12-2019

# Exploring Emotion Recognition for VR-EBT Using Deep Learning on a Multimodal Physiological Framework

Nicholas Dass
*Sheridan College*, nichnich@sheridancollege.ca

# EXPLORING EMOTION RECOGNITION FOR VR-EBT USING DEEP LEARNING ON A MULTIMODAL PHYSIOLOGICAL FRAMEWORK

A Thesis

Presented to

The Faculty of Applied Science and Technology, School of Applied Computing

of

Sheridan College, Institute of Technology and Advanced Learning

by

NICHOLAS DASS

In partial fulfillment of the requirements

for the degree of

Bachelor of Computer Science (Mobile Computing)

December 2019

ABSTRACT

**EXPLORING EMOTION RECOGNITION FOR VR-EBT USING A DEEP LEARNING ARCHITECTURE ON A MULTIMODAL PHYSIOLOGICAL FRAMEWORK**

**Nicholas Dass**
**Sheridan College, 2019**

**Advisor:**
**Dr. Khaled Mahmud**

Post Traumatic Stress Disorder is a mental health condition that affects a growing number of people. A variety of PTSD treatment methods exist, however current research indicates that virtual reality exposure-based treatment has become more prominent in its use. Yet the treatment method can be costly and time consuming for clinicians and ultimately for the healthcare system. PTSD can be delivered in a more sustainable way using virtual reality. This is accomplished by using machine learning to autonomously adapt virtual reality scene changes. The use of machine learning will also support a more efficient way of inserting positive stimuli in virtual reality scenes. Machine learning has been used in medical areas such as rare diseases, oncology, medical data classification and psychiatry. This research used a public dataset that contained physiological recordings and emotional responses. The dataset was used to train a deep neural network, and a convolutional neural network to predict an individual's valence, arousal and dominance. The results presented indicate that the deep neural network had the highest overall mean bounded regression accuracy and the lowest computational time.

Keywords: Machine Learning, Deep Neural Network, Virtual Reality Exposure-Based Treatment, Emotion Recognition, Physiological Signals, Multimodal.

***ACKNOWLEDGEMENTS***

I would like to thank Dr. Khaled Mahmud for his efforts and patience in guiding the research process, and the thesis advisory committee for their feedback; Dr. El Sayed Mahmoud, Dr. Abdul Mustafa, Dr. Edward Sykes, and Dr. Haya El Ghalayini. I would also like to thank fellow researcher Daniel Picott for laying the model groundwork from our research in Ubiquitous Computing – Smart Health, which I was able to build upon.

To my fellow co-workers at the Bank of Montreal, thank you for your support and interest in my research. To my manager Jan Demek, thank you for your encouragement and support in helping me juggle the demands of both school and work. And lastly, I'd like to thank my close friends and family for their support throughout my academic progress, which has culminated in my contribution to the computer science community.

## *TABLE OF CONTENTS*


| | |
|---|---|
| Acknowledgements | i |
| Table of Contents | ii |
| List of Tables | iii |
| List of Figures | iv |
| Chapter One 1. Introduction | 1 |
| 1.1 The Problem Context | 1 |
| 1.2 Terms and Definitions | 1 |
| 1.3 Problem Statement | 4 |
| 1.4 Purpose | 4 |
| 1.5 Motivation | 6 |
| 1.6 Proposed Work | 7 |
| 1.7 Thesis Statement | 8 |
| 1.8 Contributions | 8 |
| 1.9 Organization of Thesis | 9 |
| Chapter Two 2. Literature Review | 10 |
| 2.1 Virtual Reality Exposure-Based Treatment | 11 |
| 2.2 Machine Learning Applications in Medical Diagnosis | 13 |
| 2.3 Physiological Indicators and Emotional States | 16 |
| 2.3.1 Heart Rate | 18 |
| 2.3.2 Respiration | 20 |
| 2.3.3 Eye Movement | 24 |
| 2.3.4 Brain Waves | 26 |
| Chapter Three 3. Methodology | 35 |
| 3.1 Data Source | 36 |
| 3.2 Processing Data | 38 |
| 3.3 Research Design | 41 |
| Chapter Four 4. Results & Analysis | 50 |
| 4.1 CNN & DNN Training – 25 Epochs | 50 |
| 4.2 CNN & DNN Training – 50 Epochs | 55 |

4.3 CNN & DNN Training – 100 Epochs 63
4.5 Computational Time 65
4.4 Mapping DNN & CNN Model Predictions to PTSD Patients 67
Chapter Five 5. Conclusion 68
5.1 Conclusion 68
5.2 Limitations 68
5.3 Future Work 69
References 70

## *LIST OF TABLES*


Table 1. Terms and Definitions 1

Table 2. Data Summary 36

Table 3. Deep Neural Network Model Architecture 46

Table 4. Convolutional Neural Network Model Architecture 48

Table 5. CNN Performance Metrics – 25 Epochs 51

Table 6. DNN Performance Metrics – 25 Epochs 53

Table 7. CNN Performance Metrics – 50 Epochs 56

Table 8. DNN Performance Metrics – 50 Epochs 58

Table 9. CNN Performance Metrics – 100 Epochs 60

Table 10. DNN Performance Metrics – 100 Epochs 63

Table 11. Specifications for System Hardware 65

Table 12. Specifications for NVIDIA GeForce GTX 960M 66

## *LIST OF FIGURES*


Figure 1. Overview of an autonomous virtual reality exposure-based system. 5

Figure 2. Critical problem area. 6

Figure 3. Circumplex theory of emotions 17

Figure 4. Critical section. 35

Figure 5. Respiration physiological reading 38

Figure 6. Plethysmograph physiological reading 39

Figure 7. Horizonal Eye Movement physiological reading 40

Figure 8. Vertical Eye Movement physiological reading 40

Figure 9. Brain Waves physiological reading 41

Figure 10. Circumplex theory of emotions with 4 quadrant description 43

Figure 11. PAD 3-D emotional representation model 44

Figure 12. Deep Neural Network Architecture 45

Figure 13. Convolutional Neural Network Architecture 47

Figure 14. Convolutional Neural Network – 25 Epochs 50

Figure 15. CNN prediction accuracy – 25 Epochs 52

Figure 16. Deep Neural Network – 25 Epochs 53

Figure 17. DNN prediction accuracy – 25 Epochs 54

Figure 18. Convolutional Neural Network – 50 Epochs 55

Figure 19. CNN prediction accuracy – 50 Epochs 56

Figure 20. Deep Neural Network – 50 Epochs 57

Figure 21. DNN prediction accuracy – 50 Epochs 59

Figure 22. Convolutional Neural Network – 100 Epochs 60

Figure 23. CNN prediction accuracy – 100 Epochs 61

Figure 24. Deep Neural Network – 100 Epochs 62

Figure 25. DNN prediction accuracy – 100 Epochs 64

## *CHAPTER ONE*

## *1. Introduction*

### *1.1 The Problem Context*

The current version of virtual reality exposure-based treatment (VR-EBT) for treating PTSD, is a manual process that involves setting up a traumatic virtual reality scene that a patient will experience. A clinician inserts positive stimuli at trauma points in the scene to help a patient desensitize from the traumatic event. The patient undergoes this treatment several times for the treatment to be effective. While VR-EBT is widely acknowledged and validated as a method of treatment for PTSD, as supported by (Urella, et al, 2017) (Botella, et al., 2015) (Gavhane, et al., 2016), the delivery of VR-EBT is still based on manual processes (Urella, et al, 2017), which can make this treatment exceedingly expensive (Botella, et al., 2015) and therefore, unsustainable in the long run.

### *1.2 Terms and Definitions*

**Table 1. Terms and Definitions**

| | |
|---|---|
| **Affective States** | An individual's experience of emotion. |
| **Arousal** | Level of awareness based on psychological indicators, which also affects perception. |
| **Brain Waves** | Patterns of neural activity that represent electrical impulses in the brain. |

| | |
|---|---|
| **Circumplex Theory of Emotion** | Two-dimensional emotional space. Dimensions include arousal and valence. Intersection is the neutral state. |
| **Clinician** | A health care practitioner who works with a patient who suffers from PTSD. In the context of virtual reality exposure-based treatment, the clinician delivers the treatment to the patient. |
| **Convolutional Neural Network** | A type of deep learning network that uses a mathematical operation called convolution. |
| **DEAP Dataset** | Dataset that contains physiological indicator recordings of several participants. |
| **Deep Neural Network** | A machine learning method that can be supervised or un-supervised. |
| **Dominance** | The level of dominance or submissiveness of an emotional state. |
| **ECG** | Electrocardiography is used to record electrical activity of the heart. |
| **EEG** | Electroencephalogram is used to monitor brain waves. |
| **Emotion** | State of mind that is affected by current environment. |
| **Eye Movement** | The frequency of eye movement. |

| | |
|---|---|
| **Heart Rate** | The number of times the heart contracts per minute. |
| **MAHNOB-HCI Tagging Dataset** | Dataset that contains physiological indicator recordings and emotion responses of several participants. |
| **Patient** | A person who suffers from PTSD and receives virtual reality exposure-based treatment. |
| **Physiological Indicators** | Physiological indicators refer to heart rate, respiration rate, rapid eye movement, and brain waves data that will be used to train the machine learning algorithms. |
| **PPG** | Photoplethysmography measures blood volume changes in microvascular skin tissue. |
| **PTSD** | Post Traumatic Stress Disorder is a mental health disease that a patient suffers from. PTSD is triggered by a traumatic event. |
| **Respiration Rate** | The number of times the lungs contract, or the number of times a person breathes per minute. |
| **Signal Channel** | Signals produced from recorded physiological indicator activity, such as brain waves. |
| **SVM** | Support Vector Machine uses machine learning algorithms for classification and regression. |

| | |
|---|---|
| **Valence** | Refers to positive or negative feelings that a person experience in their current environment. |
| **Virtual Reality Exposure Based Treatment (VR-EBT)** | A treatment method that uses virtual reality to treat patients who suffer from PTSD. |
| **Vivo Exposure** | A form of Cognitive Behaviour Therapy to reduce fears. Two types are flooding and systematic desensitization. |

### *1.3 Problem Statement*

Patients who suffer from PTSD seek the use of virtual reality exposure-based treatment. However, the labor-intensive nature of clinician intervention hinders the ability to scale the use of this treatment. The labour-intensive nature can be attributable to two main areas – manual deduction of human emotion and insertion of positive stimuli. Deduction of human emotion by a clinician involves reviewing a patient's vitals to determine how the patient feels. Manually inserting positive stimuli by adapting a virtual scene, helps a patient desensitize from trauma This requires the clinician to distribute time between assessing the patient's vitals and responding with the insertion of positive stimuli. The scarcity of research applying deep learning in VR-EBT does not provide medical or health practitioners an opportunity to translate research into an improved delivery system.

### *1.4 Purpose*

The purpose of this research is to contribute to a larger effort in realizing an autonomous way of delivering VR-EBT. The high-level overview of an autonomous VR-EBT system is depicted in

Fig 1. It illustrates how an autonomous VR-EBT system would intuitively create VR scenes that are specific to each patient's trauma. Furthermore, it depicts how physiological data is collected from a patient using wearables. For the purpose of this research, it is important to note that an individual's emotional state affects their physiological state and vice versa. In light of this, individuals who suffer from PTSD experience more pronounced emotional and physiological states than individuals who are not diagnosed with PTSD. The physiological data collected from the patient is used to train a deep neural network to predict a patient's approximate grouping of emotional states. The results of the prediction can then be used to adapt the virtual reality traumatic scene, which can also be described as automatically inserting positive stimuli.

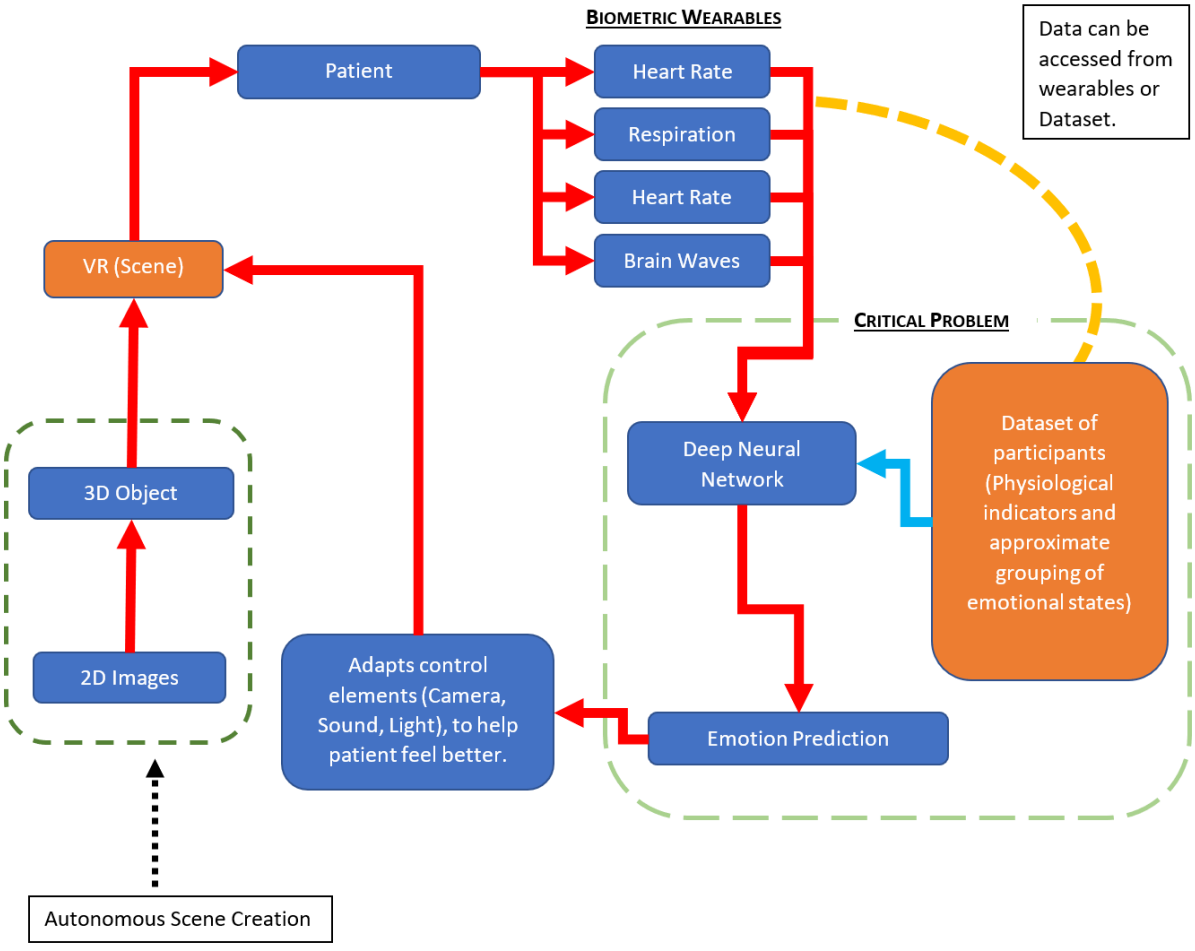


**Figure 1.** A high-level overview of an autonomous virtual reality exposure-based system. The light green dotted area highlights the critical problem that will be explored in this research.

The critical part of this research is illustrated in Fig 2. Data that depicts physiological indicators was used to train a deep neural network to predict the valance, arousal and dominance of a test subject.

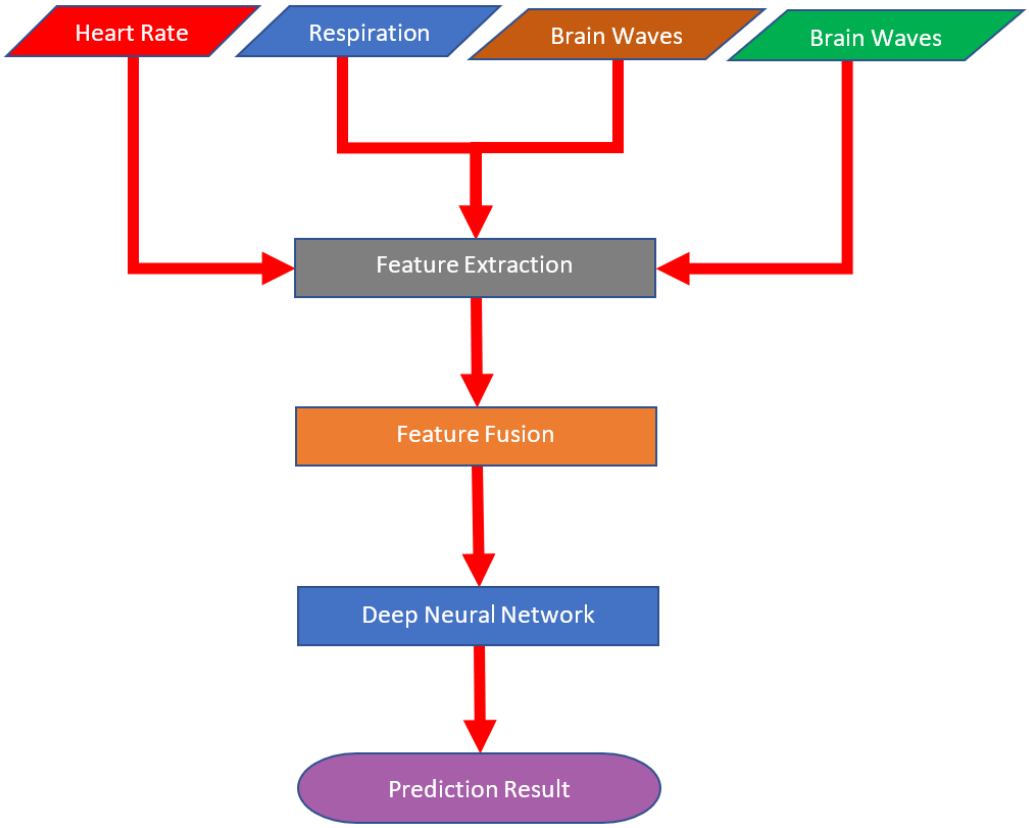


**Figure 2.** The critical problem area. Features from a dataset will be used to train the deep neural network.

### *1.5 Motivation*

This research aims to contribute to a larger effort to improve the delivery of virtual reality exposure-based treatment, which will improve the quality of life for patients who suffer from PTSD. As discussed previously, this is achieved through automating virtual reality exposure-based treatment. However, the current process of developing exposure-based treatment for virtual reality is tedious and costly. The motivation to pursue this research is driven by three reasons – improve the delivery method, reduce manual tasks required by the clinician, and successful applications of

machine learning in medicine. Improving the delivery method can be accomplished by creating a sustainable version of virtual reality exposure-based treatment that is scalable. This can have a greater impact on treating a growing number of patients who suffer from PTSD, while lowering costs related to sustaining virtual reality exposure-based treatment within the healthcare system.

Reducing manual tasks such as the assessment of emotional state or modifying virtual reality scenes will help the clinician. It will help clinicians focus more of their attention on other parts of the treatment process. Successful applications of machine learning in medical diagnosis can be found in oncology (Turki, 2018), medical data classification (Seera & Lim, 2014), psychiatry (Omurca & Ekinci, 2004), as well as the assessment of Parkinson's (Eskofier B. M. et al., 2016). Therefore, it is worthy to explore the use of machine learning to determine its efficacy in its use to predict human emotion, which is explored in the rest of this research.

### *1.6 Proposed Work*

This research did investigate the prediction accuracy of a deep neural network using multimodal physiological signals. The first part reviewed existing research that discuss physiological indicators and its correlation to human emotion. The second part used a public dataset, which consists of participants physiological signals and an approximate grouping of emotional states. The dataset was used to train a deep neural network and a convolutional neural network. The prediction accuracy of the deep neural network will be compared to the convolutional neural network to determine its performance with emotion recognition.

### *1.7 Thesis Statement*

An individual's valance, arousal and dominance affect their physiological state and vice versa. A deep neural network can be used to predict human valence, arousal and dominance using physiological indicators such as heart rate, respiration rate, rapid eye movement and brain waves. This research will develop a deep neural network and investigate the prediction accuracy of the network in emotion recognition.

### *1.8 Contributions*

This research explored the use of a deep neural network and compared it with a convolutional network to determine the prediction accuracy for emotion recognition. The following was executed as part of this research;

- Explored current research to determine an overview of physiological indicator bounds for; heart rate, respiration rate, rapid eye movement, brain waves that correlate to emotional states.
- Extracted physiological data and an approximate grouping of emotional states from a dataset to train and test a deep neural network.
- Constructed a deep neural network and a convolutional neural network.
- Presented performance results on the DNN and CNN.

### *1.9 Organization of Thesis*

This research will first begin with an overview of virtual reality exposure-based treatment. The critical problem that this research paper seeks to address will then be discussed. More specifically, exploring the use of a deep neural network with multimodal physiological signals and an approximate grouping of emotional states for emotion recognition. Furthermore, the prediction

accuracy of the deep neural network will be examined. The structure of this thesis will consist of a literature review, methodology, results, analysis and conclusion. The literature review will discuss previous work in the area of virtual reality exposure based-treatment, applications of machine learning in medical diagnosis, and the relationship between physiological indicators and emotional states in the context of emotion recognition.

The methodology section of this research will discuss how the critical problem can be addressed. More specifically, how the public dataset that contains physiological data and an approximate grouping of emotional states will be used to train a deep neural network for emotion recognition. In addition, the prediction accuracy of the deep neural network will be evaluated. The results and analysis section will discuss the performance of the deep neural network in emotion recognition, and limitations of the methodology. Furthermore, it will also discuss future work that this research can be used to build upon.

## *CHAPTER TWO*

## *2. LITERATURE REVIEW*

The use of virtual reality exposure-based treatment has become a validated approach in treating post traumatic stress disorder. A keyword search for the treatment will generate numerous journals that discuss the validity of VR-EBT. In a study by (Srivastava, et. al., 2014), the authors state that VR related scientific articles grew from 45 in 1995 to 3,203 in 2010. While the treatment method will provide long-term benefits for the patient, the ability to scale this treatment to cater to a growing number of patients who suffer from PTSD is costly (Botella, et al., 2015). Furthermore, psychiatric treatment can be expensive for patients in a fee for service healthcare system and can also be a burden within a universal healthcare system. Even with a semi-autonomous system which attempts to automatically re-create virtual reality scenes based on real life images of a patient's trauma (Urella, et al, 2017), the treatment method is still labour intensive, and only targets one part of the delivery framework.

To reach a more autonomous method of delivering VR-EBT, machine learning can play an important role in emotion recognition, which can work alongside semi-autonomous virtual reality scene creation. Emotion recognition can act as positive stimuli, by adapting virtual reality scenes to desensitize the negative feelings that a patient has towards a traumatic scene. This is an important component to realize a more autonomous VR-EBT system, as the efficacy of this critical part can either work in favor of acting as positive stimuli or be a detriment. Support for the use of machine learning in medical applications is a growing trend, some examples are; the assessment of Parkinson's disease (Eskofier B. M. et al., 2016), cancer identification (Turki, 2018), medical data classification (Seera & Lim, 2014), and closer to the domain of this research paper is the

evaluation of PTSD in individuals using machine learning (Omurca & Ekinci, 2004). The literature reviewed will cover three main areas. The first will discuss literature that supports the use of VR-EBT as valid treatment method for PTSD. This is the foundation to which the goal of this research aims to improve upon. The second will discuss medical applications that use machine learning as part of prediction and assessment. The third will discuss literature that presents the relationship between physiological indicators and human emotion in the context of using emotion recognition. Understanding this relationship will be paramount in training a deep neural network, which will use physiological signals as data inputs. Physiological indicators that are explored are; hear rate, respiration, eye movement, and brain waves.

### *2.1 Virtual Reality Exposure Based-Treatment*

Traditional methods of treating post-traumatic stress disorder, include prolonged-exposure therapy, and cognitive processing therapy. According to (Gavhane, et. al., 2016), prolonged-exposure therapy refers to a delivery method where a clinician instructs a patient to recall traumatic memories. This invokes feelings associated with the trauma. With respect to cognitive processing therapy, (Gavhane, et. al., 2016) states that this treatment method involves altering a person's thought process, as a result of the traumatic event. While traditional methods are effective, there are some limitations that (Mishkind, et. al., 2017) and (Banerjee D. et al., 2017) discuss. They include;

- Limited control of the traumatic experience.
- Privacy concerns during the delivery of treatment.
- Difficulty in accessing stimuli.
- Financial and time resource scarcity.

- Patient inability and reluctance to recall traumatic memories.
- Patient is reluctant to use treatment method.

In contrast, virtual reality exposure-based treatment is accomplished by helping a patient recall traumatic events through virtual reality scenes. According to (Jerdan, et. al., 2018), the authors refer to this as an illusion of reality. The authors state that even though patients are aware of their virtual environment, patients perceive the images and sound to be real stimuli, which is still effective in the treatment of PTSD. This is consistent with (Freeman, et. al., 2017), who describe VR-EBT treatment as a substitution of real-world sensory experience with virtual stimuli.

The use of this treatment method provides clinicians the ability to monitor traumatic events, and manually insert positive stimuli by altering virtual reality scenes. This will help a patient desensitize from the traumatic event. VR-EBT typically involves 6 to 13 sessions or more, which is dependent on the profile of the patient, their related mental health condition, as well as frequency in terms of how consistent a patient is with treatment (Rothbaum, et. al., 2014) (Georgina, et. al., 2013). Virtual reality exposure-based treatment addresses most of the limitations of traditional treatment methods and is proven to be successful in its use (Gavhane, et. al., 2016) (Wiederhold, et. al., 2018). This is consistent with a study by (Rizzo, et. al., 2014), where one trail of 20 active duty service members from the Iraq war, experienced a 50% decrease in PTSD related symptoms. In another trial of 24 active service members, 45% of participants no longer experienced PTSD, and 62% had made modest improvement.

Similarly, (Srivastava, et. al., 2014) states that after 6 months, patients who followed up, experienced a reduction in PTSD symptoms that ranged between 15% to 67%. Furthermore, (Rothbaum, et. al., 2014) reports that in their study of Iraq and Afghanistan War veterans, PTSD symptoms significantly improved after 6 sessions of virtual reality treatment. The improvement

reported was maintained at 3, 6 and 12 months. In another study that explores the use of VR-EBT on Mexican victims of criminal violence, (Georgina, et. al., 2013) states that after 13 sessions of treatment, a group of 20 victims (55% had acute PTSD and 45% had chronic PTSD) no longer meet the diagnostic criteria of PTSD.

### *2.2 Machine Learning Applications in Medical Diagnosis*

The assessment of Parkinson's disease using deep learning on sensor data by (Eskofier B. M. et al., 2016), focused on detecting bradykinesia, which is a Parkinson's disease that affects the motor system. According to the authors, the algorithms include boosting, decision trees, k-nearest neighbours, and support vector machines. The classification method that (Eskofier, et. al., 2016) used, includes pre-processing of data, feature extraction, and classifier training. For pre-processing, the authors extract non-overlapping segments from sensor data, which relates to individual tasks. Feature extraction involved the use of 8 standard machine learning features, and for classifier training the authors used AdaBoost.M1, PART, k-nearest neighbors (kNN), and support vector machines (SVM). According to (Eskofier, et. al., 2016), the deep learning framework consists of an input layer, two convolutional neural networks layers that use rectified linear units (ReLUs), with max pooling, and a soft-max output layer for classification.

The classification accuracy based on the observed results by (Eskofier, et. al., 2016) are; AdaBoost.M (86.3%), PART (81.7%), kNN (67.1%), SVM (85.6%), and Deep Learning (90.9%). The authors also discuss some of the following advantages of deep learning;

- Expert defined features for signal classification are not required.
- Resembles analysis done by human experts, as the input signal is assessed as one output.

- Adapting deep learning network to a single patient is possible.
- Deep learning networks produce better classification results with large datasets, which is consistent with findings by (Banerjee, et al. (2017).

Another crucial part of this study that (Eskofier, et. al., 2016) discuss, is the limitation that researches faced. The first limitation pertains to a non-optimized machine learning pipeline and deep learning parameters, which was done to carry out a fair comparison with deep learning. The second limitation relates to a limited database. The reason this is crucial, is that machine learning algorithms in general require a sufficiently large dataset to be trained with, so that more accurate classification can be made. Similarly, it would be crucial to gather as much input data as possible on physiological indicators, so that classification is more accurate with emotion recognition.

The research that relates to cancer identification (Turki, 2018), uses machine learning algorithms such as AdaBoost, Deepboost, Xgboost, and Support Vector Machines (SVM) on datasets pertaining to thyroid cancer, colon cancer, and liver cancer. Some of the algorithms are used by (Eskofier, et. al., 2016) in their study. According to (Turki, 2018), support vector machines showed promising results over other machine learning algorithms explored.

With respect to medical data classification, (Seera & Lim, 2014) state that medical knowledge and treatment therapy, such as new diseases and available drugs, are advancing at a rapid pace. In light of this, physicians find it increasingly difficult to keep up. The authors propose a hybrid intelligent system, that helps physicians focus on pertinent medical data, to make more informed decisions on medical prognosis and diagnosis. The proposed system by (Seera & Lim, 2014), employ the use of neural networks. The authors state that the advantages of neural networks support medical decision applications. This is in contrast with expert systems, that can be taxing in its ability to establish relationships between ever growing input symptoms and target diseases.

According to (Seera & Lim, 2014), the neural network has three models in their hybrid intelligent system, which include; Fuzzy Min-Max (FMM), Classification and Regression Tree (CART), and Random Forest (RF). In the context of the proposed hybrid system, the authors explain that FMM is primarily concerned with medical decision support by learning from data samples, CART and RF are incorporated to strengthen FMM, to explain predicted output, and attain high classification performance respectively.

The authors state that the proposed hybrid system achieves two practical key objectives; the first relates to the system's ability to explain and justify the predictions; this is key as medical practitioners need to be confident in the system's ability to accurately provide medical prognosis and diagnosis, which is critical with respect to safety. The second, relates to accuracy; if the system is not accurate in its classification, it can put patients at risk by either denying them medical attention or receiving improper medical prognosis and diagnosis. It can also place undue stress on patients and end up being a burden on resources. Otherwise, the authors state that the system can reduce costs, and provide efficiencies in healthcare delivery. Observed results by (Seera & Lim, 2014) note that the combination of the three models that are employed in the hybrid system range between 95% and 99% in terms of accuracy, sensitivity and specificity when examining the Breast Cancer Wisconsin data set. The authors also state that FMM-CART-RF offer promising results in other data sets.

In another study, (Omurca & Ekinci, 2004) evaluate the use of machine learning in diagnosing post-traumatic stress disorder. They attribute the use of machine learning in diagnosis as a result of its high capability and effective classification ability. The authors explore three different machine algorithms; Sequential Minimal Optimization, Multilayer Perceptron, and Naïve Bayes. And offer useful insight into characteristics of each of these machine learning algorithms.

According to (Omurca & Ekinci, 2004), the proposed system that evaluates PTSD, begins with processing data from a PTSD database. This step entails feature selection that uses chi-square, principal component analysis, and correlation based-feature selection. This includes class conditional independence, which refers to features being independent of each other. The next step in the system, classifies the pre-processed data set using three classification algorithms; Sequential Minimal Optimization, Multilayer Perceptron, and Naïve Bayes. The following is an overview of the characteristics of each classification algorithm discussed by (Omurca & Ekinci, 2004);

*Sequential Minimal Optimization*: Is an advanced version of support vector machines, solves real-time problems, is easy to implement, offers fast classification, and is best suited for large datasets.

*Multilayer Perceptron*: Is highly accurate and can generalize well. The authors note that it is crucial to define network structure, functions and parameters for this classification algorithm. Once this is done, training can be executed.

*Naïve Bayes*: Is known as an inductive learning algorithm. Performs fast and can classify complex dimensional data sets. It is preferred in real world applications, particularly in medical diagnosis.

(Omurca & Ekinci, 2004) observed an accuracy range between 74% and 79% in the system's ability to distinguish between individuals with or without PTSD. Furthermore, the authors describe the proposed assessment system as flexible and easily adaptable to other use cases in medical diagnosis.

### *2.3 Physiological Indicators and Emotional States*

The purpose of investigating physiological indicators are two-fold; 1. Physiological indicators determine emotions expressed by individuals, which is consistent with studies by (Verma & Tiwary, 2014) and (Vijayan, et. al., 2015) and 2. Data extracted from the relationship

between physiological indicators and human emotion, will be used to train a deep neural network for emotion recognition. There are several physiological indicators that can be used for emotion recognition. However, this research will focus on heart rate, respiration, eye movement, and brain waves. With respect to emotional indicators, arousal and valence will be explored.

Based on circumplex model of emotion, (Posner, et. al., 2009) state that valance (Pleasure to displeasure | x-axis) and arousal (Activated to deactivated | y-axis) form a two-dimensional axis to depict four main emotional areas seen in Fig 3.

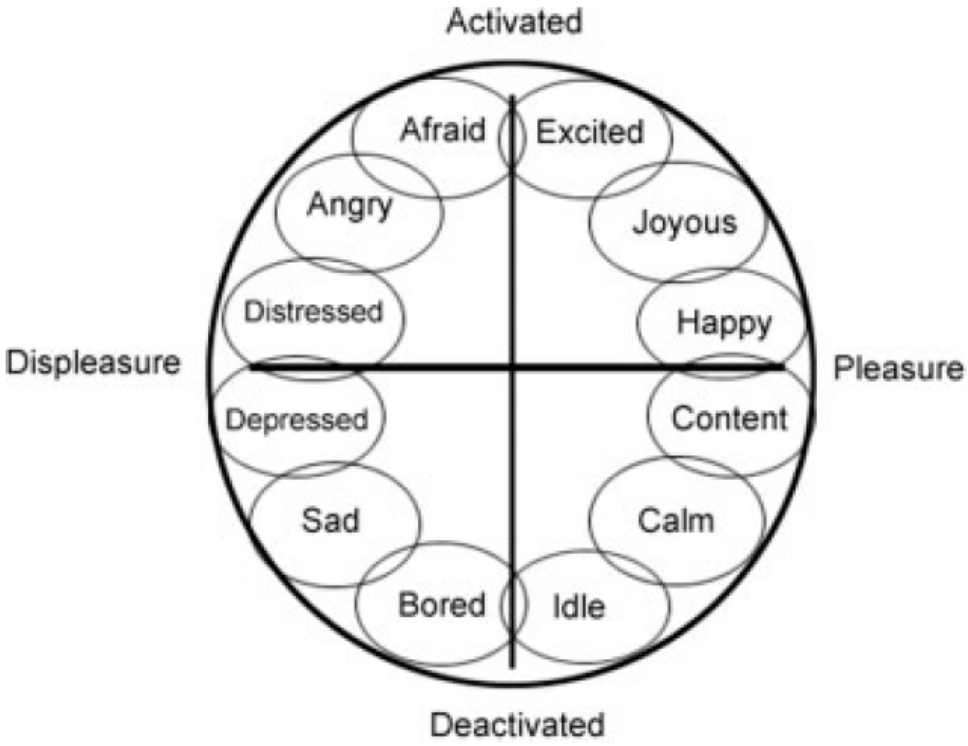


**Figure** 3. Circumplex theory of emotions segregates the approximate grouping of emotions into four quadrants (Posner, et. al., 2009).

- *High Valance & High Arousal* (High intensity of positive feelings, and high awareness of emotional and physiological responses) – Top right of the x and y axis correspond to being excited, joyous, and happy.

- *High Valence & Low Arousal* (High intensity of positive or negative feelings and low awareness of emotional and physiological responses) – Bottom right of the x and y axis correspond to being content, calm and idle.
- *Low Valance & High Arousal* (Low intensity of positive or negative feelings and high awareness of emotional and physiological responses) – Top left of the x and y axis correspond to being afraid, angry and distressed.
- *Low Valance & Low Arousal* (Low intensity of positive or negative feelings and low awareness of emotional and physiological responses) – Bottom left of the x and y axis correspond to being depressed, sad, and bored.

### *2.3.1 Heart Rate*

Heart rate is one of four physiological indicators explored in this research as part of a multimodal framework in emotion recognition. In the DEAP dataset (Koelstra, S., et al., 2012), heart rate is extracted using plethysmography which is accomplished by using a probe and light source to detect cardio-vascular pulse waves (Elgendi, 2012). Before delving into emotion recognition using heart rate in the context of machine learning, it is fundamentally important to understand how the acceleration and deceleration of heart rate corelates to emotion. Research presented by (Shi, et. al., 2017), discuss heart rate variability (HRV) on two emotional states; happiness and sadness. The research explores how the automatic nervous system would affects six emotional states; disgust, surprise, anger, fear, and sadness. According to (Shi et. al., 2017), heart rate variability refers to tiny variations between intervals in sinus heart beats. Furthermore, the authors state that time-domain, frequency domain, as well as non-linear indices of heart rate variability are used in establishing a relationship between HRV and emotion.

The base line heart rate according to (Shi, et. al, 2017) is approximately 71 beats per minute for females, and approximately 74 beats per minute for males. This base line is important, as it is used to establish the context in which the authors describe the increase and decrease in heart rate with respect to emotion. The results described by (Shi, et. al., 2017) show that when participants were happy, their heart rate decelerated, and when they were sad, their heart rate accelerated. These results are consistent with (Ekman, et. al., 1983). However, research presented by (Etzel, et. al., 2006) indicate otherwise. The authors hypothesize that cardiovascular and respiration rate induced by music, can affect mood changes. Based on this, (Etzel, et. al., 2006) observed that heart rate decelerated during sadness and accelerated during fear.

With respect to emotion recognition in the context of machine learning, the relationship between heart rate and emotion is more evident. Research presented by (Kim, et. al., 2004) propose a physiological signal-based emotion recognition system. The first step in the system according to the authors, is to implement characteristic waveform detection, and obtain relevant data features for pattern classification. Pre-processed data is then feed into a support vector machine, which is used to classify patterns. The authors state that the purpose of using SVM in their study, is to resolve issues with pattern classification. This was a result of feature variation within the same class, as well as overlapping classes. Subsequently, the results shown by (Kim, et. al., 2004) state that SVM had a classification accuracy of 78.4% for 3 emotional states; sadness, anger and stress, and 61.8% for four emotional states; sad, stressed, angry and surprised.

A study by (Yu & Chen, 2015), present a method for emotion recognition using ECG, more specifically heart rate variability (HRV), to recognize feeling neutral, happy, stressed and sad. The authors state that four features of heart rate variability (HRV) are used; time-domain, frequency-domain, Poincare plot, and differential features. With respect to a classification algorithm, (Yu &

Chen, 2015) use support vector machine (SVM) for the following reasons; it is a statistical classifier, it uses the structural risk minimization principle of machine learning and can handle linearly inseparable data. Furthermore, the authors evaluate each feature against each emotion, which is known as a one-against-all approach.

Based on the results of the study, (Yu & Chen, 2015) note that when Genetic Algorithm (GA) was not used with feature selection, classification accuracy ranged between 37.5% to 52.2% when features were not combined. The results were the same when features were combined without the use of GA. However, when GA was used with feature selection, classification accuracy ranged between 42.5% to 60% for each feature. When features were combined, and GA was used, classification accuracy increased to 90%. In light of the observed results, (Yu & Chen, 2015) conclude that Genetic Algorithm feature selection must be part of an emotion recognition system to achieve high classification accuracy.

### *2.3.2 Respiration*

Emotional state has been known to affect respiration and vice versa (Jerath & Crawford, 2015) (Wu, et. al., 2012). Like other physiological indicators discussed in this review, respiration can be used in emotion recognition (Zhang, et. al., 2017). In a study that examines the relationship between respiration and psychological activity, (Zhang, et. al., 2017) propose a deep learning framework to classify emotion using respiration data. The authors focus on the use of a deep neural network because of its ability to extract features automatically and its low computational resource outlay. This contrasts with manually extracting features, which have a few limitations. According to (Zhang, et. al., 2017) these limitations include;

- Poor domain knowledge in feature creation to capture properties of a signal channel.

- No certainty that an algorithm used for feature selection will produce an optimal feature set.
- Loss of information as manual feature selection uses statistics, which can't depict signal channel details.

(Zhang, et. al., 2017) also provide scenarios that describe the relationship between the pace and intensity of breathing to emotion, which are also consistent with (Jerath & Crawford, 2015). The authors state the following;

- *Deep & Fast Breathing* – Relates to excitement, which can depict the following emotional states; happy, angry or afraid.
- *Shallow & Fast Breathing* – Relates to tension.
- *Deep & Slow Breathing* – Relates to relaxation
- *Shallow & Slow Breathing* – Relates to clam or negative states.

With respect to respiration rates, (Zhang, et. al., 2017) state that individuals who experience excitement typically respire between 40 – 50 times per minute, while individuals who respire 20 times per minute are typically calm. Feeling tension and relaxion fall between the upper range of 40 – 50 times per minute and 20 times per minute respectively. To conduct the evaluation, (Zhang, et. al., 2017) use the dataset for emotion analysis using EEG, physiological and video signals (DEAP) database (Koelstra, S., et al., 2012), which is evaluated against the circumplex theory of emotion.

With respect to classification, (Zhang, et. al., 2017) hypothesize that the use of a sparse auto encoder, which is a deep learning method part of a broader deep learning framework, will extract the best characteristics of automatic feature selection from unlabeled data. The authors state

that this will provide a high level of classification accuracy. The next step in the process according to (Zhang, et. al., 2017) is pre-training, which occurs on the first sparse auto encoder hidden layer. Further training can be applied to the second layer, as well as other hidden layers. The other part of the deep learning framework involves feeding features into two logistic regression algorithms, that will classify valence and arousal. The observed results by (Zhang, et. al., 2017) indicate that the use of a deep learning framework (Sparse auto encoder with logistic regression algorithms), provide classification accuracy of 73.06% for valence, and 80.76% for arousal. These results are similar to a study by (Zheng, et. al., 2019), who use a Deep Neural Network (DNN) to classify emotion using EEG and Eye movement.

In another study by (Wu, et. al., 2012), the authors propose a more accurate method in determining the relationship between respiration and the following affective states (emotion); love, sadness, joy, fear, and anger. According to (Wu, et. al., 2012), the proposed method focuses on extracting Emotion Elicited Segments (EES) from the respiration signal. According to (Wu, et. al., 2012), this involves the use of Mutual Information-Based Emotion Relevance Feature Ranking based on Dynamic Time Warping Distance (MIDTW), and Constraint-based Elicited Segment Density (CESD) analysis. The authors state that the purpose of proposing a more accurate method is to remove distortion from the respiration signal. (Wu, et. al., 2012) state that this is especially true when non-invasive biosensors, used to track physiological indicators, are more susceptible to distortion. The authors note that this may negatively affect accuracy results. (Wu, et. al., 2012) continue to state that respiration is not only affected by affective states, but also by the following;

- Transition between emotion
- Level of body self-regulation
- Effects of voluntary breathing,

- Bodily motions; laughing, talking, sneezing, coughing etc.

In the context of distortion, (Wu, et. al., 2012) refer to bodily motions as motion artifacts that distort the affective state. Furthermore, motion artifacts are represented as low-frequency noise in the respiration signal, which cannot be removed. To remedy this, (Wu, et. al., 2012) focus on separating the respiration signal into quasi-homogenous segments. This enables the authors to detect and remove ambiguous emotion transition periods and motion artifacts. Once this is done, (Wu, et. al., 2012) are then able to use emotion recognition on the non-distorted segments. The authors state that they can accomplish this by using Respiratory Quasi-Homogeneity Segmentation (RHS), which is a three-step iterative process;

- Step 1: Signal Transformation – Sum the acceleration values of the squared original signal.
- Step 2: Top-down Splitting – Identify change points in the transformed signal.
- Step 3: Quasi-homogenous Test: Test sub-intervals that are coupled with nearby change points, against significance conditions described in the study.

According to (Wu, et. al., 2012), a comparative analysis using two nearest neighbor classifiers; k-Nearest Neighbor (KNN) and Probabilistic Neural Network (PNN), were used to evaluate the Emotion Elicited Segments. The approach applied by (Wu, et. al., 2012) involve the use of Emotion Elicited Segments (EES), for a specific emotion from a single subject, for each experiment as test data. The remaining clusters were used to train the classification algorithms. The authors also distinguish between length voting; emotion with greatest total segment length, and number-voting; greatest number of segments. Both cases provided the dominant emotion.

Observed results by (Wu, et. al., 2012) show that with Emotion Elicited Segments (EES), both k-Nearest Neighbor (KNN) and Probabilistic Neural Network (PNN) perform well, with PNN performing slightly better than KNN. The authors note that on average the classification accuracy for KNN on length-voting and number-voting is 86.09% and 90.43% respectively. In addition, the classification accuracy for PNN on length-voting and number-voting is 87.83% and 92.17% respectively. In contrast, (Wu, et. al., 2012) observe that when Emotion Elicited Segments (EES) are not applied, both k-Nearest Neighbor (KNN) and Probabilistic Neural Network (PNN) perform slightly lower. The classification accuracy for KNN on length-voting and number-voting is 86% and 82% respectively. In addition, the classification accuracy for PNN on length-voting and number-voting is 86% and 88% respectively.

### *2.3.3 Eye Movement*

Eye movement has been identified as another important physiological indicator of human emotion. In a study by (Zheng, et. al., 2019), the authors investigate a multimodal framework to predict human emotion using EEG and eye movement. The authors describe brain signals and eye movement as an encouraging way to model cognitive states. They investigate the strength of classification on emotional states such as being neutral, sad, fearful and happy. With respect to eye movement, the authors state that the pupil is a window into the brain. More specifically, the pupil responds by expanding or contracting based on an emotional response to a given situation. This is described as a natural way to observe human emotion. This view is supported by (Schurgin, et. al., 2014), who state that the eyes widen when facial expression of fear is perceived, and the eyes contract when facial expression of joy is perceived. In addition, (Wang, et. al., 2018) state that a person's eye movement reflects their perceived emotion from a given scene.

With respect to success rates in the accuracy of determining emotion through eye movement alone, (Schurgin, et. al., 2014) state that participants in their study had an overall accuracy range between 66% to 86.8% when judging emotional states of joy, disgust, fear, anger, sadness, and shame. A study by (Shields, et. al., 2012) discuss findings in a related study within their research, of emotion recognition using eye tracking. The study found an accuracy rate of 86.9% in classifying disgust and 86.1% in classifying anger, when a participant responded to corresponding facial expressions.

Some features that (Zheng, et. al., 2019) use as parameters in corelating dilation of the pupil with emotion include; pupil diameter, saccade (constant movement of the eyes), fixation (eyes are in fixed position), blink (rapid eye movement), and event statistics, which is also consistent in a study by (Wang, et. al., 2018). In terms of classification, the authors compare the efficacy of predicting emotion using the following scenarios;

- Eye movement (SVM).
- EEG using (SVM).
- EEG & Eye Movements into one feature (SVM).
- EEG & Eye Movements into one feature (DNN).

Similarly, (Wang, et. al., 2018) implement the use of SVM for feature and decision level fusion to classify emotional states of positive, neutral and negative from EOG and eye movement. The results observed by (Wang, et. al., 2018) show classification success ranged between 84.62% to 90.82%. However, the results provided by (Zheng, et. al., 2019) are more robust and offer a comparative analysis between the use of unimodal and multimodal physiological indicators;

*Eye Movement (SVM)*: Confusion matrix values for the target to predicted class are; sadness = 0.58, fear = 0.67, happy = 0.67, and neutral = 0.80.

*EEG (SVM)*: Confusion matrix values for the target to predicted class are; sadness = 0.63, fear = 0.65, happy = 0.80, and neutral = 0.78.

*Eye Movement & EEG (SVM)*: Confusion matrix values for the target to predicted class are; sadness = 0.69, fear = 0.79, happy = 0.73, and neutral = 0.82.

*Eye Movement & EEG (DNN)*: Confusion matrix values for the target to predicted class are; sadness = 0.85, fear = 0.85, happy = 0.74, and neutral = 0.92.

Based on the results, using a Deep Neural Network performed better overall than other scenarios. The authors note that using a multimodal approach to emotion recognition showed an 85.11% accuracy rate in emotion recognition. In contrast, a unimodal approach for EEG alone showed a 70.33% classification accuracy rate, and eye movement showed a 67.82% classification accuracy rate. Extracting multimodal physiological data provides a more wholesome approach to emotion recognition over a unimodal approach. The research done by (Zheng, et. al., 2019) supports the research conducted in this paper, where 4 different physiological indicators are used to predict human emotion.

### *2.3.4 Brain Waves*

Compared to other physiological indicators such as facial, vocal and body indicators, brain waves are far more transparent and concrete in its correlation to emotion. According to (Liu, et. al., 2018), brain waves offer an immediate response to emotional stimuli, which is difficult for a person to mask based on their EEG signals. With respect to emotion, (Liu, et. al., 2018) and (Mouhannad, et. al., 2018) describe two models;

- Discrete – Emotional space is limited to several basic emotions that consist of; joy, sadness, surprise, fear, anger and disgust.
- Dimensional – Two-dimensional space consists of valence (Negative/Positive characteristics)-arousal (Cognisant level of emotion and physiological indicators), and three-dimensional space consists of pleasure (Positive characteristics)-arousal-dominance (In control or being controlled).

(Liu, et. al., 2018) state that the use of the above models, are dependent on the number of emotions explored, as well as the level of difference between emotions. For example, if two emotions are significantly different such as being happy or sad, the dimensional model would work, as each of these emotions occupy different dimensional space. However, when emotions such as anxiety and anger are explored, the use of the discrete model its best suited, as the dimensional model is un-reliable in distinguishing between two emotions that are very close in dimensional space. This is fundamental, particularly with respect to how responsive and accurate a proposed emotion recognition system is.

Similar to previous studies discussed, the proposed framework by (Liu, et. al., 2018) consists of the following components; data acquisition, data processing to remove distortion, and feature extraction. According to the authors, support of a multiclass classification requires Library for Support Vector Machines (LIBSVM) to classify eight discrete emotions; joy, amusement, tenderness, anger, sadness, fear and disgust.

The emotions explored are also consistent in a study by (Shin, et. al., 2017). Furthermore, (Liu, et. al., 2018) improves upon the classification algorithm by creating three levels of classification. The authors first group similar emotions, three positive (joy, amusement, tenderness) and four negative emotions (anger, sadness, fear and disgust). Level 1 consists of

neural and non-neural emotions. Level 2 consists of non-neural positive and negative emotions. And level 3 consists of two groups positive and negative emotions. According to (Liu, et. al., 2018), the improvement to LIBSVM by including three levels of classification, showed a significant difference in classification accuracy. The following were the observed results;

- Default LIBSVM classification of eight emotions - 32.31%.
- Level 1 - 92.26 percent
- Level 2 – 86.63%
- Level 3 – Positive Emotion – 86.43%, Negative Emotion – 65.09%

A wealth of information can be found in a study by (Ismail, et. al., 2016), who explore the detection of human emotion through brain waves. The authors make a distinction between different types of brain waves, which is crucial in determining which parts of the brain are associated with human emotion experienced;

*Delta Wave*. Frequency Range: 0.5 – 3 Hz – Related to recovery and good sleep.

*Theta Wave*. Frequency Range: 3 – 8 Hz – Deep relaxation, mediation, and improved Memory.

*Alpha Wave*. Frequency Range: 8 – 12 Hz – Creativity, relaxation, and visualization.

*Beta Wave*. Frequency Range: 12 – 27 Hz – Awareness and concentration.

*Gamma Wave*. Frequency Range: 27 Hz – Regional learning, memory language preprocessing, and ideation.

The above brain wave categorization is also consistent with a study conducted by (Heraz, et. al., 2007) (Liu, et. al., 2018) (Vijayan, et. al., 2015). According to (Ismail, et. al., 2016) the process of gathering data begins with pre-processing data by eliminating unwanted data. This

involves removing disorder and movement, which refers to noise from static electricity or electromagnetic fields, and reference data. The next step that (Ismail, et. al., 2016) describe, is feeding the pre-processed brain wave data into a neural network to classify emotion. The following are observed results provided by (Ismail, et. al., 2016) with respect to the correlation between emotion and the brain wave activity;

*Anger*: High amount of activity in the theta brain wave. The brain wave activity here presents the opposite of deep relaxation, which is pressure placed on an individual.

*Sadness*: The delta brain wave had a higher amount of activity than the theta brain wave. With this emotion, the authors attribute the delta brain wave activity with males who experience empathy and emotion. On the other hand, the theta brain wave is associated with sadness experienced from memories.

*Happiness*: High amount of activity in the alpha brain wave. The authors attribute this to activities related to being well rested, which increases the flow of energy, and as a result increases an individual's ability to be creative.

*Surprised*: The authors note that brain wave activity is present in most regions of the brain. However, the delta and theta brain wave show the highest amount of activity. With this emotion, the authors attribute the activity in the delta brain wave to healing. And the activity in the theta brain wave to varying levels of stress.

In a similar study by (Heraz, et. al., 2007), the authors explore the use of brain waves to classify emotional states that include anger, boredom, confusion, contempt, curious, disgust, eureka, and frustration. In this study, (Heraz, et. al., 2007) use IBK, which is also known as instance-based learning with parameter k. According to (Heraz, et. al., 2007), IBK is a k-nearest neighbour

classification algorithm. Based on the implemented classifier, classification precision on each emotional state ranges between 79.2% to 83.5%.

- Anger = 81%, Boredom = 80%, Confusion = 79.2%, Contempt = 79.8%, Curious = 81.4%, Disgust = 82.9%, Eureka = 84%, and Frustration = 82.5%.

(Heraz, et. al., 2007) also note that the performance of their implemented classifier ranges between 79.2% to 83.5%. The overall classification of all emotional states is relatively equal. Based on the confusion matrix observed by (Heraz, et. al., 2007), the values in the diagonal are greatest for each emotion predicted.

An alternative approach to EEG emotion recognition is described by (Vijayan, et, al., 2015). The authors in this study propose the use of a statistical measure known as Shannon entropy with cross correlation and auto regressive modeling. A multi-modal dataset consists of EEG signals from 32 participants. The process of the novel approach described by (Vijayan, et. al., 2015), begins with pre-processing data to remove distortion. The next step is feature extraction for classification, which consist of wavelet packets, also known as wavelet decomposition. According to (Vijayan, et. al., 2015), wavelet packets (pre-processed data) are then passed to a multi-class Support Vector Machine classifier. The classification falls into four categories of emotion; excited, happy, sad, and hate. Based on ML-SVM, (Vijayan, et. al., 2015) observed classification accuracy of 94.097%. Th authors also note that the observed accuracy level, is comparable to classification accuracy levels in other studies where the following were employed;

- Common Spatial Patterns using two emotions – 93.5%
- Linear SVM on two emotions – 93%
- K-Nearest Neighbours (KNN) on five emotions (62 channels of data) – 83.26%

- Short Time Fourier Transform (STFT) & Multi-Class SVM (MC-SVM) on three emotions – 93.85%

However, (Vijayan, et. al., 2015) conclude that the epochs (emotion characteristic data) obtained using Shannon entropy, along with auto-regression coefficients from cross correlation that were used in the Multi-Class Support Vector Machine, are a high-caliber algorithm compared to other algorithms used.

To further understand the use of machine learning algorithms in emotion recognition, (Shin, et. al., 2017) develop a complex biological emotion recognition system. This system blends two physiological indicators using ratios; ECG (Records activity of the heart) and EEG (Brain waves). The system aims to classify the following emotions; amusement, fear, sadness, joy, anger, and disgust, which is consistent with previous studies discussed. Input data (Blended ratio of ECG & EEG) are split into two streams; a training stream and a test stream. The data in the training stream is used to create a user profile in the Data Map Model List, which contains correlation information between a user's EEG and ECG activity and their emotion.

Furthermore, (Shin, et. al., 2017) explain that a weight is attached to a specific channel to improve the classification accuracy. The purpose of doing this, will accommodate how each user develops an emotion. As a result of this difference, the Data Map revaluates the EEG and ECG input, and only applies a new probability when it receives new EEG data. The new probability, which accounts for the user's highest active emotion, is used to update the probability weights on each channel on the data map. According to (Shin, et. al., 2017), this process repeats, which improves the accuracy of the classification.

The types of machine learning algorithms (Shin, et. al., 2017) use to evaluate their emotion recognition system consist of Multilayer Perceptron (MLP), Support Vector Machines (SVM), and Bayesian Networks (BN). Based on observed results, (Shin, et. al., 2017) note that when EEG is only considered, classification accuracy for each of the algorithms performed poorly; MLP (44.86%), SVM (24.99), and BN (62.28). However, when the complex bio-signal system (ECG & EEG) are considered, the classification accuracy performs significantly better; MLP (83.97%), SVM (63.97%), and BN (98.06%). The confusion matrix for MLP and SVM, had a diagonal with the highest values, and some outliers. The confusion matrix for BN has a similar diagonal with extremely minimal outliers. The observed outcomes are consistent with (Wang, et. al., 2018) and (Zheng, et. al., 2019), in that higher accuracy classification rates are seen in multimodal systems, that use more than one physiological indicator for emotion recognition.

In the literature reviewed, the DEAP dataset (Koelstra, S., et al., 2012) which is the dataset for emotion analysis using EEG, physiological and video signals, was used by (Zhang, et. al., 2017) and (Vijayan, et. al., 2015). This dataset provides a wealth of information with respect to physiological signals and emotional responses by participants. Furthermore, the DEAP dataset (Koelstra, S., et al., 2012) will be examined to determine how it can be used to train a deep neural network. The dataset consists of;

1. 14 – 16 volunteers who rated one-minute extracts of 120 music video recordings on arousal, valence and dominance, using an online-self assessment platform.
2. Of the 120 one-minute music videos, 32 participants watched and rated a subset of 40 videos. Each participant had their EEG and physiological signals recorded. Of the 32 participants, 22 participants had their face reaction recorded.

Furthermore, the DEAP dataset (Koelstra, S., et al., 2012) includes the following files described from the DEAP website. The files illustrate information that is valuable in this research;

1. *Online ratings* – Individual ratings form online self-assessment.
2. *Video list* – Names and links of YouTube videos from online self assessment, experiment and stats of individuals who rated the videos using the online self-assessment tool.
3. *Participant ratings* – Video ratings by participants.
4. *Participant questionnaires* – Questionnaire answers that participants provided before the experiment.
5. *Face video recordings* – Front facial records of 22 participants.
6. *Original data* – Physiological data that was unprocessed from the experiment (BioSemi.bdf format).
7. *Pre-processed data* – Preprocessed data that includes; down-sampling, EOG removal, filtering, segmenting etc. Physiological recordings are available in MATLAB and Python-NumPy formats.

Comparing various prediction algorithms and their application in emotion recognition is crucial. The reason for this is that it provides a fundamental understanding of the types of scenarios that a prediction algorithm might be best suited for. Furthermore, the comparative analysis between unimodal and multimodal emotion recognition system, provides insight into the relationship to prediction accuracy. Based on the literature reviewed, Deep Neural Networks, and Bayesian Networks offer promising results with respect to prediction algorithm accuracy. The studies by (Zhang, et. al., 2017), (Eskofier B. M. et al., 2016), and (Zheng, et. al., 2019), show that the prediction accuracy of using a Deep Neural Network range between 73.06% to 92.9%.

With respect to Bayesian Networks, (Shin, et. al., 2017) observe a prediction accuracy rate of 98.06%. The prediction accuracy rates described above, are dependent on the number of physiological signals and the type of feature enhancers used. As reviewed in the literature, a multimodal physiological framework provides a more robust system, which has a positive effect on prediction accuracy. The focus on using a Deep Neural Network relates to its prediction performance when using a large dataset, and its ability to learn using high-level data. This is beneficial when using the DEAP dataset (Koelstra, S., et al., 2012).

## *CHAPTER THREE*

## *3. METHODOLOGY*

As mentioned in Chapter 1, this research concentrates on the critical section depicted in Fig 4. The methodology describes the process of developing and testing the deep neural network and convolutional neural network. The use of a deep neural network is attributable to the strong prediction rates observed by (Zhang, et. al., 2017), (Eskofier B. M. et al., 2016), and (Zheng, et. al., 2019), its effectives when using a large dataset, and its ability to learn using high-level data. The first part of the methodology discuses the dataset used. Thereafter, the following sections – processing data and research design will discuss the remaining parts of the methodology.

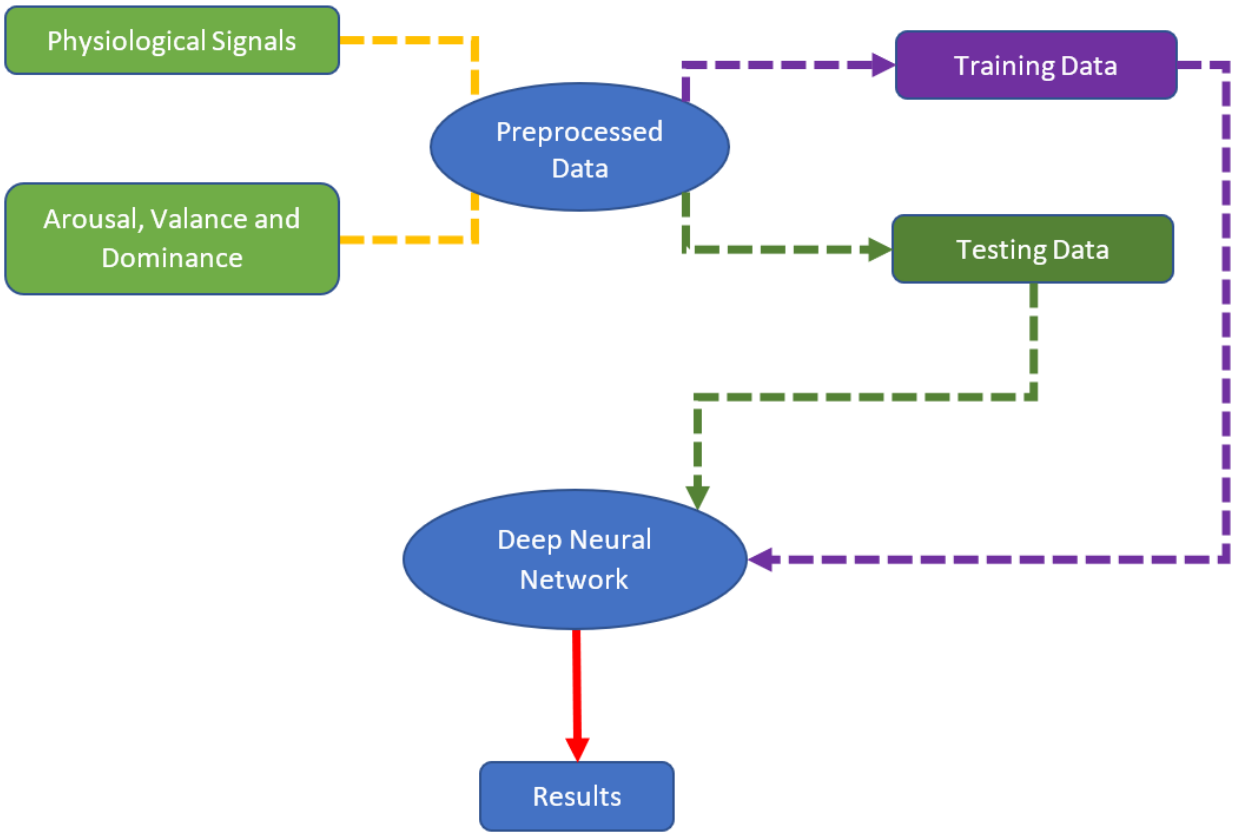


**Figure 4.** Critical section explored in this research.

### *3.1 Data Source*

The DEAP dataset (Koelstra, S., et al., 2012) was used as the data source to train the neural networks. This dataset was complied by researchers from the following academic institutions;

- Queen Mary University of London – United Kingdom
- University of Twente – Netherlands
- University of Geneva – Switzerland
- EPLF (École polytechnique fédérale de Lausanne) – Switzerland

Access to the dataset was granted once an end user license agreement was submitted by an established senior research faculty member who supervised this research. The use of the dataset is only permitted for academic research use. Each of the 32 participants also indicated their consent to redistribute their physiological recordings, and whether their audio-visual recordings may be published. The physiological indicators used from this dataset are heart rate (plethysmograph), respiration, eye movement, and brain waves by real-life participants. Emotional responses from videos watched by real-life participants were also used.

The linkage between the physiological data and emotional responses from these participants will prove crucial in this research. While participants watched a 1-minute video clip, their physiological readings were recorded. Each video provides 8,064 data samples which have been down sampled to 128Hz. A summary of the DEAP (Koelstra, S., et al., 2012) data files is seen in Table 2.

**Table 2. Data Summary**

| Title | Description |
|---|---|
| Number of participants | 32 |
| Total number of physiological signal channels | 40 (32 – EEG and 12 – Remaining Physiological channels) |
| Number of physiological channels used as inputs | 36 (32 – EEG, 2 – EOG, 1 – Respiration, 1 – Plethysmograph) |
| Sample rate of original signals | 512Hz |
| Down sampled rate of preprocessed signals | 128Hz |
| Arousal range | 1 – 9 |
| Valence range | 1 – 9 |
| Dominance range | 1 – 9 |
| Length of each channel and video recording | 63 seconds |
| Data samples from channel recording | 8,064 |
| Array shape of participant data file | Data – 40 (video/trial) x 40 (channel) x 8064 (data). Labels – 40 (video/trial) x 4 (labels – valence, arousal, dominance and liking). |
| Array shape of participant data file used | Data – 40 (video/trial) x 36 (channel) x 8064 (data). Labels – 40 (video/trial) x 3 (labels – valence, arousal, dominance). |

### *3.2 Processing Data*

As the DEAP dataset (Koelstra, S., et al., 2012) offers pre-processed data, it provides an opportunity to expedite the use of the dataset for regression without processing all the data first. The preprocessed and segmented version of the data files has been down sampled from 512Hz to 128Hz. To get a sense of the physiological readings, the data file for participant number 6 was used to depict readings for respiration, plethysmograph, horizontal and vertical eye movement, and brain waves. The x-axis represents samples taken at 128Hz. Each of the physiological recordings provide 63 seconds of measurements, which produce 8,064 data samples. In Fig 5. the participant was fitted with a respiration belt that monitored the participants lung capacity. The blue wave represents tidal volume, and the area under the blue line is functional residual capacity.

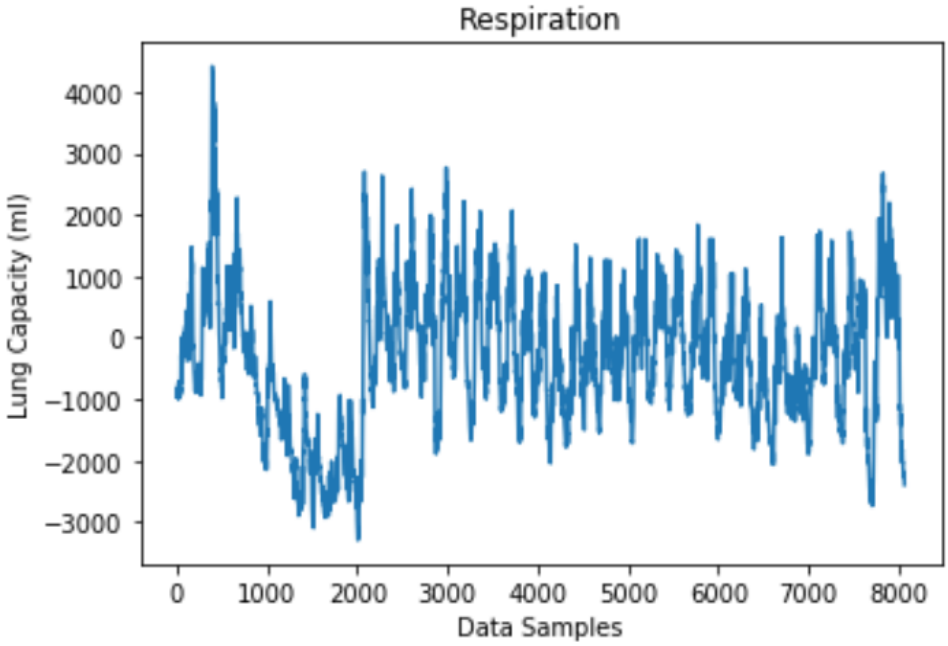


**Figure 5.** Respiration readings over 63 seconds of participant number 6.

Heart rate and cardiac cycle is measured using finger plethysmography. This is typically accomplished by using a pulse oximeter, which can monitor fluctuations in blood volume through pressure on the finger.  Participant number 6 has plethysmography readings depicted in in Fig 6.

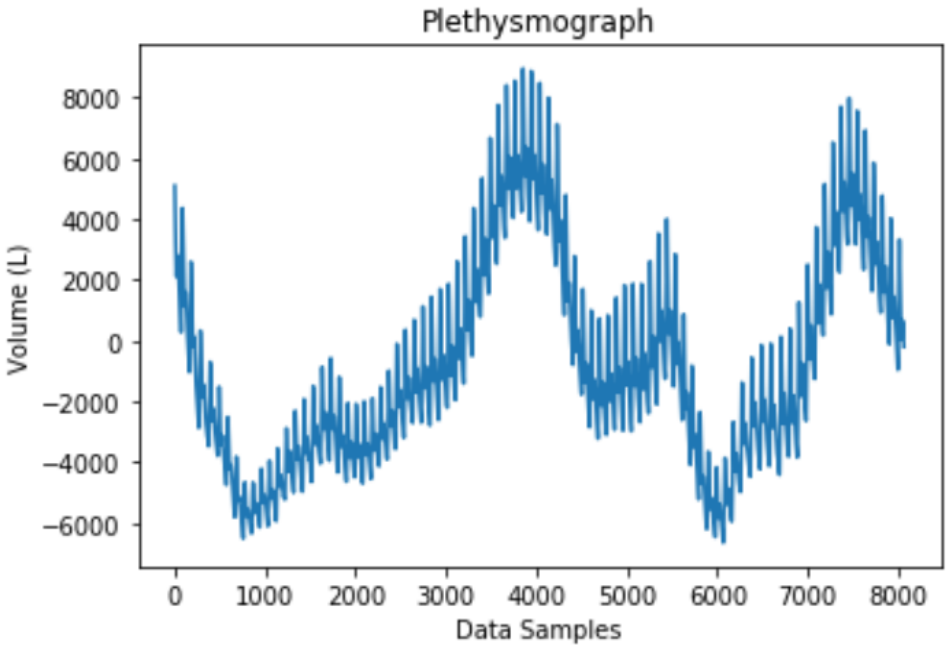


**Figure 6.** Plethysmograph readings over 63 seconds of participant number 6.

Both horizontal and vertical eye movements were recorded for 22 out of the 32 participants. This is accomplished by using electrodes that are placed around the eye. To record horizontal eye movement, electrodes are placed on both the right and left temple. For vertical eye movement, electrodes are placed above and below the eye to record horizontal eye movements. This setup enables eye tracking for visual objects. Horizontal and vertical eye movements recordings for participant number 6 are seen in Fig 7 and Fig. 8.

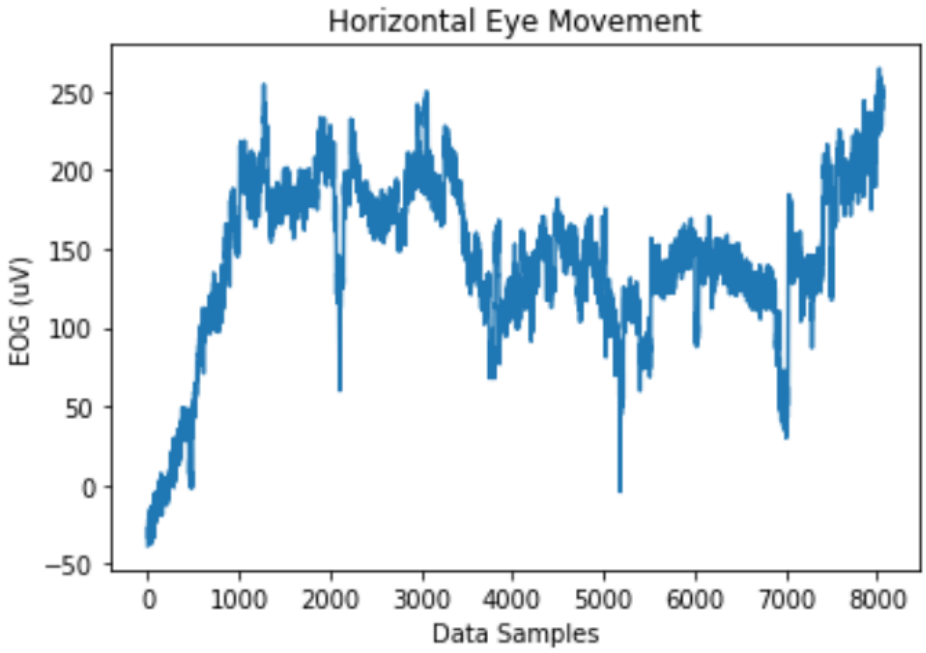


**Figure 7.** Horizontal eye movement readings over 63 seconds of participant number 6.

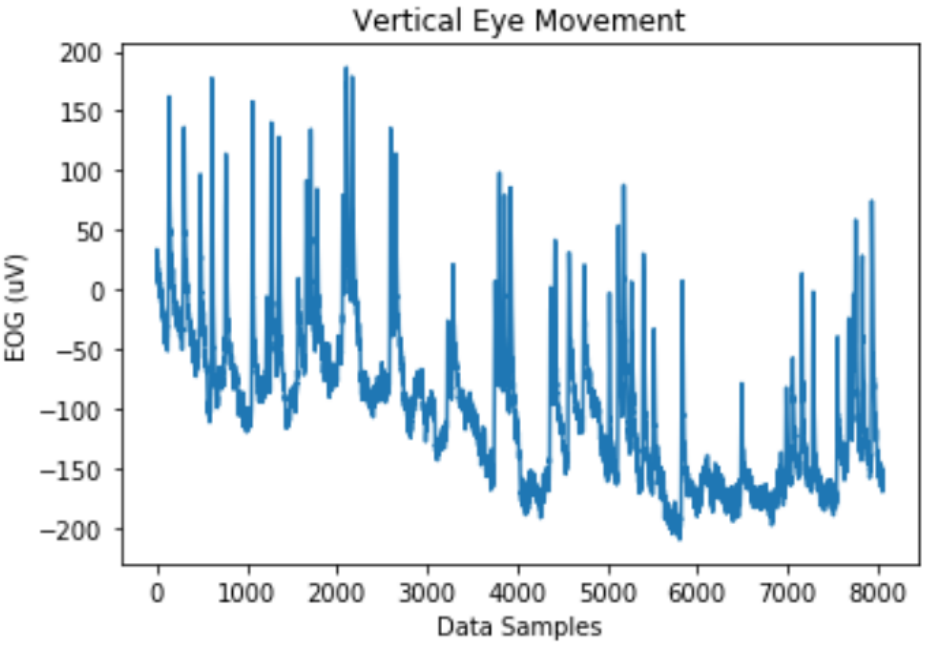


**Figure 8.** Vertical eye movement readings over 63 seconds of participant number 6.

Each of the 32 participants had their brain waves recorded over 63 seconds. A total of 32 EEG channels were used to record delta, theta, alpha, and beta brain waves. The data from the recordings were down sampled to 128 Hz. Brain wave recordings for participant number 6 is seen in Fig 9.

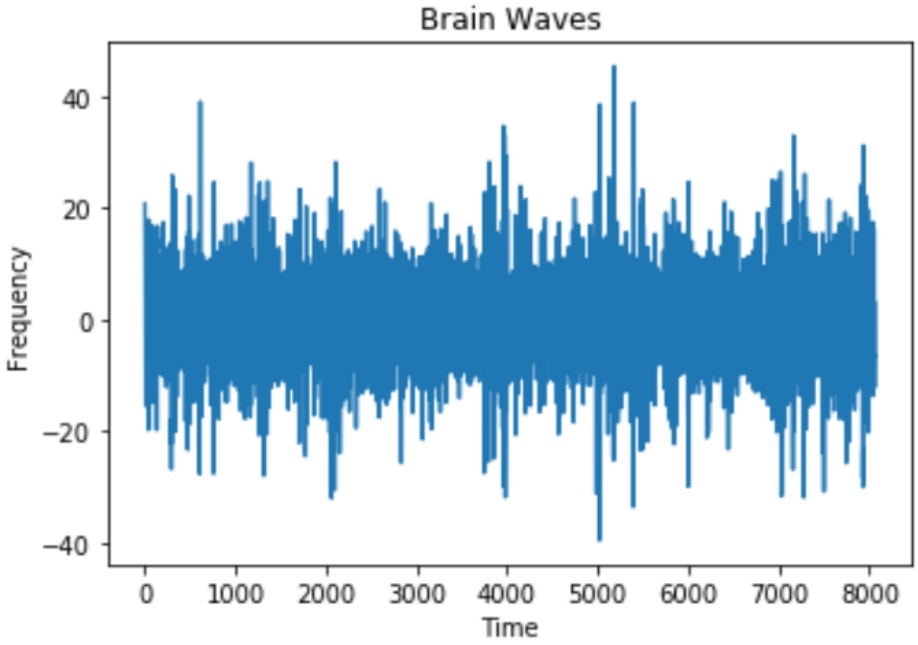


**Figure 9.** Brain wave readings from 32 channels over 63 seconds of participant number 6.

### *3.3 Research Design*

The research design section illustrates steps taken to execute the development and testing of both the deep neural network and convolutional neural network. The remaining part of this section will discuss the architecture of the deep neural network and convolutional neural network, and justifications behind the architecture design choices. Summary of steps taken to execute the research design;

1. Selected features that were used for prediction.
2. Concatenated all 32 participant data files.

3. Developed a deep neural network and convolutional neural network to compare prediction performance.
4. Trained the deep neural network and a convolutional neural network to determine prediction accuracy at;
    a. 25 Epochs
    b. 50 Epochs
    c. 100 Epochs

In the first step, features selected include physiological indicators – heart rate, respiration, eye movement, and brain waves. And approximate grouping of emotional states – arousal, valence and dominance**.** Based on the number of features selected, a total of 36 physiological recordings, 3 approximate grouping of emotional states (arousal, valance and dominance), and 8,064 data points were used as inputs into the deep neural network and convolutional neural network.

The second step combined all 32 participant data files. This was done to enable both neural networks to generalize predictions on arousal, valence and dominance. Up till now, arousal and valence are understood to exist on the 2-dimensional circumplex theory of emotion seen in Fig 10. However, dominance requires an additional dimension, which is illustrated in the PAD 3-dimentional emotion representation seen in Fig 11.

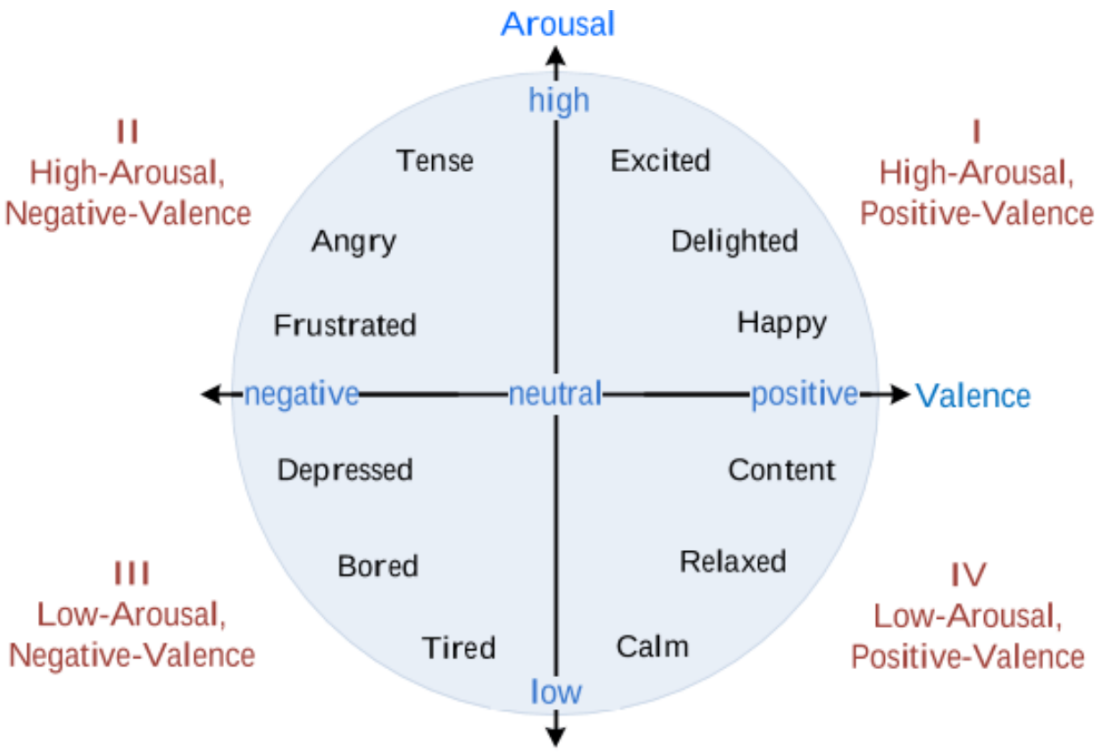


**Figure 10.** Circumplex theory of emotions 2-dimension model (Liu, et. al., 2018). The 2-dimensional representation does not depict dominance.

Dominance in the 3-dimensional model is represented by the D(Z)-axis. This axis represents how dominant (positive values) or submissive (negative values) an emotion is. For example, if a participant experiences high arousal, low pleasure (valance) and high dominance, the individual may be feeling hostile. This feeling is a dominant feeling compared to feeling anxious, which is a submissive feeling seen in Fig 11.

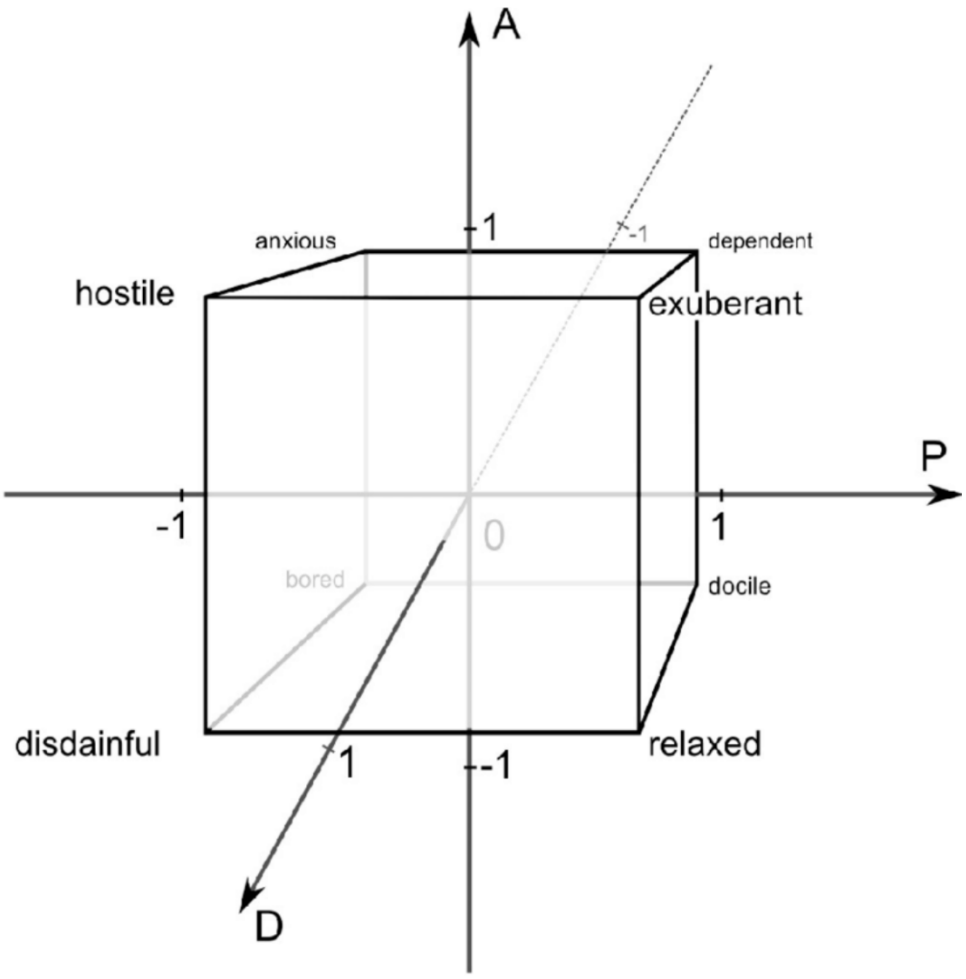


**Figure 11.** PAD (Pleasure/Valence Arousal Dominance) emotional 3-dimension representation model. (Kołakowska, A., et. al., 2015).

The third step of the research design involved developing a deep neural network and a convolutional neural network. The implementation of both the DNN and CNN were created using TensorFlow with Keras in Python. The deep neural network developed used 8 hidden layers seen in Fig 12. The first 7 layers used a ReLU (Rectified Linear Unit) activation function and the 8$^{th}$ layer used a linear function. The use of the ReLU function primarily performs best with prediction problems.

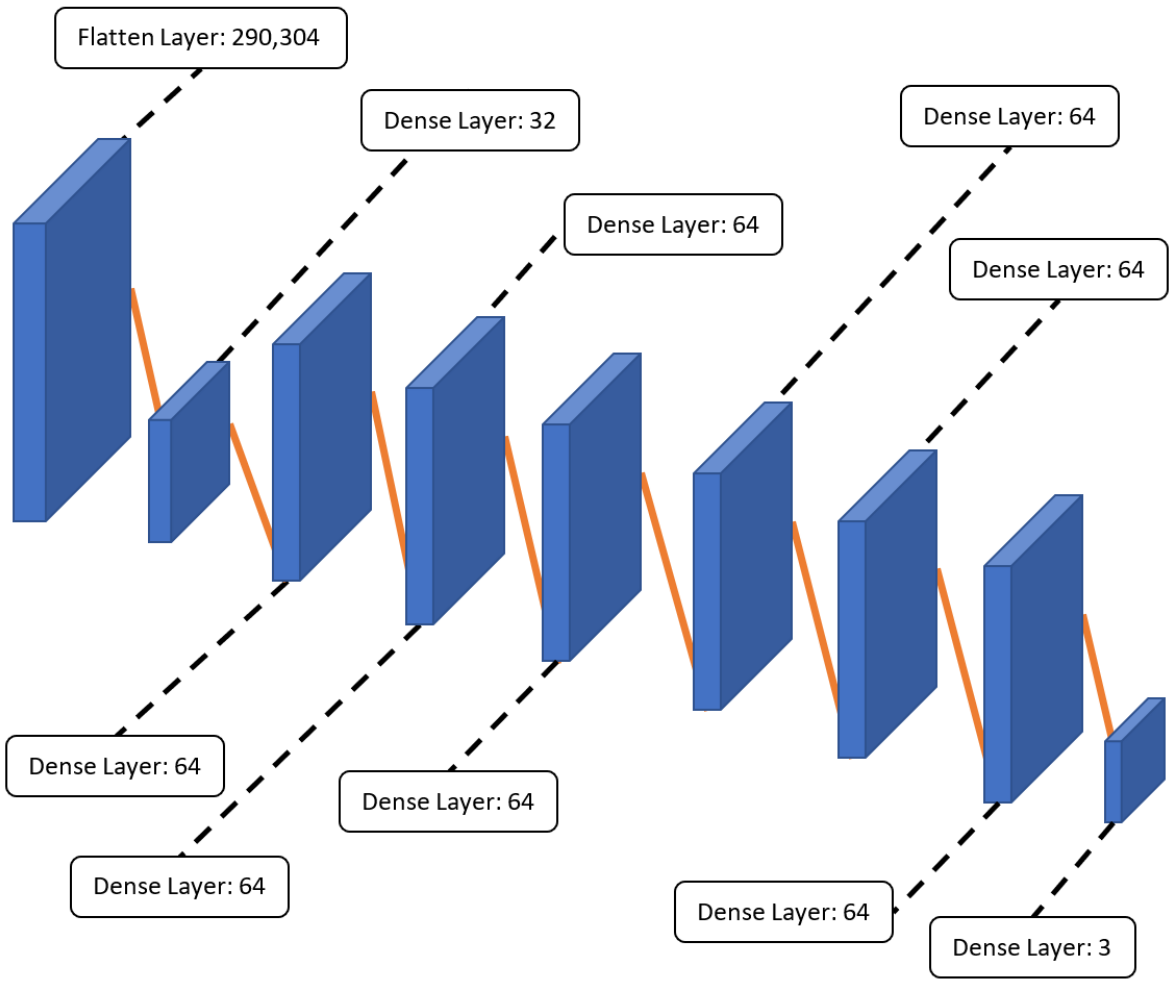


**Figure 12.** Deep Neural Network Architecture.

Each neuron in the deep neural network is connected to every output from the previous layer. The flatten layer initially flattens all the data into a single dimensional array of numbers. Since there are 36 physiological channels with 8,064 samples each, it provides 298,368 (36 * 8,064) data points as seen in Table 3. Each of the 298,368 has 32 connections, one for each neuron in the next layer. In the first dense layer there are 9,289,728 (32 * 290,304) parameters. The number of parameters reduces at the second layer as the filter capacity is set at 64.

The number of parameters is constant from the 3rd layer till the 7th layer as the filter capacity is 64. On the last layer, the filter capacity is reduced to 3 to account for the three labels (valence,

arousal and dominance. This reduces the number of parameters to 195 (3 * 64). A batch size of 10 was used during model compilation. This would randomly process 10 training samples at a time until a total of 896 samples were processed on each training epoch. The use of batch size minimized error values and variance between loss and validation loss functions.

**Table 3. Deep Neural Network Model Architecture**

| Layer (type) | Output | Parameters | Activation |
|---|---|---|---|
| **Flatten** | 290,304 | 0 | - |
| **Dense** | 32 | 9,289,728 | ReLU |
| **Dense** | 64 | 2,048 | ReLU |
| **Dense** | 64 | 4,096 | ReLU |
| **Dense** | 64 | 4,096 | ReLU |
| **Dense** | 64 | 4,096 | ReLU |
| **Dense** | 64 | 4,096 | ReLU |
| **Dense** | 64 | 4,096 | ReLU |
| **Dense** | 3 | 192 | Linear |

A convolutional neural network was chosen as a comparison to the deep neural network as it has been identified by (Yang, H., et al., 2019) as a strong contender in predicting valence and arousal using pattern recognition. The dataset that was used to in the study by (Yang, H., et al., 2019) also used the DEAP dataset (Koelstra, S., et al., 2012). The CNN developed in this research features 5 convolutional layers, 4 max pooling layers and 2 fully connected dense output layers. Similar to the deep neural network, a batch size of 10 training samples was set during model

compilation. The convolutional neural network does not use a multi-column structured model as presented in the study by (Yang, H., et al., 2019).

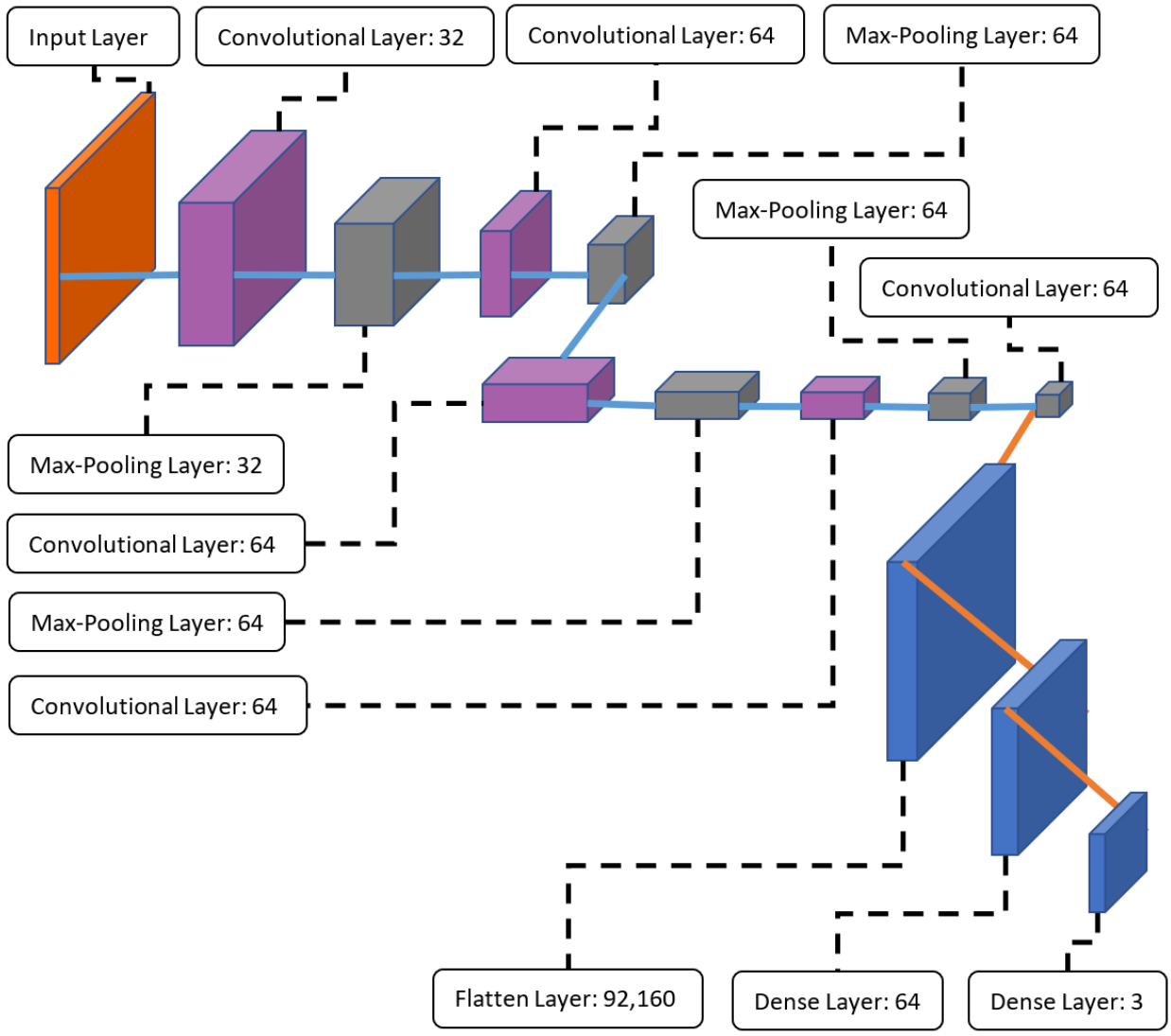


**Figure 13.** Convolutional Neural Network Architecture.

The filter in the first convolutional layer (32) is applied to the input (36 x 8064, along with the activation function) as the filter is moved along. The stride determines how large of a movement the filter is moved along by. The max-pooling layer reduces the size of the input. This process

continues until the flatten layer. The flatten layer converts the pooled layer representation into a column that is passed to the dense layer.

**Table 4. Convolutional Neural Network Model Architecture**

| Layer (type) | Kernel/Pool | Stride | Outputs | Activation |
|---|---|---|---|---|
| **Reshape** | - | - | 36 x 8064 x 1 | - |
| **Convolution** | 1 x 4 | 1 x 1 | 36 x 8061 x 32 | ReLU |
| **Max Pooling** | 1 x 2 | 1 x 2 | 36 x 4030 x 32 | - |
| **Convolution** | 1 x 8 | 1 x 2 | 36 x 2012 x 64 | ReLU |
| **Max Pooling** | 1 x 2 | 1 x 2 | 36 x 1006 x 64 | - |
| **Convolution** | 1 x 8 | 1 x 2 | 36 x 500 x 64 | ReLU |
| **Max Pooling** | 1 x 2 | 1 x 2 | 36 x 250 x 64 | - |
| **Convolution** | 1 x 64 | 1 x 2 | 36 x 94 x 64 | ReLU |
| **Max Pooling** | 1 x 2 | 1 x 2 | 36 x 47 x 64 | - |
| **Convolution** | 1 x 8 | 1 x 1 | 36 x 40 x 64 | ReLU |
| **Flatten** | - | - | 92,160 | - |
| **Dense** | - | - | 64 | ReLU |
| **Dense** | - | - | 3 | Linear |

The fourth step of the experiment involved training and testing the deep neural network and the convolutional neural network at 25, 50 and 100 epochs. The length of the epochs was selected to compare prediction accuracy over time and to analyze convergence and divergence of loss and validation loss functions. Based on the prediction results and the MAE (Mean Absolute Error), the ratio for training and validating that minimized the MAE, was set at 70% for training (896

samples), and 30% for validation (384 samples) for a total of 1,280 samples (32 participants * 40 videos). This was applied to both the deep neural network and convolutional neural network.

To analyze prediction accuracy, the following metrics were used;

- *Loss* – The level of variation from actual values.
- *Mean Absolute Error* – Mean of the absolute values of prediction errors.
- *Validation Loss* – Error value reported after running validation dataset against training dataset.
- *Validation Mean Absolute Error* – Mean of the absolute values of prediction errors based on the validation dataset.
- *Difference between predicted and real values* – How close predicted values are from actual values with respect to arousal, valence and dominance.
- *Bounded Regression Accuracy* – Determines prediction accuracy as a percentage of the range of possible y-values. The formula is given by;

$$Bounded\ Regression\ Accuracy = \frac{1 - |\ Prediction\ Value - Real\ Value\ |}{Range\ Value * 100\%}$$

## *CHAPTER FOUR*

## *4. RESULTS AND ANALYSIS*

This section will present findings from compiling the deep neural network and convolutional neural network. It will discuss prediction performance on arousal, valence and dominance. Three training time intervals were used – 25, 50 and 100 epochs, to compare the prediction performance of the deep neural network against the convolutional neural network.

### 4.1 CNN & DNN Training – 25 Epochs

The convolutional neural network trained at 25 epochs initially shows the loss and validation loss functions converging at approximately 2 epochs seen in Fig 14. This convergence is not consistent as the loss and validation loss functions start to diverge. Based on the divergence of the loss and validation loss functions, it can be inferred that the convolutional neural network requires more data to train the model. It also requires more than 25 epochs to narrow the gap between the loss and validation loss functions.

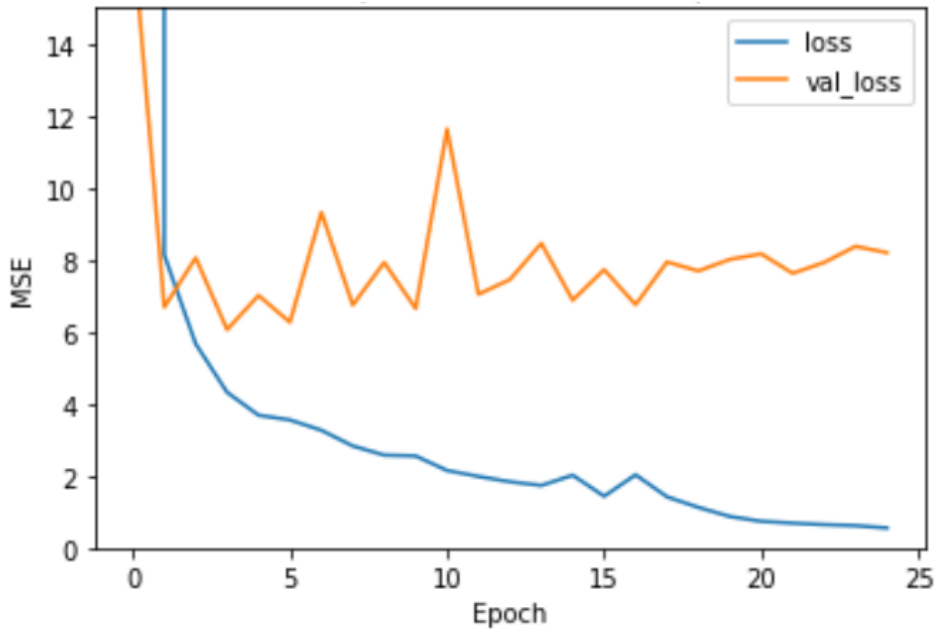

**Figure 14.** Convolutional Neural Network trained for 25 epochs.

Table 5. summarizes performance metric values for the convolutional neural network at 5, 10, 15 and 25 epochs. Values in the loss function reduce while values in the validation loss function increase. The difference between the loss function and validation loss function at 25 epochs is 7.6498. While the mean absolute error values reduce with each epoch up till 25 epochs, the significantly large difference between the loss and validation loss functions can be explained with overfitting.

**Table 5. CNN Performance Metrics – 25 Epochs**

| Epoch | Loss | MAE | Validation Loss | Validation MAE |
|---|---|---|---|---|
| **5** | 3.6841 | 1.5663 | 7.0124 | 1.8905 |
| **10** | 2.5524 | 1.1897 | 6.6501 | 1.9625 |
| **15** | 2.0178 | 0.9673 | 6.8810 | 1.9440 |
| **25** | 0.5502 | 0.4887 | 8.2000 | 1.9631 |

Prediction accuracy as a percentage is seen in Fig. 15. Each set of valence, arousal and dominance prediction represents an approximate group of emotional states felt by the participant while watching a video for 63 seconds. While overfitting is evident at 25 epochs, the overall mean bounded regression accuracy for valance arousal and dominance is 74.13%, 75.30% and 76.96% respectively. The overall mean bounded regression accuracy for the model at 25 epochs is 75.46%. This can be explained by the decline in the mean absolute error values with every epoch cycle. It is important to note that given overfitting is evident with 25 epochs, the model needs more data to train on to reduce the likelihood of overfitting.

```
                Pred | Real | Diff | Acc
============================================
Valence         5.50 | 5.01 | 0.49 | 93.9%
Arousal         5.84 | 1.03 | 4.81 | 39.8%
Dominance       6.14 | 6.04 | 0.10 | 98.8%
============================================
Valence         4.54 | 1.67 | 2.87 | 64.2%
Arousal         6.28 | 5.21 | 1.07 | 86.6%
Dominance       4.94 | 1.92 | 3.02 | 62.2%
============================================
Valence         5.24 | 2.99 | 2.25 | 71.9%
Arousal         5.40 | 3.03 | 2.37 | 70.4%
Dominance       5.22 | 1.99 | 3.23 | 59.6%
============================================
Valence         2.80 | 3.88 | 1.08 | 86.5%
Arousal         3.78 | 2.14 | 1.64 | 79.5%
Dominance       2.34 | 1.42 | 0.92 | 88.6%
============================================
Valence         4.73 | 5.09 | 0.36 | 95.5%
Arousal         4.45 | 8.04 | 3.59 | 55.1%
Dominance       4.13 | 8.05 | 3.92 | 51.0%


Mean valence bounded regression accuracy: 74.13%
Mean arousal bounded regression accuracy: 75.30%
Mean dominance bounded regression accuracy: 76.96%
Overall mean bounded regression accuracy: 75.46%
```

**Figure 15.** CNN prediction accuracy of valance, arousal and dominance at 25 epochs. Valence, arousal and dominance value range from 1 – 9.

In comparison, the deep neural network indicates a momentary convergence between the loss and validation loss functions at approximately 3 epochs seen in Fig 16. This convergence does not stay consistent and overfitting occurs immediately. While the difference between the loss and validation functions are significantly large, both functions show a downward trend after 15 epochs. Similar to the convolutional neural network, more data and training at 25 epochs would alleviate overfitting seen in the deep neural network.

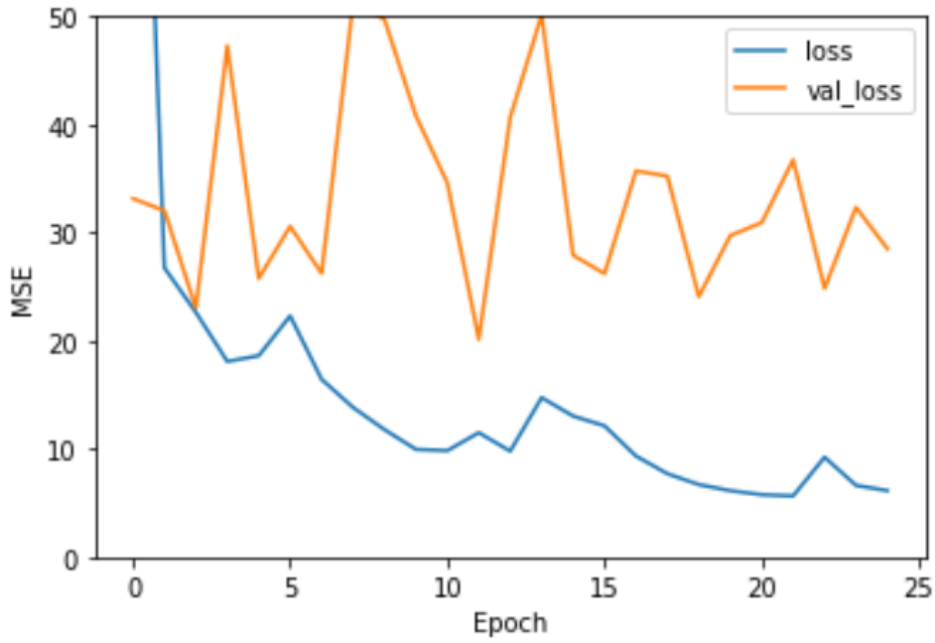


**Figure 16.** Deep Neural Network trained for 25 epochs.

The values depicted in Table 6. are consistent with the graph in Fig 16. The loss and validation loss function show significant swings upwards and downwards from epochs 5 to 15. The values for the loss and validation loss function start a downward trend after 15 epochs. The mean absolute error steadily declines through each epoch cycle. In comparison to the convolutional neural network, the values are significantly higher for the deep neural network for each epoch cycle up till 25 epochs.

**Table 6. DNN Performance Metrics – 25 Epochs**

| Epoch | Loss | MAE | Validation Loss | Validation MAE |
|---|---|---|---|---|
| **5** | 18.6222 | 3.3918 | 25.7804 | 3.7487 |
| **10** | 9.9603 | 2.5228 | 40.8092 | 3.4722 |
| **15** | 13.0411 | 2.8688 | 27.9236 | 3.4605 |
| **25** | 6.1428 | 1.8942 | 28.5146 | 3.0575 |

In terms of prediction accuracy, the deep convolutional neural network does not perform as well as the convolutional neural network at 25 epochs. The mean bounded regression accuracy for valance, arousal and dominance is 59.69%, 62.79% and 62.86% respectively. In the convolutional neural network at 25 epochs, the mean bounded regression accuracy for valance arousal and dominance is 74.13%, 75.30% and 76.96%. The overall mean bounded regression accuracy for the deep neural network model is 61.78% which does not perform as well as the convolutional neural network at which has an overall mean bounded accuracy of 75.46%.

```
                Pred | Real | Diff | Acc
============================================
Valence         6.48 | 7.06 | 0.58 | 92.8%
Arousal         6.74 | 5.96 | 0.78 | 90.3%
Dominance       6.57 | 6.13 | 0.44 | 94.5%
============================================
Valence         2.54 | 8.24 | 5.70 | 28.7%
Arousal         2.94 | 5.08 | 2.14 | 73.3%
Dominance       3.10 | 6.01 | 2.91 | 63.7%
============================================
Valence         10.25 | 4.00 | 6.25 | 21.9%
Arousal         9.55 | 1.00 | 8.55 | -6.9%
Dominance       10.70 | 9.00 | 1.70 | 78.7%
============================================
Valence         5.83 | 1.86 | 3.97 | 50.4%
Arousal         5.73 | 2.04 | 3.69 | 53.9%
Dominance       6.61 | 1.00 | 5.61 | 29.9%
============================================
Valence         3.04 | 4.96 | 1.92 | 76.0%
Arousal         3.27 | 1.99 | 1.28 | 84.0%
Dominance       3.49 | 4.08 | 0.59 | 92.6%


Mean valence bounded regression accuracy: 59.69%
Mean arousal bounded regression accuracy: 62.79%
Mean dominance bounded regression accuracy: 62.86%
Overall mean bounded regression accuracy: 61.78%
```

**Figure 17.** DNN prediction accuracy of valance, arousal and dominance at 25 epochs. Valence, arousal and dominance value range from 1 – 9.

### 4.2 CNN & DNN Training – 50 Epochs

At 50 epochs the loss and validation loss functions of the convolutional neural network converges just after epoch 0 seen in Fig 18., which is slightly earlier than the 25-epoch scenario. However, both the loss and validation loss functions diverge immediately after the initial convergence. Similar to the convolutional neural network trained for 25 epochs, the model experiences overfitting. Even as loss and validation loss functions approach the 50-epoch cycle, there is no clear indication that overfitting converges.

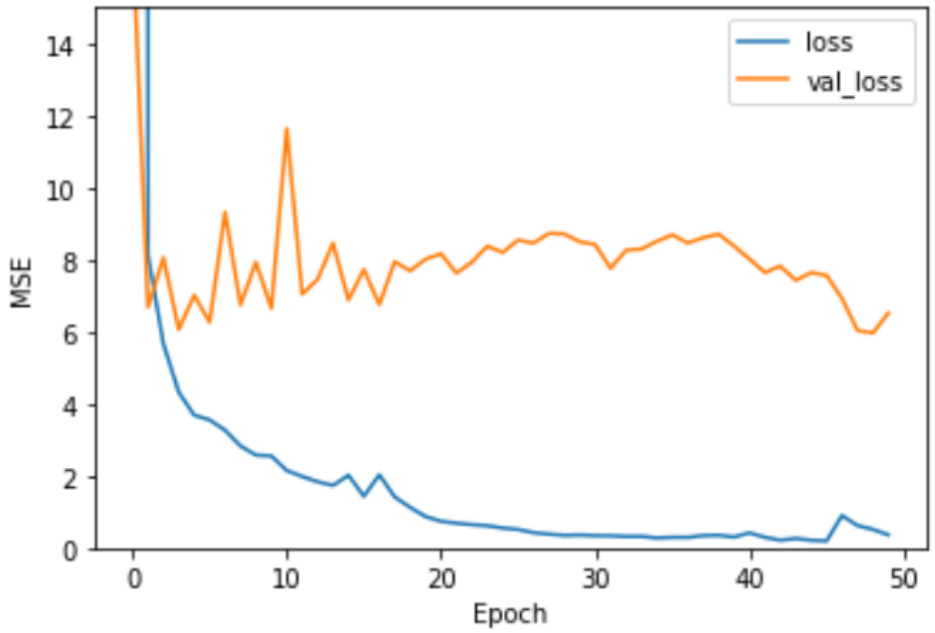


**Figure 18.** Convolutional Neural Network trained for 50 epochs.

The performance metric values for the convolutional neural network trained for 50 epochs is seen in Table 7. The values in the loss column decline at approximately epoch 30 but increase after approximately 35 epochs. However, these values are comparatively lower than the loss values in the convolutional neural network trained for 25 epochs. This is explained by the increase in training from 25 to 50 epochs. Similar to the loss values, the validation loss function experienced fluctuations in its values, which corresponds to the validation loss function trend seen in Fig. 18.

**Table 7. CNN Performance Metrics – 50 Epochs**

| Epoch | Loss | MAE | Validation Loss | Validation MAE |
|---|---|---|---|---|
| **30** | 0.3579 | 0.3592 | 8.4978 | 1.9588 |
| **35** | 0.2736 | 0.2955 | 8.5119 | 1.9528 |
| **40** | 0.3003 | 0.3452 | 8.3764 | 1.9448 |
| **50** | 0.3612 | 0.3777 | 6.5174 | 1.8795 |

The overall bounded regression accuracy for valence, arousal, and dominance is 75.43%, 76.67%, and 77.42% respectively. The overall mean bounded regression accuracy for the model trained for 50 epochs is 76.51%.

```
                Pred | Real | Diff | Acc
=============================================
Valence         5.40 | 5.01 | 0.39 | 95.2%
Arousal         6.19 | 1.03 | 5.16 | 35.5%
Dominance       5.99 | 6.04 | 0.05 | 99.4%
=============================================
Valence         4.80 | 1.67 | 3.13 | 60.9%
Arousal         6.26 | 5.21 | 1.05 | 86.9%
Dominance       4.64 | 1.92 | 2.72 | 66.0%
=============================================
Valence         5.53 | 2.99 | 2.54 | 68.2%
Arousal         5.69 | 3.03 | 2.66 | 66.8%
Dominance       6.05 | 1.99 | 4.06 | 49.3%
=============================================
Valence         2.92 | 3.88 | 0.96 | 88.0%
Arousal         4.71 | 2.14 | 2.57 | 67.8%
Dominance       4.16 | 1.42 | 2.74 | 65.8%
=============================================
Valence         4.89 | 5.09 | 0.20 | 97.5%
Arousal         4.75 | 8.04 | 3.29 | 58.9%
Dominance       4.53 | 8.05 | 3.52 | 56.0%


Mean valence bounded regression accuracy: 75.43%
Mean arousal bounded regression accuracy: 76.67%
Mean dominance bounded regression accuracy: 77.42%
Overall mean bounded regression accuracy: 76.51%
```

**Figure 19.** CNN prediction accuracy of valance, arousal and dominance at 50 epochs. Valence, arousal and dominance values range from 1 – 9.

This represents a slight increase in performance when compared to training the convolutional neural network for 25 epochs. The slight increase in prediction accuracy when the model is trained for 50 epochs can be explained by the further reduction in mean absolute error values seen in Table 7.

The deep neural network trained for 50 epochs indicates that its first convergence of the loss and validation loss functions occur approximately 25 epochs after the convolutional neural network seen in in Fig 20. Unlike the convolutional neural network trained for 50 epochs, the loss and validation loss functions in the deep neural network trained for 50 epochs converge again at approximately 30 epochs. While overfitting exists, the variance between the loss and validation loss function is significantly less than the convolutional neural network trained for 50 epochs.

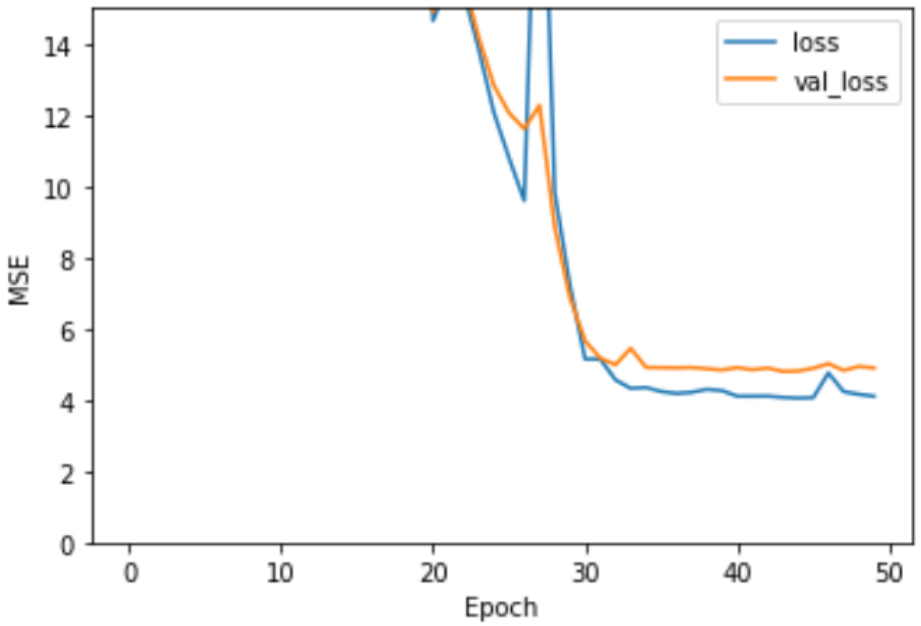


**Figure 20.** Deep Neural Network trained for 50 epochs.

The performance metrics seen in Table 8. supports the overfitting difference between the convolutional neural network and the deep neural network trained for 50 epochs. Based on the metrics, the difference in loss and validation loss at 30, 35, 40 and 50 epochs are 0.3425, 0.567, 0.5778 and 0.7966 respectively. In comparison the difference in loss and validation loss for the convolutional neural network at 30, 35, 40 and 50 epochs are 8.1399, 8.2383, 8.0761, and 6.1562 respectively. The variance in loss is wider than the deep neural network.

**Table 8. DNN Performance Metrics – 50 Epochs**

| Epoch | Loss | MAE | Validation Loss | Validation MAE |
|---|---|---|---|---|
| **30** | 7.2251 | 2.1854 | 6.8826 | 2.0860 |
| **35** | 4.3482 | 1.7104 | 4.9152 | 1.8268 |
| **40** | 4.2627 | 1.6913 | 4.8405 | 1.8146 |
| **50** | 4.1026 | 1.6648 | 4.8992 | 1.8300 |

The prediction accuracy of the deep neural network is seen in Fig 21. The bounded regression accuracy for valance, arousal and dominance is 77.22%, 77.36%, and 76.79% respectively. The overall mean bounded regression accuracy for the deep neural network trained for 50 epochs is 77.12%. In comparison the bounded regression accuracy for valance, arousal and dominance for the convolutional neural network trained for 50 epochs is 75.43%, 76.67%, 77.42%. Dominance is the only category that performed slightly better in the convolutional neural network trained at 50 epochs. The overall mean bounded regression accuracy for the convolutional neural network is 76.51%, which doesn't perform as well at the deep neural network. It is also worthy to note that the variance in loss with the deep-neural network is significantly less than the convolutional neural network. For these reasons, the deep neural network performs better than the convolutional neural network trained for 50 epochs.

```
                Pred | Real | Diff | Acc
============================================
Valence         5.19 | 7.09 | 1.90 | 76.2%
Arousal         5.09 | 6.42 | 1.33 | 83.4%
Dominance       5.14 | 3.49 | 1.65 | 79.4%
============================================
Valence         5.19 | 2.32 | 2.87 | 64.1%
Arousal         5.09 | 2.83 | 2.26 | 71.7%
Dominance       5.14 | 2.40 | 2.74 | 65.7%
============================================
Valence         5.19 | 3.81 | 1.38 | 82.8%
Arousal         5.09 | 3.85 | 1.24 | 84.5%
Dominance       5.14 | 4.78 | 0.36 | 95.5%
============================================
Valence         5.19 | 5.06 | 0.13 | 98.4%
Arousal         5.09 | 1.05 | 4.04 | 49.5%
Dominance       5.14 | 5.04 | 0.10 | 98.7%
============================================
Valence         5.19 | 6.99 | 1.80 | 77.5%
Arousal         5.09 | 3.79 | 1.30 | 83.7%
Dominance       5.14 | 3.83 | 1.31 | 83.6%


Mean valence bounded regression accuracy: 77.22%
Mean arousal bounded regression accuracy: 77.36%
Mean dominance bounded regression accuracy: 76.79%
Overall mean bounded regression accuracy: 77.12%
```

**Figure 21.** DNN prediction accuracy of valance, arousal and dominance at 50 epochs. Valence, arousal and dominance values range from 1 – 9.

### 4.3 CNN & DNN Training – 100 Epochs

The convolutional neural network trained for 100 epochs seen in Fig 22., appears to have no significant improvement since its last training boundary at 50 epochs. Both the loss and validation functions continue to indicate a significant difference in variance up till epoch 100. This infers that any additional training does not significantly improve the model.

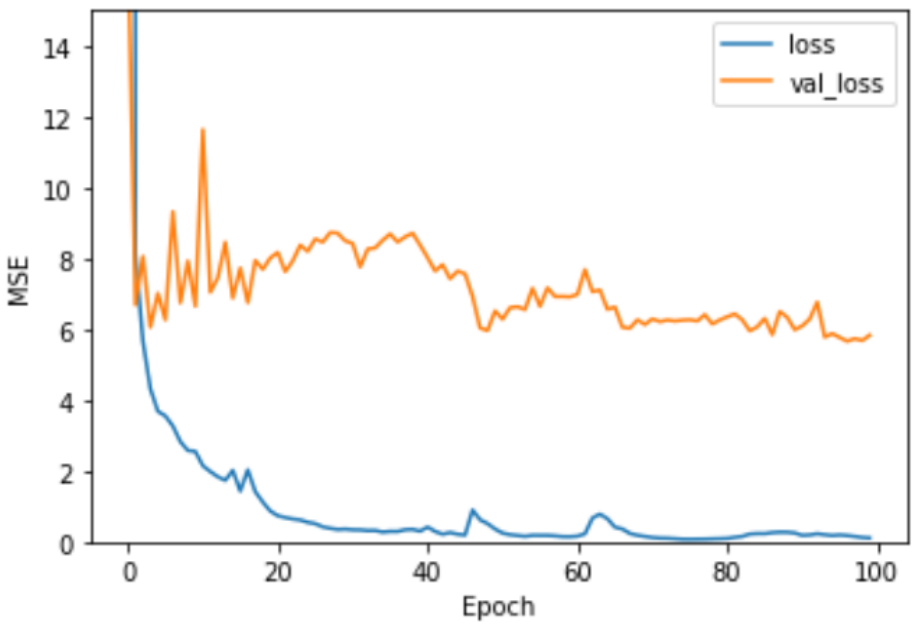


**Figure 22.** Convolutional Neural Network trained for 100 epochs.

The performance metrics seen in Table 9. Indicate a modest improvement in the convolutional neural network trained for 100 epochs compared to 50 epochs. While the mean absolute values are lower, the variance between the loss and validation loss functions are still significant. More training appears to have made a slight improvement to overfitting. It is possible that the model requires more data in order to make a significant reduction in overfitting with an increase in training epochs.

**Table 9. CNN Performance Metrics – 100 Epochs**

| Epoch | Loss | MAE | Validation Loss | Validation MAE |
|---|---|---|---|---|
| **70** | 0.1515 | 0.2580 | 6.1339 | 1.8792 |
| **80** | 0.0878 | 0.1966 | 6.2619 | 1.8676 |
| **90** | 0.2494 | 0.3566 | 5.9862 | 1.8708 |
| **100** | 0.1078 | 0.2315 | 5.8295 | 1.8774 |

The prediction performance for the convolutional neural network trained for 100 epochs is seen in Fig 23. It supports the explanation that no significant improvement was made by increasing the number of training epochs to 100. The mean regression accuracy for valance, arousal and dominance is 74.83%, 76.88%, 77.89% and 76.53% respectively when trained for 100 epochs. In comparison, the mean regression accuracy for valance, arousal and dominance is 75.43%, 76.67%, 77.42% and 76.51% respectively when trained for 50 epochs. Apart from valence, there is no significant improvement in performance for arousal and dominance.

```
                Pred | Real | Diff | Acc
============================================
Valence         5.53 | 5.01 | 0.52 | 93.5%
Arousal         5.71 | 1.03 | 4.68 | 41.4%
Dominance       5.83 | 6.04 | 0.21 | 97.4%
============================================
Valence         4.17 | 1.67 | 2.50 | 68.7%
Arousal         5.66 | 5.21 | 0.45 | 94.4%
Dominance       4.12 | 1.92 | 2.20 | 72.5%
============================================
Valence         5.40 | 2.99 | 2.41 | 69.9%
Arousal         5.34 | 3.03 | 2.31 | 71.2%
Dominance       5.06 | 1.99 | 3.07 | 61.6%
============================================
Valence         2.51 | 3.88 | 1.37 | 82.9%
Arousal         4.11 | 2.14 | 1.97 | 75.4%
Dominance       4.11 | 1.42 | 2.69 | 66.3%
============================================
Valence         5.15 | 5.09 | 0.06 | 99.2%
Arousal         4.42 | 8.04 | 3.62 | 54.7%
Dominance       4.41 | 8.05 | 3.64 | 54.4%


Mean valence bounded regression accuracy: 74.83%
Mean arousal bounded regression accuracy: 76.88%
Mean dominance bounded regression accuracy: 77.89%
Overall mean bounded regression accuracy: 76.53%
```

**Figure 23.** CNN prediction accuracy of valance, arousal and dominance at 100 epochs. Valence, arousal and dominance values range from 1 – 9.

The overall mean bounded regression accuracy between 50 and 100 epochs is 76.51% and 76.53% respectively. This also supports the explanation that an increase in training epochs from 50 to 100 does not improve the convolutional neural network.

The deep neural network trained for 100 epochs seen in Fig 24. appears to have significantly improved with an increase in training from 50 epochs. This is supported by the loss and validation function converging at approximately 20 epochs. The variance between the loss and validation function is minimal. This also infers that overfitting is minimal. In comparison, overfitting in the convolutional neural network trained for 100 epochs is significantly more and does not improve with every new training epoch.

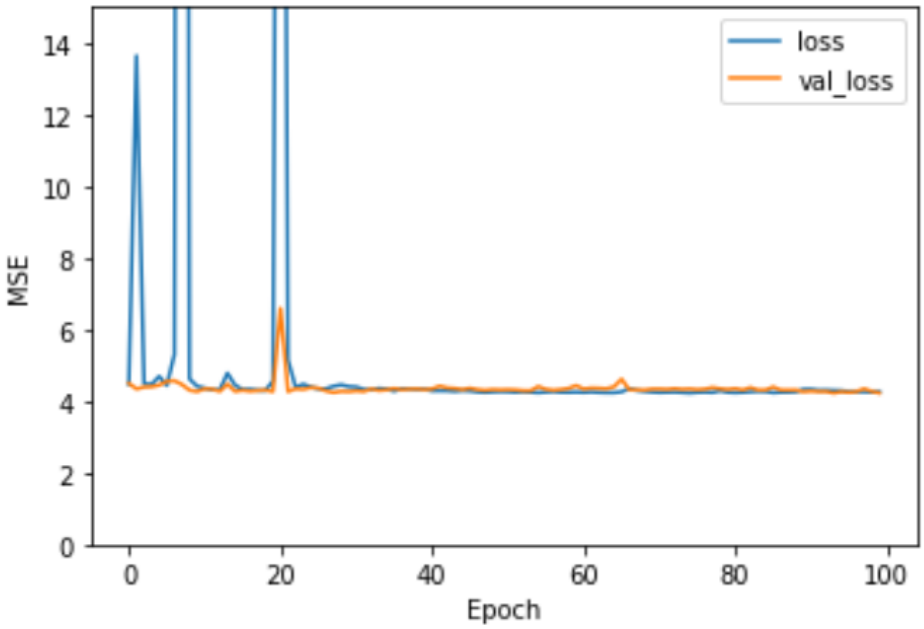


**Figure 24.** Deep Neural Network trained for 100 epochs.

The performance metrics for the deep neural network is seen in Table 10. The difference in variance for the loss and validation loss function at 70, 80, 90 and 100 epochs are 0.0832, 0.0806, 0.068, and 0.0348 respectively. This infers that over fitting and underfitting is at a minimal. In

comparison, the loss and validation loss function at 70, 80, 90 and 100 epochs for the convolutional neural network are 5.9824, 6.1741, 5.7368, and 5.7217 respectively, which are significantly higher than the deep neural network. While the mean absolute error values are higher in the deep neural network, the deep neural network performs better in in prediction accuracy.

**Table 10. DNN Performance Metrics – 100 Epochs**

| Epoch | Loss | MAE | Validation Loss | Validation MAE |
|---|---|---|---|---|
| **70** | 4.2541 | 1.7076 | 4.3373 | 1.7277 |
| **80** | 4.2466 | 1.7045 | 4.3272 | 1.7238 |
| **90** | 4.3320 | 1.7301 | 4.2640 | 1.7149 |
| **100** | 4.2616 | 1.7076 | 4.2268 | 1.7089 |

The mean bounded regression accuracy for valance, arousal and dominance for the deep neural network trained for 100 epochs is 77.58%, 79.22% and 79.22% respectively. In comparison, the mean bounded regression accuracy for valance, arousal and dominance for the convolutional neural network is 74.83%, 76.88% and 77.89%, which does not perform as well as the deep neural network. This is substantiated by the higher overall mean bounded regression accuracy for the deep neural network (78.64%), compared to the convolutional neural network (76.53%).

```
                Pred | Real | Diff | Acc
=============================================
Valence         5.14 | 3.49 | 1.65 | 79.4%
Arousal         5.15 | 6.54 | 1.39 | 82.7%
Dominance       5.28 | 2.04 | 3.24 | 59.5%
=============================================
Valence         5.14 | 8.06 | 2.92 | 63.5%
Arousal         5.15 | 5.05 | 0.10 | 98.7%
Dominance       5.28 | 8.03 | 2.75 | 65.6%
=============================================
Valence         5.14 | 9.00 | 3.86 | 51.8%
Arousal         5.15 | 2.99 | 2.16 | 72.9%
Dominance       5.28 | 6.17 | 0.89 | 88.9%
=============================================
Valence         5.14 | 7.06 | 1.92 | 76.0%
Arousal         5.15 | 5.96 | 0.81 | 89.9%
Dominance       5.28 | 6.13 | 0.85 | 89.4%
=============================================
Valence         5.14 | 7.60 | 2.46 | 69.3%
Arousal         5.15 | 6.78 | 1.63 | 79.7%
Dominance       5.28 | 7.86 | 2.58 | 67.7%


Mean valence bounded regression accuracy: 77.58%
Mean arousal bounded regression accuracy: 79.22%
Mean dominance bounded regression accuracy: 79.11%
Overall mean bounded regression accuracy: 78.64%
```

**Figure 25.** DNN prediction accuracy of valance, arousal and dominance at 100 epochs. Valence, arousal and dominance values range from 1 – 9.

With an overall mean bounded regression accuracy of 78.64%, the deep neural network has the highest prediction accuracy at 100 epochs over all other scenarios discussed. In comparison, the best prediction accuracy rate for the convolutional neural network over 25, 50 and 100 epochs is 76.53%. The mean absolute error is lower in the convolutional neural network at 0.2315, compared to 1.7076 in the deep neural network when both models are trained for 100 epochs. This training cycle provides the lowest mean absolute error for both neural networks. This infers that the deep neural network is unable to accurately predict an emotion when it comes across new data by 1.7076. Based on the DEAP dataset (Koelstra, S., et al., 2012), the absolute mean error for the deep neural network of 1.7076 is on the lower end of the range for valance, arousal

and dominance, which is from 1 – 9. For these reasons, the error value for the deep neural network is reasonable. In addition, the deep neural network has minimal overfitting and underfitting compared to the convolutional neural network.

### 4.4 Computational Time

The convolutional neural network and deep neural network were compiled on two separate machines. The system hardware specifications that the models were tested on are seen in Table 11. The convolutional neural network was compiled on machine 1, which had the TensorFlow GPU running on CUDA (Compute Unified Device Architecture). CUDA is NVIDIA's parallel computing platform, which increases computing performance. This was used to reduce the processing time of the convolutional neural network. The GPU specifications are seen in Table 12. The deep neural network was compiled on machine 2, which did not utilize CUDA.

**Table 11. Specifications for system hardware used to compile CNN and DNN.**

| Specification | Machine 1 – CNN | Machine 2 – DNN |
|---|---|---|
| **Processor** | Intel® Core™ i7-6700HQ CPU @ 2.60GHz | Intel® Core™ i7-6600U CPU @ 2.60GHz |
| **Cores** | 4 | 2 |
| **Logical Processors** | 8 | 4 |
| **Installed Memory (RAM)** | 16 GB | 16 GB |
| **System Type** | 64-bit Operating System, x64-based processor | 64-bit Operating System, x64-based processor |

Based on the hardware specifications in Table 11., machine 1 has the capacity to perform better than machine 2. The convolutional neural network took approximately 1 hour and 25 minutes to train 100 epochs (45 seconds on average to process each epoch). If CUDA was not enabled on machine 1, the computing time to process the convolutional neural network for 100 epochs would increase substantially. The GPU specifications for machine 1 where CUDA was enabled are seen in Table 12.

**Table 12. Specifications for NVIDIA GeForce GTX 960M (CUDA Enabled)**

| Specification | GeForce GTX 960M- CNN |
|---|---|
| **Cuda Cores** | 640 |
| **Base Clock** | 1096 + Boost |
| **Memory Clock** | 2500 MHz |
| **Memory Interface** | GDDR5 |
| **Memory Interface Width** | 128-bit |
| **Memory Bandwidth** | 80 |

In comparison, the computing time for the deep neural network was approximately 13 minutes for 100 epochs (8 seconds on average to process each epoch) on machine 3, which did not utilize CUDA. In terms of computational time, the deep neural network preforms better as a model for prediction compared to the convolutional deep neural network.

### 4.5 Mapping DNN & CNN Model Predictions to PTSD Patients

The prediction results produced from the deep neural network and convolutional neural network are based on participants who are not diagnosed with PTSD. This is a crucial point as physiological readings and emotional responses may be more pronounced in participants who are diagnosed with PTSD. The pronounced physiological readings and emotional responses may also produce prediction results that defer from the overall bounded regression accuracy prediction results for valence, arousal and dominance observed in each of the three epoch training scenarios (25, 50 and 100) in this research.

To scale the prediction results from the deep neural network and convolutional neural network, a dataset that uses physiological readings and emotional responses from participants with PTSD should be used. The physiological readings and emotional responses will be used as inputs to train both the deep neural network and convolutional neural network discussed in this research.

# *CHAPTER FIVE*

# *5. CONCLUSION*

## 5.1 Conclusion

This research has discussed existing literature that supports VR-EBT as an effective and popular method used to treat post traumatic stress disorder. The research has also provided a high-level overview of an autonomous virtual reality exposure-based system. A deep neural network and a convolutional neural network was developed to compare prediction performance. Based on the results the deep neural network provides the highest overall mean bounded regression accuracy of 78.64% of all compared scenarios. While the mean absolute error rate was higher in the deep neural network, overfitting and underfitting was at a minimal compared to the convolutional neural network. The computational time was also found to be significantly lower for the deep neural network.

## 5.2 Limitations

Some of the limitations and restrictions faced in this research are concerned with the use of the DEAP dataset (Koelstra, S., et al., 2012). The first limitation is with respect to the number of data samples. For the model to make better predictions, more data is needed to improve prediction accuracy. The second relates to participants involved in the databased. As the participants are not diagnosed with PTSD, physiological readings and emotional responses may be more pronounced than participants who are diagnosed with PTSD. This may affect prediction results from the DNN and CNN. The dataset also relies on self-reporting, which is susceptible to emotional ambiguity. While the model was developed to generalize predictions based on a broader group of people, results produced may differ for individualized models.

## 5.3 Future Work

Future work could be expanded by classifying specific emotional states based on the prediction values of valance, arousal and dominance identified in this research. The use of wearables to extract physiological signals or an alternative data source could also be explored such as the MAHNOB HCI-Tagging database, to determine performance accuracy of both the deep neural network and convolutional neural network discussed in this research. Individualized models can also be explored and compared against the generalized models discussed in this research, and how well individualized models work on mobile devices such as health applications.

## *REFERENCES*


Banerjee D. et al. (2017). A Deep Transfer Learning Approach for Improved Post-Traumatic Stress Disorder Diagnosis. *2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA*, pp. 11-20. doi: 10.1109/ICDM.2017.10

Botella, C., Serrano, B., Baños, R. M., & Garcia-Palacios, A. (2015). Virtual reality exposure-based therapy for the treatment of post-traumatic stress disorder: A review of its efficacy, the adequacy of the treatment protocol, and its acceptability. *Neuropsychiatric Disease and Treatment*, *11*(default), pp. 2533-2545. doi:10.2147/NDT.S89542

DEAPdataset. *Queen Mary University of London, 2012, https://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html. Accessed 22 May. 2019.*

Eskofier, B. M., et al. (2016). Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL,* pp. 655-658. doi: 10.1109/EMBC.2016.7590787

Elgendi, M. (2012). On the Analysis of Fingertip Photoplethysmogram Signals. *Current Cardiology Reviews*, *8*(1), pp. 14-25

Etzel, J. A., Johnsen, E. L., Dickerson, J., Tranel, D., Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology*, *61*(1), pp. 57-69, doi: https://doi.org/10.1016/j.ijpsycho.2005.10.025

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, *47*(14), pp. 2393-2400. doi:10.1017/S003329171700040X

Gavhane, A., Kokkula, G., Shinde, S., Monghal, T., & Sisodia, J. (2016) Virtual reality: A possible technology to subdue disorder and disability. *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, *Jalgaon*, pp. 546-550. doi: 10.1109/ICGTSPICC.2016.7955361

Georgina, C., Anabel, D., I., R., Lorena, F., & Ximena, D. (2013). A controlled trial for PTSD in mexican victims of criminal violence. *2013 International Conference on Virtual Rehabilitation (ICVR), Philadelphia, PA*, pp. 41-45. doi: 10.1109/ICVR.2013.6662102

Heraz, A., Razaki R., & Frasson, C. (2007). Using Machine Learning to Predict Learner Emotional State from Brainwaves. *IEEE*. doi:10.1109/ICALT.2007.277.

Ismail, W.O. A.S., W., Hanif, M., Mohamed, S., B., Hamzah, N., & Rizman, Z., I. (2016). Human Emotion Detection via Brain Waves Study by Using Electroencephalogram (EEG). *International Journal on Advanced Science, Engineering and Information Technology*, *6*(6), pp. 1005-1011. doi: http://dx.doi.org/10.18517/ijaseit.6.6.1072.

Jerath, R., & Crawford., M. (2015). How does the Body Affect the Mind? Role of Cardiorespiratory Coherence in the Spectrum of Emotions. *Advances in Mind-Body Medicine*, *29*(4), pp. 4.

Jerdan, S.W., Grindle, M., Van, W. H. C., Kamel Boulos, M., N. (2018). Head-Mounted Virtual Reality and Mental Health: Critical Review of Current Research. *JMIR Serious Games*. *6*(3): e14. doi: 10.2196/games.9226

Kim, K. H., Bang., S., W., & Kim S., R. (2004). Emotion Recognition System using Short-Term Monitoring of Physiological Signals. *Medical & Biological Engineering & Computing*, *42*(3), pp. 419-427. doi: 10.1007/BF02344719

Koelstra, S., et al. (2012) DEAP: A Database for Emotion Analysis; using Physiological Signals. *IEEE Transactions on Affective Computing*, *3*(1), pp. 18-31. doi: 10.1109/T-AFFC.2011.15

Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., & Wrobel, M. R. (2015). Modeling emotions for affectaware applications. In Information Systems Development and Applications (pp. 55–69). Faculty of Management, University of Gdańsk, Poland.

Liu, Y., J., et al. (2018). Real-Time Movie-Induced Discrete Emotion Recognition from EEG Signals. *IEEE Transactions on Affective Computing*, *9*(4), pp. 550-562. doi: 10.1109/TAFFC.2017.2660485

Liu, Zhe., Xu, Anbang., Guo, Yufan., Mahmud, Jalal., Liu, Haibin., Akkiraju, Rama. (2018). Seemo: A Computational Approach to See Emotions, pp. 1-12. doi: 10.1145/3173574.3173938.

MAHNOB-HCI Tagging Database. *Imperial College London, 2012, https://mahnob-db.eu/hci-tagging/. Accessed 22 May. 2019.*

Mishkind, M. C., Norr, A. M., Katz, A. C., & Reger, G. M. (2017). Review of virtual reality treatment in psychiatry: Evidence versus current diffusion and use. *Current Psychiatry Reports*, *19*(11), 1-8. doi:10.1007/s11920-017-0836-0

Mouhannad, A., et al. (2018). A Globally Generalized Emotion Recognition System Involving Different Physiological Signals. *Sensors (Basel, Switzerland)*, 18(6), pp. 1905. doi: doi:10.3390/s18061905

Omurca, S. İ., Ekinci E. (2004). An alternative evaluation of post traumatic stress disorder with machine learning methods. *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA), Madrid, 2015*, pp. 1-7. doi: 10.1109/INISTA.2015.7276754

Posner, J., Russell, J. A., Gerber, A., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., Kangarlu, A., Zhu, H., & Peterson, B. S. (2009). The Neurophysiological Bases of Emotion: An fMRI Study of the Affective Circumplex Using Emotion-Denoting words. *Human Brain Mapping*, *30*(3), pp. 883-895. doi:10.1002/hbm.20553

Rizzo, A., Hartholt, A., Grimani, M., Leeds, A., & Liewer, M. (2014). Virtual Reality Exposure Therapy for Combat-Related Posttraumatic Stress Disorder. *in Computer*, *47*(7), pp. 31-37. doi: 10.1109/MC.2014.199

Rothbaum, B., O., et al. (2014). A Randomized, Double-Blind Evaluation of D-Cycloserine Or Alprazolam Combined with Virtual Reality Exposure Therapy for Posttraumatic Stress Disorder in Iraq and Afghanistan War Veterans. *The American Journal of Psychiatry*, *171*(6), pp. 640.

Soleymani, M., et al. (2012). A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing*, *3*(1), pp. 42-55. doi: 10.1109/T-AFFC.2011.25

Schurgin, M. W., et al. (2014). Eye Movements during Emotion Recognition in Faces. *Journal of Vision*, *14*(13), pp. 14. doi: 10.1167/14.13.14

Seera, M., and Lim, C. P. (2014). A Hybrid Intelligent System for Medical Data Classification. *Expert Systems with Applications*, *41*(5), 2014, pp. 2239-2249. doi: https://doi.org/10.1016/j.eswa.2013.09.022

Shields, K., Engelhardt, P., E., & Ietswaart M. (2012). Processing Emotion Information from both the Face and Body: An Eye-Movement Study. *Cognition & Emotion*, *26*(4), pp. 699-709. doi: 10.1080/02699931.2011.588691

Shi, H., Yang, L., Zhao, L. et al. (2017). Differences of Heart Rate Variability Between Happiness and Sadness Emotion States: A Pilot Study. *Journal of Medical and Biological Engineering*, *37*(4). Pp. 527-539. doi: 10.1007/s40846-017-0238-0

Shin, D., Shin, D., & Shin, D. (2017). Development of Emotion Recognition Interface using Complex EEG/ECG Bio-Signal for Interactive Contents. *Multimedia Tools and Applications*, *76*(9), pp. 11449-11470. doi: 10.1007/s11042-016-4203-7

Srivastava, K., Das, R. C., & Chaudhury, S. (2014). Virtual reality applications in mental health: Challenges and perspectives. *Industrial Psychiatry Journal*, *23*(2), pp. 83-85. doi:10.4103/0972-6748.151666

Turki, T. (2018). An empirical study of machine learning algorithms for cancer identification. *IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, 2018*, pp. 1-5. doi: 10.1109/ICNSC.2018.8361312

Urella, N., Hughes, J., Conrad, E., Zhang, J., & Li, X. (2017). A VR scene modelling platform for PTSD treatment. *12th International Conference on Computer Science and Education (ICCSE)*, pp. 257-262. doi: 10.1109/ICCSE.2017.8085499

Verma, G., K., Tiwary, U. S. (2014) Multimodal Fusion Framework: A Multiresolution Approach for Emotion Classification and Recognition from Physiological Signals. *Neuroimage*, vol. 102, pp. 162-172. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.11.007

Vijayan, A., E., Sen, D., & A., P., Sudheer. (2015). EEG-Based Emotion Recognition using Statistical Measures and Auto-Regressive Modeling. *IEEE*. doi:10.1109/CICT.2015.24.

Wang, Y., Zhao Lv., & Zheng, Y. (2018). Automatic Emotion Perception using Eye Movement Information for E-Healthcare Systems, *Sensors (Basel, Switzerland)*, *18*(9), pp. 2826. doi: http://dx.doi.org/10.3390/s18092826

Wiederhold, B. K., Miller, I. T. and Wiederhold, M. D. (2018). Using Virtual Reality to Mobilize Health Care: Mobile Virtual Reality Technology for Attenuation of Anxiety and Pain. *IEEE Consumer Electronics Magazine*, *7*(1), pp. 106-109. doi: 10.1109/MCE.2017.2715365

Ekman, P., Levenson, R., W., & Friesen W., V. (1983). Autonomic Nervous System Activity Distinguishes among Emotions. *Science*, *221*(4616), pp. 1208-1210.

Wu, C., K., Chung, P., C., & Wang, C., J. (2012). Representative Segment-Based Emotion Analysis and Classification with Automatic Respiration Signal Segmentation. *IEEE Transactions on Affective Computing*, *3*(4), pp. 482-495. doi: 10.1109/T-AFFC.2012.14

Zhang, Q., et al. (2017). Respiration-Based Emotion Recognition with Deep Learning. *Computers in Industry*, *92*(93), pp. 84-90. doi: http://dx.doi.org/10.1016/j.compind.2017.04.005

Yang, H., Han, J., & Min, K. (2019). A Multi-Column CNN Model for Emotion Recognition from EEG Signals. *Sensors*, *19*(21), pp. 4736. doi: 10.3390/s19214736

Yu, S., N., & Chen, S., C. (2015). Emotion State Identification Based on Heart Rate Variability and Genetic Algorithm. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2015, pp. 538-541. doi: 10.1109/EMBC.2015.7318418

Zheng, W., L, et al. (2019). EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. *IEEE Transactions on Cybernetics*, *49*(3), pp. 1110-1122. doi:

10.1109/TCYB.2018.2797176