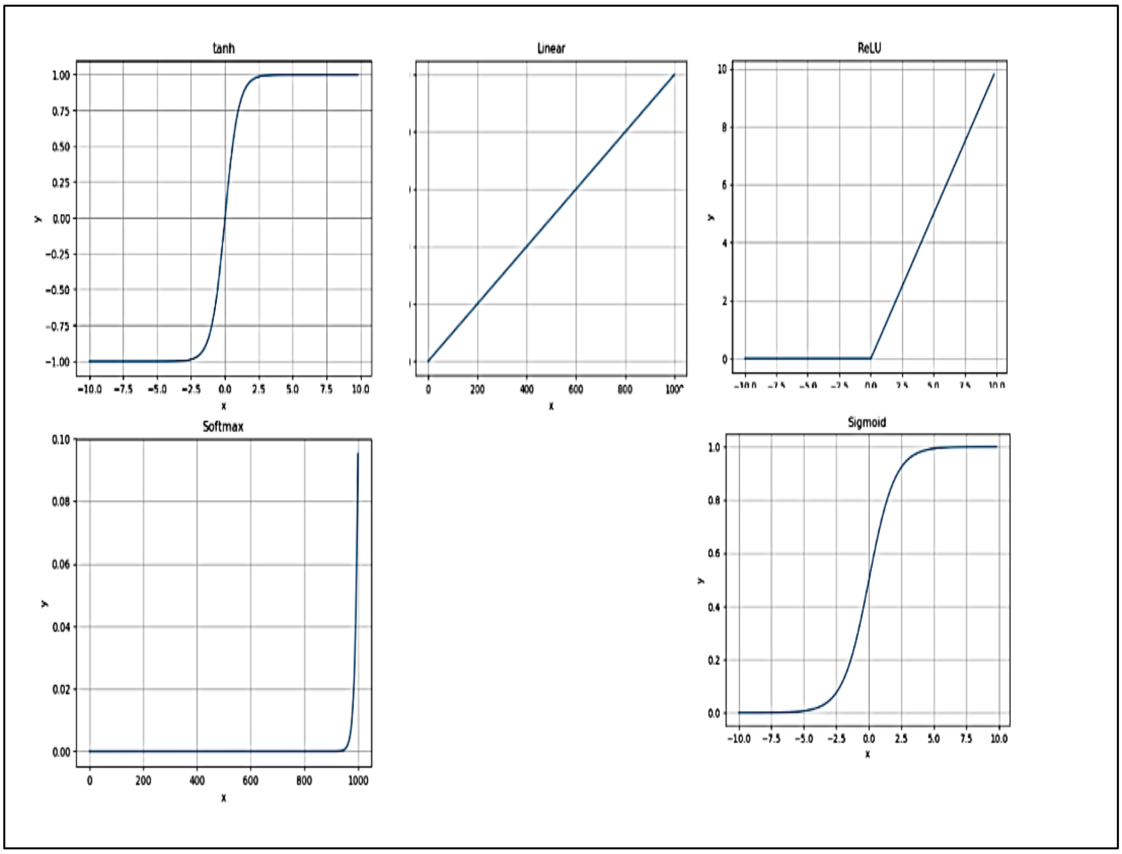

Figure (2.2): Popular Activation Function in DL

4. Loss Function: is a metric to determine the error in producing true predictions between the class target label and the predicted class label. In other words, it is the amount of misclassification error.

5. Optimization Methods: model training in machine learning aims to produce model logic which can predict the best features from a parameter vector. Optimization process finds the best model parameters (weights) with lowest loss value and higher prediction accuracy.

6. Hyper parameter: is a configuration of parameters settings. Choosing hyper parameters values is based on their effect in model performance. The general hyper parameters in any deep learning model are:

- Layer size.

- Magnitude like step size, gradient, momentum and learning rate.

- Regularization like Dropout and batch-normalization.

- Type of activation functions. - Number of epochs during the training phase and mini-batch size for each step.

- Loss function.

- Vectorization.

- Weights initialization.

The above section showed the popular components of DL, the next sections will state common architectures of deep networks used in this thesis.

2.1.4 Convolutional Neural Network (CNN)

CNN (LeCun et al., 2010) is a type of NN to represent the data with a grid structure. CNN has an ability to deal with different dimensions of data. For example, on-dimension(1D) grid for data like time series, 2D grid for image pixels with width and height, and videos as 3D. Any CNN structure has input layer, number of hidden layers, and output layer. The hidden layer can be formed in CNN from convolutional layers, pooling layers, flatten layers, dropout layers, and fully connected layers.

Convolutional layer of CNN refers to the mathematical operations of the merging two function, it represents the core of CNN. This operation works as a filter for the input information. Figure (2.3) below stated the convolutional layer concepts.

Pooling layer: in another important layer in CNN, it works to reduce features map dimensionality during convolutional process. That will lead to reduce number of training parameters. Pooling has an important role on time of training and overcome the overfitting of network. As well as, features reduction improve model performance towards unseen data (LeCun et al., 2010).

There are two types of pooling layer, first one based on choosing the maximum value of features by sliding a window with specific side and stride (skip step) to reduce the features map, this type called by Max Pooling. Second type, is Average Pooling which try to calculate mean value for the scanned window. The purpose of the two pooling is in addition to reduction features map, also can return the important value. As stated that CNN is a best structure for spatial features engineering, spatial features can be extracted from CNN while temporal features can be abstract by using another type of Recurrent neural network (RNN) is called Long Short Term Memory(LSTM).

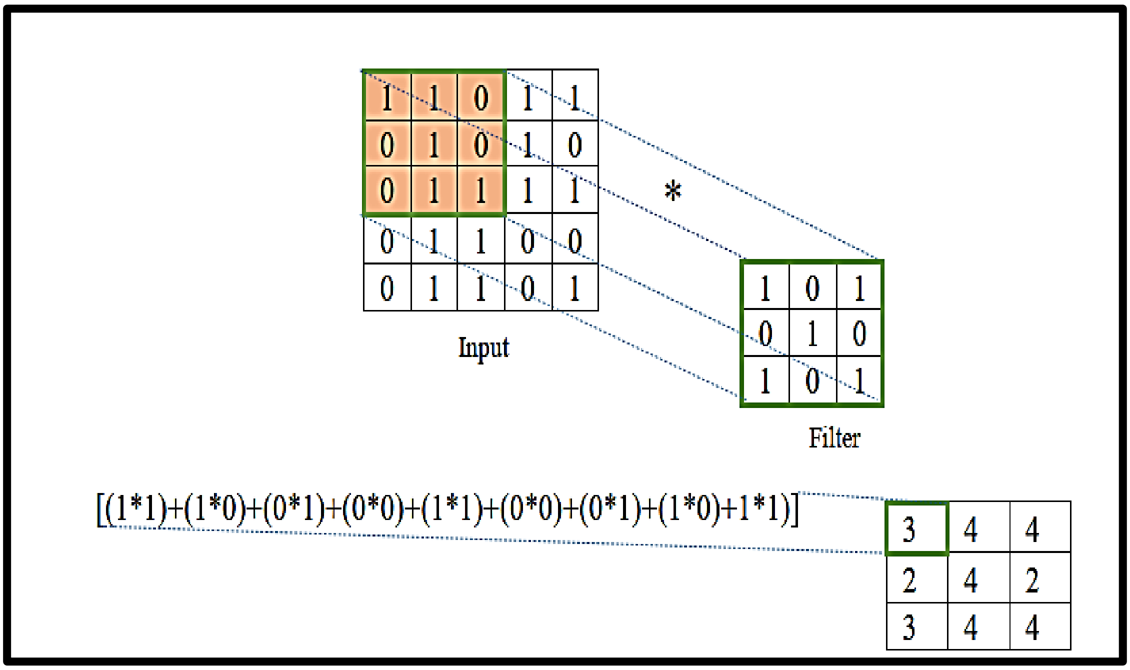

Figure (2.3): Convolutional Operation in CNN to extract features map.

2.1.5 Long Short Term Memory (LSTM)

The limitations of RNN that recognized by it is inability to store information for long time. To overcome these limitations LSTM was introduced by (Hochreiter & Schmidhuber, 2014) as an efficient machine to deal with long length input sequences for long time. LSTM structure consists of input, memory celled, and output gates, the input and output function to prevent unrelated data to pass to the memory. Forbidden concept added for learning process, it is meaning deleted far history of data by using new gate called by Forbidden gate. For this power of LSTM used widely in deep learning applications ((Fernández, Graves, & Schmidhuber, 2012; He & Droppo, 2016; Hsu, Zhang, Lee, & Glass, 2016).

Tanh activation function was used for input and output units, while gate activation is Sigmoid function.

Emotion recognition

Any algorithm in deep or machine learning, which are used for classification or emotion recognition should follow three steps:

A. **Face detection in space**
B. **Features extraction**
C. **Emotions classification**

A. Face detection

Face detection is a process used to find and extract face region from background in video clips or still image. This step is important for any automated model in facial expressions recognition and classification. The purpose of this step is to estimate new position of face in frames sequences or images.

B. Feature extraction

Is the next step in facial and emotions modelling. This process tries to abstract important data to determine the person image, the goal is to reduce the information size and extract the important meaning behind these information or data.

The best methods to extract facial features points depended on a processing face model like a whole face (holistic face) or part of face or specific regions of face (local face), Also, the methods depended on type of data like using still images or video stream. Based on these methods can be recognized two types of feature extraction methods. First one, is appearance based methods, which are depended on physical changes to extract and determine features such as wrinkles, eye color, skin color, and furrowed. This method tracked face locally and find suitable appearance. (Tian et al., 2013; Happy et al., 2012). Examples of these methods are Gabor Features (Viola et al., 2020; Tian et al., 2016; Yang et al., 2018; Donato et al., 2012); Haar-like features (Viola & Jones, 2020; Yang et al., 2010), and Local binary pattern(LBP) (Ojala et al., 2017; Zhao & Pietikainen, 2017; Ojala et al., 2012).

Other methods is called by geometrical features which are track whole face and depend on shape and location of facial points on face parts like eyes, mouth, chin, eyebrows, and top of nose (Tian et al., 2016).

Choosing the proper method based on algorithm which are used, the purpose of extracting these features, and type of data.

C. Emotions classification

It is the last step in emotion recognition and classification. This step depended on the features map that extracted by previous steps. These methods can be divided based on type of algorithms techniques that be used in classification. These types are traditional approach, probabilistic approach, deep learning approach.

Traditional approach like if-then-rule, k-means, fuzzy mean, and active appearance model (AAM) (Pantic et al., 2018; Jurado et al., 2008). These approaches used questionnaires form and manipulate them statically. The limitations of these methods done manually and don't provide best response or correct affect.

While, probabilistic methods were developed to overcome the traditional methods limitations, like HMM, and DBN (Baker et al., 2010; Afzal & Robinson, 2010; Conati, 2019). These methods need a prior knowledge for features, and use handcrafted feature expressions to classify emotions.

From above, it is clear that features extractions and emotions classification consider is a difficult task. As the difficulty was identified by selecting the suitable algorithm for features extraction, then feed to specific classifier for emotions classification. Deep learning techniques developed like a popular convolution neural network (LeCun et al., 2017), and RNN-LSTM to overcome the limitations of above methods .

As this paper depended on CNN and LSTM to build emotional model that can extract engagement and its level.

2.2 Dataset

In order to build emotional model need to emotional dataset for this purpose. Building affective and emotional dataset consider is difficult mission. As it requires natural environment without any controlling or user intervention, also needs a lot of efforts. Generally, most researchers agreed that the universal emotions in the world are six basic emotions are (happy, sad, disgust, anger, fear, and surprise) with face in plane . almost available dataset is acted data. Acted data means the emotions were collected and annotated based on acted movements of emotions aby professional actors.

Increasing in the applications that employ real time space, there is growing need to natural data in real environment and time. Therefore, the efforts were directed towards natural-spontaneous data. Due that, the researchers approved that the natural data are different in appearance and intensity from acted data (Cohn & Schmidt, 2018; Schmidt et al., 2013). Moreover, acted data can't be used for applications in real life. Typically, the datasets in the emotional field classified into three categories: acted data, induced data, and natural-spontaneous data.

- Acted data: is a data that collected and annotated based on professional actors, as they are acted basic emotions based on specific instructions. It is easy to collect and annotate, and there is no need to ethical approval. However, it is not willing for applications in real time (Pelachaud, 2013).
- Induced data: the emotions in this type were collected in controlled environment based on induced specific behaviors during a session such as watching a movie, or hearing a story (Gross & Levenson, 2016).
- Natural-spontaneous data: this type of data were collected in natural and an uncontrolled environment. As the participants in this

data express emotions naturally without any instruction. Such type of these data in Human-computer interaction (HCI), data for call center, or videos from webpage or TV (Lee & Narayanan, 2021; Devillers & Vidrascu, 2018; Busso et al., 2013; Morency et al., 2011).

- These datasets recognize by their collecting and annotating difficulty, and they need to ethical consideration.

Generally, each type of these dataset has advantages and disadvantages, and choosing a proper type depends on the context of problem and where the data will be used. In this study, DAiSEE dataset were used for model training and testing purposes (Gupta, et al., 2016). The dataset was created in online learning session for students in uncontrolled environment. It contains on specific types of emotions which related with engagement effects like bored, happy, sad, anger, joy, fear, and disgust. Engagement affect were labeled under very high level, high, and low level. To evaluate the proposed model, other videos were used from another dataset to approve model generality toward unseen data that different in environment and other conditions like illumination, occlusion, and control**Invalid source specified.**.

2.4 Performance Measurements

The statically measurements for model efficiency based on confusion matrix. The raw in confusion matrix describes the class prediction by the model, while the column presents the actual class or label in the ground truth dataset. In addition, other metrics can be used to measure model generality. These metrics, True Positive Rate (TP) also called by (Recall) which is mean the number of correct detection that match the true label in dataset. As well as, False Negative Rate (FN) which detected the false class. Another metric used was Precision or Positive Prediction Value (PPV). The mathematical meaning of these

measurements is illustrated Equations 2.5, and 2.6 (Davis, 2020).

$$Recall = \frac{\mathrm{TP}}{(\mathrm{TP+FN})} \tag{2.5}$$

$$Precision = \frac{\mathrm{TP}}{(\mathrm{TP+FP})} \tag{2.6}$$

$$\text{Accuracy} = \frac{\text{number of detection}}{\text{total number of videos}} \tag{2.7}$$

TP= is the number of true detection

FN= is the number of false negative detection of class that doesn’t match the true label.

FP = is the number of false class that match the label in ground truth data.

Chapter Three

Model Building

Chapter Three

Model Building

3.1 Introduction

This chapter presents the significant methodologies which are used to build a model for engagement detection. First, the description of the dataset that used in this work. Then describes the CNN algorithms for model construction. As well as, description of the model training, testing, and validation.

3.2 Model Building

The structure of the proposed model can be seen in figure (3.1) below. This structure consists of CNN-LSTM-CNN.

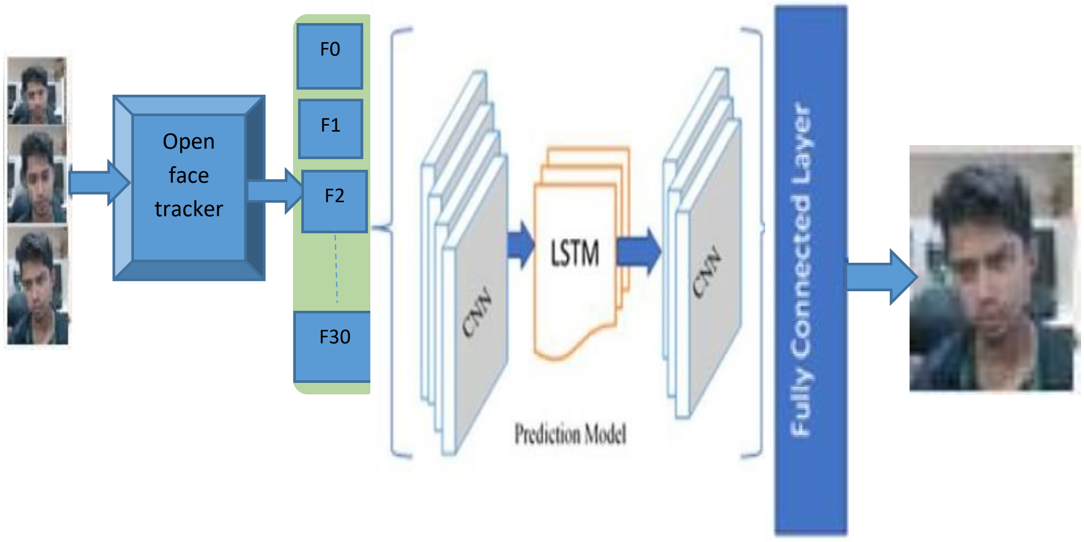

Figure (3.1): Structure of Proposed Model

3.3 Dataset Preparation

In E-learning environment, there is a lack in the dataset that concerned students during their online learning or during their interaction with computer. In this work, the dataset was used DAiSEE[16] as it is public and available datasets for scientific research purposes. DAiSEE contains 9,068 videos captured from 112 students under the online courses environment. The video was conducted while learning online courses, engagement affect was the mainly target that be taken. The video duration is 30 frame per second(30f/s) with resolution 1920 × 1080 pixels. This dataset contains different emotions like confused, bored, engaged, frustrated. All these emotions have the same features and expressions. Therefore, the proposed model goal is to recognise and classify engagement affect from others emotions.

The first step in any classification and detection model is prepare the dataset.

The form of the dataset was a set of labelled video. Face features were extracted from video clips by an Open Face tracker (Baltruvsaitis, et al., 2016). It is an open sours tracker that used to track and detect, head, and eye gaze in space in real time. The concern in this work focused on extracting facial action units (AU), and eye gaze direction. The features resulted as output from tracker for each video were filtered to get the desired features and saved in a separate file. The file contains ID subject, ID emotion, and emotion label.

Before using the dataset for training purposes some steps of processing were accomplished like data-cleaning, and outliers' detection. After the preprocessing dataset, the sequence of features will be used to build model cues.

3.4 Data-Cleaning

Involves finding the error and missing frames, and deleting the error. Error tracking happened due to the variation and occlusion in dataset environment. Like illumination changes, or miss head tracker (out of camera range) and missing in face capture in some frames. Based on open face tracker there is an average ratio (0.85) that determined the missing tracker. If the tracker ratio less than 0.85 that mean, there is error or missing tracker. The counting of values of missing frames tracking in this work was 2.56% from all dataset, so due to this insignificant value the frames were removed from file. Algorithm (3.1)stated the procedure of detecting missing tracking.

Algorithm (3.1) missing-tracking

1- For each file tracking-video in dataset

  1-track each frame in the video

    1- If tracking rate <0.85 then

      1-pointer frame for eliminating

  2-Remove pointer frame from file tracking

3.5 Missing Labeled

Another step in data preprocessing is tracking missing labeled or unlabeled data. Missing label means there is unlabeled value for video. Unlabeled means there is no emotions-id in the video file tracking. To resolve this problem, it done by tracking or checking all missing values and re-label the missing. As can be shown in Algorithm (3.2) for missing label.

Algorithm (3.2) missing-labeled

1- For each file tracking-video in dataset, track missing label
    1- S= set of unlabeled video[]
    2- For each file-video
        1- If the emotion-ID is missing then filled value
            1- Append. Unlabeled video [video-clip]
    3-Re-named all video –clip in set

3.6 Predictive Model

To extract emotions with high level of features abstraction require optimal classifier. Today, deep learning approved its solidity in image classification and object detection. The deep learning techniques which are used in this thesis is Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). The idea of choosing these techniques extracted from stacked layers for CNN which are used by a lot of researchers in the image net award (Russakovsky et al., 2015; He et al., 2016).

Using CNN to extract spatial features of video clips from the temporal series which include AU, and eye gaze. Features space is cue vector where $\boldsymbol{cue} \in R^{C*(AUs+eye\ gaze)}$, c =30 is the number of frames. AUs=18 is the number of action units in this thesis, and gaze = 4 is (up, dawn, left, right) for each eye.

The model build based on using CNN and LST techniques. The power of this model depended on CNN strong in dealing with spatial data and LSTM with time series. Convolutional layer in the proposed structure has filter of size (3x1).

The resulting map from last layer will then enter into a classification layer with (softmax) function to get the correct class.

The model parameters are listed below in table (3.1) for CNN model and LSTM

Table (3.1): CNN model and LSTM

LAYER	**SHAPE**	**PARAMS.**
Input layer	(6,27)	0
Conv 1D	(6,64)	1792
Dropout	(6,64)	0
Fully connected	(6,27)	1755
Sum		3547
LSTM		
LAYER	SHAPE	PARAMS.
LSTM	(6,162)	123120
LSTM	(6,162)	210600
Sum		333720
CNN		
Conv 1D	(6,128)	20864
Conv 1D	(6,128)	16512
Conv 1D	(6,128)	16512
Conv 1D	(6,128)	16512
Pooling 1D	128	0
Sum		70400

The model reads the videos sequentially by reading the data frame by frame. The features that taken from the tracker will be sampled and transform into stacked of frames (30f/s), then divide into five frames as each one action equals five frame. This result was included after different experiments to choose the best. The matrix of reading will transform and stacked from 1D to 2D.

The model consists of three subnetworks, first one is CNN with 256 node and T*anh*-activation function, follow by dropout layer with ratio of 20%. Then, stacked of two LSTM will use to predict long-sequence dependences. which are a particular kind of recurrent neural network (Hochreiter & Schmidhuber, 2015). The first LSTM take the 188 output from CNN. The layers of two LSTM have 972 nod for each one. The second LSTM gain the sequence of features map from first one and produce the final series of features which fed into another CNN. The purpose to these step to abstract a lot of skills of features. The recently CNN formed from four convolutions with one dimension (Conv 1D) and Tanh-activation function. As well a pooling layer to reduce features dimensionality. The final step in the proposed model is the fully connected layer which gains its input from the previous subnets to predict the emotions. This layer has five units and softmax function to extract emotions which refer to the engagement. The loss function was categorical cross-entropy and Adam optimizer.

3.7 Model Training

The previous sections were prepared dataset and stated the structure of model to extract features. Now, the data should be divided into three groups (training, validating, and testing data). Training data means the data that can be seen by the model to learn and adjust its weights. The validation data is the part that used by the model to provide un biased evaluation for the model parameters during learning step. As during validation, the model will not learn or train from the validation data, but based on the validation results the model will update its parameters. Finally, the test data uses to test the model efficiency toward unseen data after the model complete its all training epochs.

There are no standardised methods to choose optimal size of these three parts, but it depended on trial and error. For this reason, different experiments were achieved to choose the best size. The dataset split into (60% training, 40 validation), (70% training, 30 validation) , and (80% training, 20 validation). The model was trained and tested using 10-fold cross validation with 100 epochs.

Table (3.2) : Dataset splitting ratio effect on Accuracy and Loss

Split Ratio	*Accuracy*		*Error Rate*	
Training- Val.	***Mean***	***Max***	***Mean***	***Min***
60%-40%	62.34%	62.97%	0.98	0.97
70%-30%	62.53%	62.85%	0.98	0.97
80%-20%	**62.63%**	**63.12**	**0.97**	**0.96**

The experiments show that the best performance of split dataset into (80-20%). Therefore, the ratio 80% and 20% were based in this model for training and validation respectively. The ratio of 5% of total dataset were excluded for testing purposes, then the remaining data were siplite into training and validation as shown in table (3.3) below.

Table (3.3): Remaining Data were split into Training and Validation

TOTAL NUMBER OF CUES	TESTING 5% CUES	REMAINING CUES	TRAINING CUES (80%)	VALIDATION CUES (20%)
159732	7987	151745	121396	30349

Table (3.4) below describes the hyper-parameters which used in this model.

Table (3.4): Model Final Hyper-Parameters

Hyper-parameter	***Value***
Dataset parts ratio	5% testing, 80-20% training and validation
Mini-Batch size	64 cues/batches
Loss Function	Categorical Cross-entropy
Optimizer	Nadam
Learning Rate Size	0.001
No. of epochs	100

Based on the parameters which mentioned above and using 10-cross validation with 100 epochs for training and validation, the model achieved significant accuracy for model training and validation process. As shown in figure (3.2)

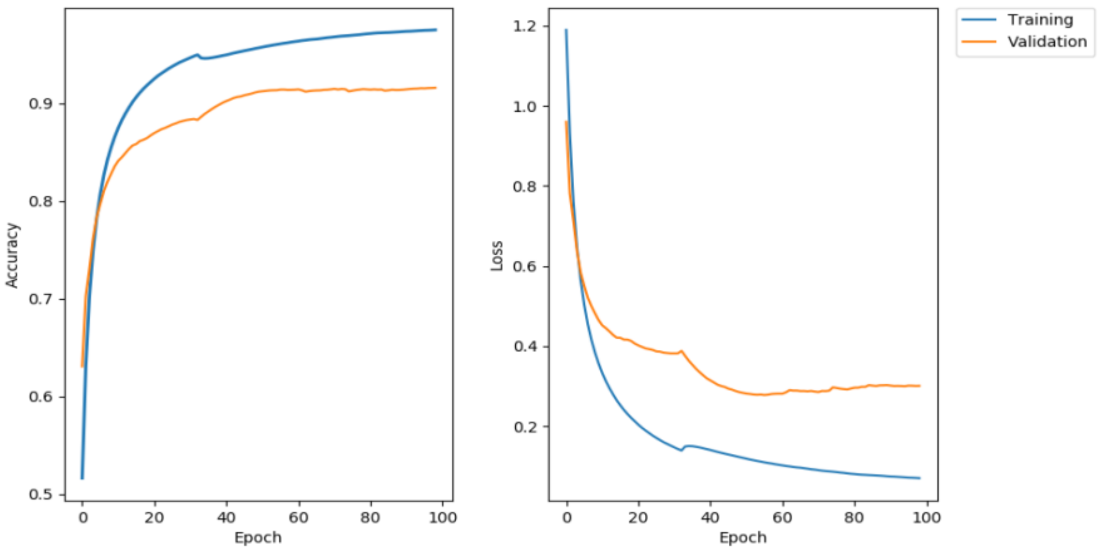

Figure (3.2) : Final model performance on training and validation data (accuracy and loss error)

Chapter Four

Results and Discussion

Chapter Four

Results and Discussion

4.1 Introduction

This chapter presents the significant results that obtained from proposed model that mentioned in previous chapter. It states the model accuracy and the confusion matrix for all model classes. In addition to, the chapter presents the model generality towed unseen data.

4.2 Results

From Chapter 4, the model was trained and validated using a DAiSee dataset. The several ratio from size of data was withheld for model testing, these data were not used during training and validity steps. The model performance and its efficiency can be measured through its accuracy, error rate, generality towards unseen data, and its speed. The results in table (4.1) and chart diagram in figure (4.1) show the model test accuracy.

Table 4.1: Accuracy for model classification

AFFECTIVE STATES	ACCURACY %
BORED	80
confus	81
ENGAGEMENT	96
frustration	88

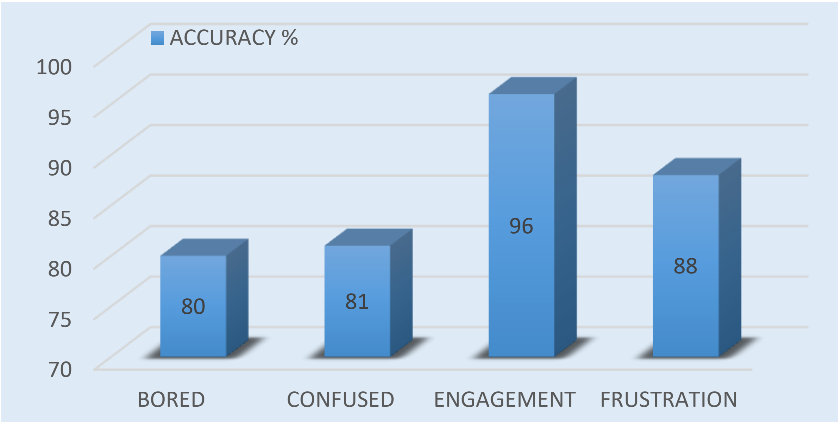

Figure (4.1): Chat Diagram for Model Accuracy.

From the above can be see that the highest accuracy is for engagement detection, while the lowest accuracy is bored. The variation in accuracy between the classes because that all videos nearly share the same features of engagement. As shown in the figure below (4.2) which states sample of the data from (DAiSEE) . As seen from figure below the different emotions share the same expressions and features, and it is difficulty to recognise between them. In spite that, the model approved its efficiency by classification and detection them. That clear and can be seen from the results in table (4.1).

Figure (4.2) Sample of Emotions from DAiSEE Dataset

Other metrics were used to measure model efficiency are shown in the table (4.2) below.

Table (4.2): Measure Model Efficiency

Metrics	Result
precision	95.33
Recall	88.72
F1 score	91.91

The accuracy measure represents the all detection classes over the total number of sample. On the other side, the metrics in table (4.2) were used for enhance model precisions.

4.3 Evaluation

To evaluate the model efficiency, another natural-spontaneous dataset was used for this goal. The data set (NACD) that mentioned in chapter 2 used to evaluate the model performance toward unseen data. the model accuracy was 88%. This rate of accuracy shows the significant model performance toward unseen data, as NACD dataset is characterized by its environment, lighting, occlusion, and cultural differences. NACD is not used in model training, validation, and testing for the proposed model. This evaluation approved the power of CNN and LSTM techniques on model performance and especially with a new environment.

4.4 Model Speed

Model speed is an important factor in any model performance especial in application like HCI. The model speed means the starting time with an input and ending at response or output of class. The proposed model was tested on a computer without GPU and with an intel ® Core™

i3-40104 CPU @1.70 GHZ processor, RAM 8 GB, and 64bit windows 8 operating system. The average loop time of model speed, which starts with the input video, tracking, extract features, and the output of affective prediction is 25.84 msec for each 30 frame/sec. The average time of affective prediction for 30 frame/sec equals 18 msec. There are no standardized methods to determine models' performance in real time, as that depends on the model's context . The time delay 45 msec is acceptable in application like HCI, and above this time will consider as a time delay .

4.5 Case Study

In this experiment the proposed system compared with another system to insure the system efficiency, using DAiSEE data set with the same conditions. The system compared with M. Murshed (2019) to classify the engagement (not engaged, normal engaged, high engaged)

The result of the proposed model achieved higher accuracy than Murshed system in different experiment and different structure of CNN

Table (4.3) The average classification accuracy in test dataset

Model	Accuracy	parameters
All-CNN	7.5	1.366 M
Proposed model	94.34	0.605 M

Table (4.4) The average classification accuracy in test dataset

Model	**not-engaged**	**Normally engaged**	**highly-engaged**
	(Level # 1)	**(Level # 2)**	**(Level # 3)**
CNN	75.75 (±0.05)	72.34 (±0.07)	79.83 (±0.03)
proposed Model	92.55 (±0.08)	90.15 (±0.05)	96.76 (±0.07)

we investigated three popular models of convolutional neural networks (CNN) and a proposed model for the students' engagement classification. In the experiments, three-level (not-engaged, normally-engaged and highly-engaged) decisions on engagement detection have been made, where the proposed model shows high accuracy in engagement classification (not-engaged 92.55%, normally-engaged 90.15%, and highly-engaged 96.76%) in the Dataset for the Affective States in E-Environments (DAiSEE, M. Murshed 2019).

Chapter Five

Conclusion and Future Works

Chapter Five
Conclusions and Future Works

5.1 Introduction

This chapter will conclude the progress that been made in building emotional model, which used to detect emotions engagement between different emotions in online learning environment. Also, it presents important future works.

5.2 Conclusions

This thesis developed an emotional model to extract students' engagement during interacting with computer. One of the objectives of this thesis that was achieved using natural data in uncontrolled environment that serve applications in real time.

CNN and LSTM techniques were used to automate spatial-temporal features of students based on their facial expressions. As based on investigating literature review, that shallow machine learning provides low generality toward unseen data. While, deep learning techniques overcomes these limitations. Moreover, general machine learning classifiers require prior knowledge about the features that related to the required class. In contrast, CNN and LSTM can extract their features from raw data and learn through their deep connected layers. This capability belongs to their complexity in layers' connection.

Best performance of model was accomplished through choosing the best hyper-parameters. Best accuracy for model based on the power of CNN and LSTM in extracting spatial-temporal features. Furthermore, model generality was evaluated towards unseen data, as these data are different completely in their characteristic from which one that used for model training.

The best results can be seen in chapter four as 80%, 96%, 88% accuracy for bored, engagement, and frustration respectively. The highest ratio for engagement as the dataset focus was on engagement affect.

5.3 Future Works

One of the important work can be applied merging the DAiSEE and NACD dataset in learning Give high power for the model performance. Also,

In future cultural differences can have impact on model performance due to the variations in facial expressions. Therefore, it is important to take culture differences in annotation and recognition affective states.

Another interesting work is combined another channel like body gestures with facial expression.

Finally, may be these works will foster for the rich modality of computer-based learning to develop further socially and emotionally intelligent user interface.

References

References

1. Sarwar, H., Akhtar, H., Naeem, M. M., Khan, J. A., Waraich, K., Shabbir, S., Hasan, A., & Khurshid, Z. (2020). Self-reported effectiveness of eLearning classes during COVID-19 Pandemic: A nation-wide survey of Pakistani undergraduate dentistry students. European Journal of Dentistry, 14(S01), S34–S43. https://doi.org/10.1055/s-0040-1717000
2. M. Murshed, M. A. A. Dewan, F. Lin and D. Wen, "Engagement Detection in e-Learning Environments using Convolutional Neural Networks," 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, 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, pp. 80-86, doi: 10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00028.
3. Poon, W. C., Kunchamboo, V., & Koay, K. Y. (2022). E-Learning Engagement and Effectiveness during the COVID-19 Pandemic: The Interaction Model. International Journal of Human–Computer Interaction, 1-15. https://doi.org/10.1080/10447318.2022.2119659.
4. Christenson, S., Reschly, A. L., & Wylie, C. (2012). Handbook of research on student engagement (Vol. 840). New York: Springer. https://doi.org/10.1007/978-1-4614-2018-7.
5. Altuwairqi, K., Jarraya, S. K., Allinjawi, A., & Hammami, M. (2021). A new emotion–based affective model to detect student's engagement. Journal of King Saud University-Computer and Information Sciences, 33(1), 99-109. https://doi.org/10.1016/j.jksuci.2018.12.008.
6. Y. Zeng and Y. . -C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," in IEEE

Transactions on Communications, vol. 57, no. 6, pp. 1784-1793, June 2009, doi: 10.1109/TCOMM.2009.06.070402.

7. Woźniak, M. (2019, May). Gamification of Eye Exercises for Evaluating Eye Fatigue. In Artificial Intelligence and Soft Computing: 18th International Conference, ICAISC 2019, Zakopane, Poland, June 16–20, 2019, Proceedings, Part II (Vol. 11509, p. 104). Springer.
8. Shen, L., Dillard, J. P., & Peng, L. (2022). Correspondence between Two Methods of Measuring Discrete Emotions: Self-report versus Machine-coded Facial Displays. Western Journal of Communication, 86(2), 194-214. https://doi.org/10.1080/10570314.2022.2040582 .
9. 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. https://doi.org/10.3390/technologies6010017.
10. 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. DOI: 10.1109/WACV.2016.7477618.
11. Hernandez, J., Liu, Z., Hulten, G., DeBarr, D., Krum, K., & Zhang, Z. (2013, April). Measuring the engagement level of TV viewers. In 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG) (pp. 1-7). IEEE. DOI: 10.1109/FG.2013.6553742
12. Bidwell, J., & Fuchs, H. (2011). Classroom analytics: Measuring student engagement with automated gaze tracking. Behav Res Methods, 49(113).
13. Amit, K. (2018). Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain. CRC press.
14. Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: going

beyond euclidean data. IEEE Signal Processing Magazine, 34(4), 18-42.

15. Koutnik, J., Greff, K., Gomez, F., & Schmidhuber, J. (2014, June). A clockwork rnn. In International conference on machine learning (pp. 1863-1871). PMLR.

16. Graves, A., & Graves, A. (2012). Long short-term memory. Supervised sequence labelling with recurrent neural networks, 37-45.

17. He, T., & Droppo, J. (2016, March). Exploiting LSTM structure in deep neural networks for speech recognition. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5445-5449). IEEE. doi: 10.1109/ICASSP.2016.7472718.

18. Ping Tian, D. (2013). A review on image feature extraction and representation techniques. International Journal of Multimedia and Ubiquitous Engineering, 8(4), 385-396.

19. Happy, S. L., George, A., & Routray, A. (2012, December). A real time facial expression classification system using local binary patterns. In 2012 4th International conference on intelligent human computer interaction (IHCI) (pp. 1-5). IEEE. doi: 10.1109/IHCI.2012.6481802.

20. Viola, R., Martin, A., Morgade, J., Masneri, S., Zorrilla, M., Angueira, P., & Montalbán, J. (2020). Predictive CDN selection for video delivery based on LSTM network performance forecasts and cost-effective trade-offs. IEEE Transactions on Broadcasting, 67(1), 145-158.

21. Yang, Y., Zhou, J., Ai, J., Bin, Y., Hanjalic, A., Shen, H. T., & Ji, Y. (2018). Video captioning by adversarial LSTM. IEEE Transactions on Image Processing, 27(11), 5600-5611.

22. Caroni, P., Donato, F., & Muller, D. (2012). Structural plasticity upon learning: regulation and functions. Nature Reviews Neuroscience, 13(7), 478-490.

23. Viola, R., Martin, A., Morgade, J., Masneri, S., Zorrilla, M., Angueira, P., & Montalbán, J. (2020). Predictive CDN selection for video delivery based on LSTM network performance forecasts and cost-effective trade-offs. IEEE Transactions on Broadcasting, 67(1), 145-158.

24. Alatalo, T., Pouke, M., Koskela, T., Hurskainen, T., Florea, C., & Ojala, T. (2017, June). Two real-world case studies on 3D web applications for participatory urban planning. In Proceedings of the 22nd International Conference on 3D Web Technology (pp. 1-9).

25. Saitoh, T., Zhou, Z., Zhao, G., & Pietikäinen, M. (2017). Concatenated frame image based CNN for visual speech recognition. In Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part II 13 (pp. 277-289). Springer International Publishing.

26. Lim, Y. K., Liao, Z., Petridis, S., & Pantic, M. (2018). Transfer learning for action unit recognition. arXiv preprint arXiv:1807.07556.

27. Afzal, S., & Robinson, P. (2010, July). Modelling affect in learning environments-motivation and methods. In 2010 10th IEEE International Conference on Advanced Learning Technologies(pp.438-442).IEEE.doi: 10.1109/ICALT.2010.127.

28. Sims, S. D., Putnam, V., & Conati, C. (2019). Predicting confusion from eye-tracking data with recurrent neural networks. arXiv preprint arXiv:1906.11211.

29. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018, October). Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM international conference on multimodal interaction (pp. 400-408).

30. Perazzi, F., Pont-Tuset, J., McWilliams, B., Van Gool, L., Gross, M., & Sorkine-Hornung, A. (2016). A benchmark dataset and evaluation methodology for video object segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 724-732).

31. Lim, Y., Toutios, A., Bliesener, Y., Tian, Y., Lingala, S. G., Vaz, C., ... & Narayanan, S. S. (2021). A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images. Scientific data, 8(1), 187.

32. Etienne, C., Fidanza, G., Petrovskii, A., Devillers, L., & Schmauch, B. (2018). Cnn+ lstm architecture for speech emotion recognition with data augmentation. arXiv preprint arXiv:1802.05630.

33. Morency, L. P., Mihalcea, R., & Doshi, P. (2011, November). Towards multimodal sentiment analysis: Harvesting opinions from the web. In Proceedings of the 13th international conference on multimodal interfaces (pp. 169-176). https://doi.org/10.1145/2070481.2070509

34. Hurt, J. A., Huangal, D., Davis, C. H., & Scott, G. J. (2020, December). enabling machine-assisted visual analytics for high-resolution remote sensing imagery with enhanced benchmark meta-dataset training of NAS neural networks. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 3914-3923). IEEE.

الخلاصة

مع التطور المتزايد في تطبيقات تفاعل الإنسان مع الحاسوب ، وخاصة في مجال التعليم. حاول العلماء تطوير واجهة ذكية تكون أكثر تفاعلية واستجابة للمستخدمين.

احتل التعليم الالكتروني دورا مهما في الكثير من الازمات وخصوصا أثناء جائحة كورونا و فرض قيود التباعد الاجتماعي، تم الاستعاضة عن طرق التعليم التقليدية (التي تكون وجها لوجه) بطرق التعلم عن بعد. لهذا السبب تولدت الحاجه الى نظام او موديل يمتلك القدرة على اكتشاف تصرفات الطلاب خلال التعليم الالكتروني والاستجابة لها تلقائيًا من تعابير الوجه وبصوره اتوماتيكية دون مقاطعة عملية التعليم. لذلك ومن خلال هذا المشروع تم اخذ هذه المتطلبات بنظر الاعتبار لتصميم وتطوير نموذج مؤثرًا لاستخلاص تفاعلات الطلاب اثناء التعليم الالكتروني باستخدام الحاسوب.

تم اعتماد تقنيات الذكاء الصناعي (التعلم العميق) نظرًا لقوة هذه الخوارزميات في استخلاص مستوى عالٍ من الخصائص والملامح الخاصة بتعابير الوجه والمأخوذة من البيانات الأولية ، عكس تقنيات تعلم الاله التي تستنبط مستوى منخفض من التفاصيل. حيث تتطلب هذه التقنيات الى خوارزميات اخرى لأغراض التصنيف او الاكتشاف للحصول على النتيجة المرغوبة.

يعتمد النموذج المقترح على تعبيرات وجوه الطلبة التي يتم التقاطها من كاميرا الكترونية أثناء التفاعل مع الحاسوب وتعتبر هذه كمدخلات الى الشبكة .

اثبتت النتائج العالية التي حققها الموديل من خلال الاختبار على بيانات غير موجوده في مرحلة التدريب والاختبار الى قدرة وكفاءة شبكات التعلم العميق على استنباط ادق التفاصيل ولعمق الترابط الداخلي بين طبقاتها.

وزارة التعليم العالي والبحث العلمي
جامعة بابل - كلية العلوم للبنات
قسم علوم الحاسوب

اكتشاف مستوى ارتباط الطلاب اثناء التفاعل مع الحاسوب باستخدام تقنيات التعلم العميق

بحث مقدم إلى
كلية العلوم للبنات، جامعة بابل
جزءا من متطلبات نيل درجة الدبلوم العالي في علوم الحاسوب

مقدمة من قبل
زهراء احمد جاسم

بإشراف
ا.م.د. امنة عطية داود

١٤٤٤هـ ٢٠٢٢م