

UNIVERSITY OF BAHRAIN
College of Information Technology



Emotion Recognition from ECG Signals Using Deep Learning Techniques

A Thesis Submitted in Partial Fulfilment of the Requirements for the
M.Sc. Degree in Machine learning and computational intelligence.

Submitted by
Mennatulla Mohammed Mohammed
University Number: 20120773

Supervised by
Dr. Faisal Alkhateeb
Professor
University of Bahrain

Co-Supervised by
Dr. Hessa Al- Junaid
Associate Professor
University of Bahrain

Kingdom of Bahrain
April 2023



Thesis Discussion Committee Decision

Student Name: Mennatulla Mohammed	Student ID: 20120773
Department: Dpt of Computer Science	College: Information Technology
Program: M.Sc. in Machine learning and computational intelligence	
Thesis Title: Emotion Recognition from ECG Signals Using Deep Learning Techniques	
Supervisor Name: Dr. Faisal Alkhateeb and Dr. Hessa Al-Junaid	Student GPA: 3.5

The Discussion Committee met on 08-Jun-23 In (S40-002). Upon discussing the student whose details are stated above in public, the discussion committee recommends the following:

The student is awarded a grade **Pass** Choose an item. in the thesis subject.

☐ Without making any changes to the thesis.

☒ After making the modifications attached with this decision.

Discussion Committee Members Approval:

Name

1. **Dr. Faisal Alkhateeb and Dr. Hessa Al-Junaid**

2. **Dr. Riadh Ksantini**

3. **Dr. Abdel-Raouf Bsoul**

(Internal Examiner)

(External Examiner)

Signature

(Advisor)

13-6-2023

11/6/2023

12/6/2023

Abstract

For over a century, researchers and scientists worldwide have been fascinated by electrical signals since the first electrocardiogram was produced. These signals have been the focus of extensive research and development efforts, leading to remarkable advancements in cardiology. It is widely known that our mental states significantly impact the activity of our Autonomous Nervous Systems, which in turn affects the signals of our electrocardiograms. This has led to the exploration of methods for identifying emotional states using deep learning algorithms to reach high degree of emotional identification accuracy. The Dreamer emotional recognition dataset is used to train the model. Due to the limitation of the dataset, a new method of dataset pre-processing is developed using four distinct lowpass filters for the purpose of data augmentation to the electrocardiogram signal. A hybrid architecture of Bidirectional Long Short-Term Memory and Convolutional Neural Network is used to detect nine emotions (amusement, excitement, happiness, peacefulness, anger, disgust, fear, and surprise). The proposed thesis delivers an impressive accuracy rate of 98.6%. Compared to other studies, this level of precision is the highest achievable accuracy. Compared to other studies, this level of precision is the highest possible accuracy.

This thesis demonstrates the potential of electrocardiogram signals for detecting emotional states. Furthermore, it highlights the importance of comprehensive pre-processing techniques and data augmentation to enhance the accuracy and precision of the models. The findings of this thesis provide a valuable foundation for future research in this area, with potential applications in fields such as mental health, psychology, and human-computer interaction.

Dedication

To my precious grandmother soul, how much I wished you to attend this with me

May Allah have mercy on your soul.

To my supporters always and at all times, my mother, father, and my family.

Acknowledgments

I am deeply grateful to my supervisors, Dr. Faisal Alkhateeb and Dr. Hessa Al-Junaid, whose unwavering support and guidance made it possible for me to successfully complete this job. Their expertise and constructive feedback helped me navigate through every stage of the writing process for my project, and I am genuinely thankful for their mentorship.

I would also like to express my sincere appreciation to Dr. Riyadh Ksantini for his invaluable assistance throughout my thesis. His valuable insights and suggestions were instrumental in shaping my work, and I am genuinely grateful for his contributions.

I am also indebted to the university for providing me with the necessary software and resources and to the IT college for giving me access to the AI laboratory. These facilities were instrumental in enabling me to carry out my duties efficiently and effectively.

I would like to express my heartfelt gratitude to God for guiding me through every obstacle and blessing me with the opportunity to pursue my degree. None of this would have been possible without his constant support and guidance. I will always put my faith in Him and trust in His plan for my future.

Table of Contents

Abstract.....	A
Dedication	B
Acknowledgments	C
Table of Contents	D
Table of Figure	F
List of Table.....	H
Table of Abbreviation	I
Chapter 1	2
1.1 Emotion Recognition	2
1.2 Emotion Recognition from ECG Signals	3
1.3 Research Purpose and Questions	4
1.4 Thesis Structure	4
Chapter 2	7
Section One: Background of ECG and Emotions	7
2.1 Autonomic Nervous System	7
2.2 ECG signals	9
2.3 Generating ECG signals.....	15
2.4 Adjusting the Settings for Acquiring Emotional Recognition	17
Section Two: Literature Review	19
2.5 Emotion Recognition from ECG Signals Using Deep Learning Techniques	19
2.6 Emotion Recognition from ECG Signals Using Machine Learning Techniques	26
2.7 Emotion Recognition Using Deep Learning and machine learning Techniques	32
2.8 ECG Signals Dataset.....	40
2.9 Conclusion	45
Chapter 3: Methodology.....	47
3.1 Introduction.....	47
3.2 The Proposed Model	48
3.3 Pre-processing	49
3.3.1 Down-sample and Lowpass.....	49
3.3.2 Baseline Removal	52
3.3.3 Z-Score Normalization	55

3.4 Convolutional Neural Network.....	56
3.5 Bidirectional LSTM.....	57
3.6 Fully Connected Layer	59
3.7 Activation functions.....	60
3.7.1 Softmax Function.....	61
3.7.2 ReLU Activation Function	62
3.7.3 Tanh Activation Function	64
3.8 Conclusion	65
Chapter 4: Experimental setting, Result and Discussion	68
4.1 Introduction	68
4.2 Data Collection	69
4.3 Evaluation metrics	74
4.3.1 Mean Square Error Loss Function	75
4.3.2 Categorical Cross Entropy Loss Function.....	75
4.3.3 Precision.....	76
4.3.4 ROC Curve and AUC Curve	77
4.3.5 F1 Score.....	77
4.3.6 Confusion Matrix	78
4.3.7 Accuracy	78
4.4 Results and Decision	79
4.4.1 Model Results	80
4.4.2 Model Results vs. Other Studies	90
4.5 Conclusion	93
Chapter 5: Conclusion and future work.....	95
5.1 Conclusion	95
5.2 Recommendations and Future Work.....	98
Bibliography	100

Table of Figure

Figure 1: Heart and brain connection by the ANS (Betts, Desaix, Johnson, Johnson, Korol, Kruse, Poe, Wise, Womble & Young 2013).....	9
Figure 2: Global and local reintegration Global reentry between the atria and ventricles may occur through an additional channel in addition to the AV node. The Bundle of Kent connecting the RA and RV is one such conduit, which may generate retrograde action pote (Klabunde, 2012).....	10
Figure 3: QRS complex variations (Luthra, 2019)	12
Figure 4: ECG elements. The P wave, QRS complex, and T wave reflect atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively. PR interval is time necessary for depolarization wave to cross atria and AV node; QT interval reflects ventricular depolarization and repolarization; ST segment is isoelectric period when whole ventricle is depolarized. 1mm squares.....	14
Figure 5: The proposed model architecture	47
Figure 6: Raw ECG signal with 256 Hz	51
Figure 7: ECG signal after down sampled to 128 Hz.....	51
Figure 8: Signals after applying lowpass filters and segmented to 1s.....	52
Figure 9: Signal before and after remove baseline	54
Figure 10: Baseline Signal segments and Average.....	54
Figure 11: Signal after applying baseline removal and Z-score normalization.	55
Figure 12: Example of Max pooling (Kuo & Huang, 2018)	57
Figure 13: The classical LSTM architecture (Wang, Zhang, Liang, & Liu, 2023)	59
Figure 14: The BiLSTM architecture (Wang, Zhang, Liang, & Liu, 2023).	59
Figure 15: The ReLU activation function presentation (Sultan, Salem, & Al-Atabany, 2019).....	64

Figure 16: Valence and Arousal emotion distribution (Dar , Akram , Khawaja, & Pujari , 2020).....	79
Figure 17: Left sensor Confusion Matrix	83
Figure 18: Right sensor Confusion Matrix	83
Figure 19: Right & Left sensor Confusion Matrix	84
Figure 20: Right sensor Model Loss.....	85
Figure 21: Left sensor Model Loss	86
Figure 22: Right& Left sensor Model Loss	86
Figure 23: Right sensor Model Precision	87
Figure 24: Left sensor Model Precision.....	88
Figure 25: Right& Left sensor Model Precision.....	88
Figure 26: Right sensor Model ROC	89
Figure 27: Left sensor Model ROC	89
Figure 28: Right& Left sensor Model ROC	90

List of Table

Table 1: Previous studies models and results	31
Table 2: Dataset Videos and targeted emotions.....	70
Table 3: Dataset summary	74
Table 4: Emotions Category	80
Table 5: Model Evaluation Results.....	82
Table 6: Other Work Model accuracies	92
Table 7: Other Work Model accuracies	92

Table of Abbreviation

Activation Function	AF
Area Under the Curve	AUC
Artificial Intelligence	AI
Artificial Neural Networks	ANN
Atrioventricular	AV
Automatic Speech Recognition	ASR
Autonomic Nervous System	ANS
Backpropagation Neural Network	BPNN
Bidirectional Long short-term memory	BiLSTM
Bluetooth Low Energy	BLE
Convolutional Neural Network	CNN
Deep 3D Convolution Networks	3DCNN
Deep Canonical Correlation Analysis	DCCA
Deep Learning	DL
Deep neural network	DNN
Discrete Fourier Transform	DFT

Electrocardiogram	ECG
Electroencephalogram	EEG
Electromyography	EMG
Emotion-specific multilevel dichotomous classification	EMDC
Empirical Mode Decomposition	EMD
Facial Action Coding System	FACS
Facial Emotion Recognition	FER
Galvanic Skin Reaction	GSR
Graphical User Interface	GUI
Heart Rate	HR
Heart Rate Variability	HRV
Hidden Markov Model	HMM
Human-Computer Interaction	HCI
K Nearest Neighbour	KNN
kernel Canonical Correlation Analysis	KCCA
Linear Canonical Correlation Analysis	LCCA
Long Short-Term Memory	LSTM
Machine learning	ML

Mean Square Error	MSE
Natural Language Processing	NLP
One-Dimensional Convolutional Neural Network	1DCNN
Parasympathetic Nervous System	PNS
Particle swarm optimization	PSO
Positive and Negative Affect Schedule	PANAS
probabilistic Linear Discriminant Analysis	pLDA
Receiver Operating Characteristic	ROC
Rectified Linear Unit	ReLU
Recurrent neural networks	RNN
Self-Assessment Manikin	SAM
Speech emotion recognition	SER
Support vector machine	SVM
Sympathetic Nervous System	SNS
Teager Energy Operator	TEO
Wearable Stress and Emotion Detection	WESAD

Chapter 1

Introduction

1.1 Emotion Recognition

1.2 Emotion Recognition from ECG Signals

1.3 Research Purpose and Questions

1.4 Thesis Structure

Chapter 1

Introduction

1.1 Emotion Recognition

Emotion recognition involves a series of actions and mechanisms that are utilized to detect human emotions. With the rapid advancement of computer science, which has merged with other fields such as engineering, sociology, and psychology, technology has now become capable of accurately detecting human emotions. AI plays a significant role in emotion recognition, supporting HCI. By implementing emotion recognition in various systems, the goal of the application can be achieved. The increasing use of intelligent technologies, such as smartphones and smartwatches, has further supported the use of emotion recognition systems. AI is now commonly used in emotion recognition across various techniques, including education, healthcare, marketing, and video games, ultimately enhancing service quality.

In education, emotion recognition can be used to gauge student feelings and emotions towards the provided services, including content, learning style, assessment tools, and the overall learning process. This can significantly improve the quality of the education service provided. In healthcare, AI applications that support emotion recognition can provide various healthcare services, such as reminding patients when to take their medicine, evaluating healthcare services based on patient emotions, and sorting patients based on their emotional state. Emotion recognition is also valuable for marketing applications, where it can significantly impact emotions, attracting emotions and improving marketing. Accurately recognizing emotions supports a better understanding of customers and targeting new customers. Additionally, video game

developers use emotion recognition systems during the testing process to detect end-users emotions and receive feedback before publishing the game.

Emotion can be recognized using various AI techniques. However, it is crucial to consider the different approaches adopted to detect emotions and how they collaborate with each other. Facial expressions are commonly used to detect emotions, whereas face reactions are analyzed to determine feelings. Speech signals can also be used to detect emotions by analyzing voice tone. However, people can hide their faces and control their voice, making it difficult to detect emotions accurately. Textual data can also be used to recognize emotions, as text can represent some human attitudes. Finally, physiological signals such as ECG signals and EEG can be utilized to detect emotions, where the data offered is more reliable as it is related to the ANS results.

1.2 Emotion Recognition from ECG Signals

Electrocardiographic (ECG) signals are the signals that keep tracking the HR activity. McCraty, in his study, proposed that the heart has a nervous system that impacts human attitude (McCraty, 2022). In addition, HRV refers to the change over time in both the intervals between successive heartbeats and the instantaneous HR. Since the ANS controls heart rhythm, it is believed that HRV reflects the activity of the sympathetic and parasympathetic branches of the ANS. It is feasible to distinguish between the sympathetic and parasympathetic contributions to the HRV using frequency analysis since both units of the ANS respond to stimuli at distinct speeds (A. Garc'ia, Otero, Vila, M'endez, Leandro & Lado 2016). The heart rhythm can have a direct effect from the emotions that will reflect on the ECG signals results. For example, human feelings influence the heart; when the person feels angry or afraid, the heart beats fast and comes down while feeling relaxed. However, the ECG signals are used to detect heart disease or any abnormal signals but can also detect human emotions. In addition, ECG signals are physiological signals composed of arousal, motivational

intensity, and valence that have powerful effects toward emotions and attitudes. As the using of AI is increasing and HCI get advanced in certain aspects, researcher founds that arousal could be assessed by sensors in the ECG signals. As well as the other ECG signal elements, motivational and valence can be easily calculated using the arousal results according to (Greene, Thapliyal, & Caban-Holt, 2016). The main objective of this thesis is to develop DL model that can recognize human emotions using ECG signals.

1.3 Research Purpose and Questions

The presented thesis investigates the use of DL algorithms for recognizing emotional states based on ECG readings. The main objective of the thesis is to propose a combination model of BiLSTM and CNN for detecting emotions from ECG signals. The outcome of the model was evaluated alongside the findings of previous research conducted using ML and DL models. When applied to the Dreamer dataset, the proposed model was able to define the following:

- How will the combination of BiLSTM and CNN be adapted for use in recognizing emotions from ECG signals?
- How will the proposed model enhance the accuracy of emotion recognition from ECG signals?
- What is the difference between the results of the developed model and the results of previous studies?
- What are the advantages and limitations of the suggested model?

1.4 Thesis Structure

In addition to the introduction, this dissertation includes a comprehensive description of fundamental concepts and theories that are crucial to understanding the principles of emotion identification and ECG signal analysis. Chapter 2 provides an in-depth discussion of the theoretical foundations of ECG signals, including signal variability and essential acquisition methods for ECGs.

Additionally, the chapter explores the various modalities used for emotion recognition, including physiological signals such as ECG signals and facial expressions, as well as the currently available emotion models.

Chapter 2 also includes state-of-the-art analysis of the most significant methodologies used in emotion identification today, including ML and DL techniques. These methodologies are analyzed and explained based on other studies, and the methods that were reproduced are shown in meticulous detail. The findings obtained from these methodologies are presented and explained to draw significant conclusions about the effectiveness of these techniques in emotion recognition.

The methodologies adopted in this thesis are discussed in Chapter 3, where the proposed model architecture for emotion recognition using ECG signals is explained in detail. The pre-processing methods used to prepare the data for analysis are also discussed.

Chapter 4 introduces the Dreamer dataset, a multimodal dataset that captures human emotions using a variety of modalities, including ECG and EEG signals. The chapter explains how the dataset was gathered, and it also discusses the model evaluation metrics, such as accuracy and F1-score. The chapter presents the results of the model, including a comparison with the results of other studies.

Finally, Chapter 5 provides concluding statements that consider an overview and evaluation of the dissertation's contributions to the field of emotion recognition. The chapter analyzes the successes and development status of the field, pointing out its limits and potential for future research. Additionally, the chapter highlights some possible avenues that should be pursued in additional research, such as exploring the use of multiple modalities for emotion recognition and developing more sophisticated ML and DL models. Overall, this dissertation provides a comprehensive analysis of ECG signals for emotion recognition, laying the groundwork for future research in this area.

Chapter 2

Background and Literature Review

2.1 Background

2.2 Literature Review

Chapter 2

Background and Literature Review

This chapter consists of a background of the ECG signals and the emotions. Also, consist of a literature review of detecting emotions using DL and ML methods. The proposed chapter has two main sections. The first section of the chapter introduces the ANS, which is the part of the neurological system that controls emotion and describes the ECG signals and some methods to generate the ECG signal. Moreover, adjusting the settings for acquiring emotional recognition. The second section of the chapter demonstrates the literature review of previous studies on the emotion recognition topic. Particularly the ECG recognition experiments using DL and ML will be shown and discussed. At the very end, a variety of approaches to emotional state identification will be discussed, and a brief description of the available ECG signals datasets.

Part One: Background of ECG and Emotions

2.1 Autonomic Nervous System

Emotions play an essential role in human survival, as they facilitate quick responses to a wide range of environmental stimuli that pose a threat to life. The emotional system can bypass complex cerebral processing and activate immediate, automatic reactions through the ANS when faced with potentially fatal situations. Levenson's idea has gained support from many theorists, although there is still no consensus on the number of distinct emotional states connected to diverse ANS patterns (Levenson, 2014).

Some researchers suggest that there are only two ANS patterns - the "on" and "off" states. However, others hypothesize that there are several activation patterns of the ANS, each associated with a specific emotion. The ANS has two branches: the SNS and the PNS (J. Bota, Wang, Fred, & Da Silva, 2019). The

SNS mediates the "on" state during physically or psychologically stressful conditions, regulating the body's response to danger. This system is responsible for increasing HR, bronchial tube dilation, muscle contraction, pupil dilation, and the release of adrenaline. In contrast, the PNS maintains homeostasis during rest, slowing the HR, increasing digestive system activity, and relaxing muscles. The PNS promotes increased activity in the salivary glands, tear ducts, stomach, and intestines, indicating that the metaphor of "fight or flight" and "rest and digest" works for the heart but not for other systems of the ANS (Ab. Aziz , Hasnul, Alelyani, Mohana , & Abd. Aziz , 2021).

The ANS serves multiple roles, including that of a regulator, an activator, a coordinator, and a sensor. The ANS regulates the internal bodily environment to maintain strict limits, activating and reallocating body resources to better respond to internal or external stimuli. It acts as a coordinator, organizing the continuous bidirectional flow of data between the somatic and brain systems. The ANS also acts as a sensor, detecting changes in the external environment (Levenson, 2014). However, physiological changes such as an increase in HR or breathing are not always associated with emotional states. These changes can be a consequence of one of the ANS's non-emotional capabilities, making it challenging to match a subject's emotional state with their present physiological signals. Figure 1 below illustrates this complexity.

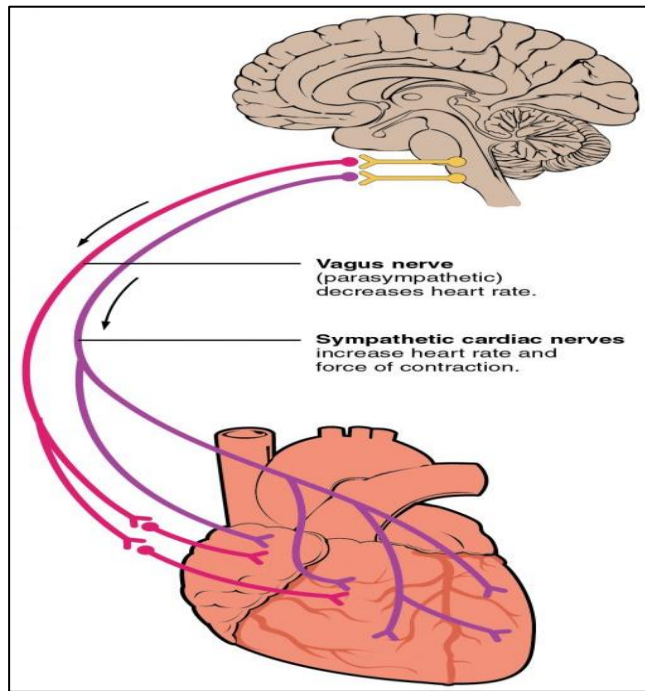


Figure 1: Heart and brain connection by the ANS (Betts, Desaix, Johnson, Johnson, Korol, Kruse, Poe, Wise, Womble & Young 2013)

2.2 ECG signals

The heart is a vital organ located in the hollow space between the lungs and to the left of the sternum (Pinto, 2017). Its primary function is to contract and move blood through the circulatory system, increasing arterial pressure and pumping blood to separate the systemic and pulmonary circulations. The heart also controls circulation by adjusting its rate and force according to metabolic needs (Tate, 2009). The heart is a muscle with three layers of protection provided by the pericardium, a thick fibrous connective tissue that secures the heart to the body and safeguards it from damage (Tate, 2009)

The heart wall has three layers: the endocardium, epicardium, and myocardium. The endocardium is responsible for preventing aberrant coagulation in the heart chambers and valves and forms the majority of the myocardium from cardiac muscle, while the epicardium interacts with the heart and covers it. The

myocardium is the thickest layer and comprises cardiac muscle (Marieb, & Hoehn, 2013).

The heart has four chambers, with the interatrial septum separating the right and left atriums and the interventricular septum separating the right and left ventricles, the heart's bottom chambers. The ventricles have thicker walls than the atriums. The right atrium receives blood from the caval veins, and the blood passes through the right AV valve, also known as the tricuspid valve, when the atrium contracts. The ventricle pumps blood to the lungs via the pulmonary artery. In contrast, the left atrium transfers blood from the four pulmonary veins to the left ventricle, which pumps the blood out via the aorta, the body's largest artery, through the mitral or bicuspid valve (Scanlon & Sanders, 2019)

The cardiac cycle is the sequence of events that occurs with each heartbeat. The heart's spontaneous contraction is caused by electrical activity in the myocardium, which follows a specific sequence and course throughout the heart, regulating the cycle. The electrical impulse governing the cycle is governed by the electrical activity of the myocardium (Kennedy, Finlay, Guldenring, Bond, Moran &McLaughlin 2016), as shown in Figure 2 below (Sampson & McGrath, 2015).

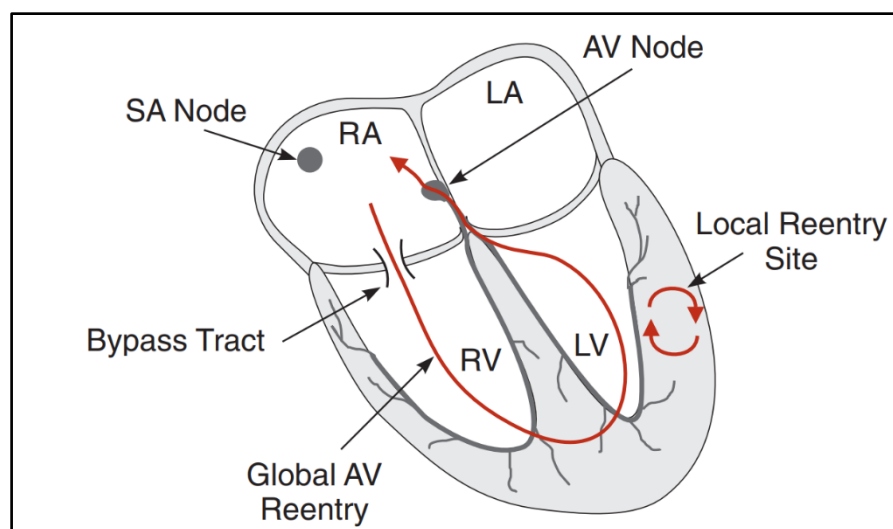


Figure 2: Global and local reintegration Global reentry between the atria and ventricles may occur through an additional channel in addition to the AV node. The Bundle of Kent connecting the RA and RV is one such conduit, which may generate retrograde action pote (Klabunde, 2012)

The cardiac cycle is a very quick process. It is comprised of an electrical signal that is generated in the heart and then transported throughout the heart. The ECG, is a graphical representation of the electrical activity of the heart that may be detected and recorded by an electrocardiograph (Marieb, & Hoehn, 2013). The three distinct waves that make up an ECG are referred to respectively as the P wave, the QRS complex, and the T wave. The P wave is formed when the AV conduction system becomes depolarized, while the QRS is caused when the ventricle becomes depolarized, and the AV conduction system becomes repolarized. At the very conclusion, the T-wave illustrates the repolarization of the ventricles. The PQ interval sometimes called the PR interval, is the period between the onset of the P wave and the onset of the QRS complex in a heart rhythm recording. The atrial walls tighten and then release throughout this phase. Depolarization of the ventricles follows the PQ phase.

The graph of the ECG is composed of waves or deflections. An alphabet letter (Luthra, 2019) has been arbitrarily assigned to each electrocardiographic interception that was performed. As a consequence of this, a wave sequence that represents a single cardiac cycle is denoted by the letters P, Q, R, S, and T and U. Also, according to tradition, P, T, and U waves are always denoted using capital letters. On the other hand, Q, R, and S waves may be denoted using either a capital letter or a small letter, depending on whether their amplitude is relative or absolute. Little waves with a wavelength of less than 5 mm are denoted by the tiny letters q, r, and s, while large waves with a wavelength of more than 5 mm are denoted by the capital letters Q, R, and S (Luthra, 2019). As a result, this indicates that the ventricles have depolarized, and the whole QRS complex is interpreted as a single entity. A R wave is a term that is often used to describe positive deflection. The negative deflection that comes before the R wave is referred to as the Q wave, while the negative deflection that comes after the R wave is referred to as the S wave. The combination of a short "q" and a long "R"

is referred to as a "qR complex," whereas a long "Q" followed by a short "r" is called a "Qr complex." The combination of a short r and a long s is known as a rS complex, whereas a tall R followed by shorts is known as a Rs complex. It would be appropriate to cite two more situations. When a QRS deflection is entirely negative, and there is no succeeding positive, this is known as a QRS complex. Second, if the QRS complex reflects two positive waves, the second positive wave is designated as R', and the complex is designated as rSR' or RsR' based on the size of the positive (r or R) and negative (s or S) waves (Luthra, 2019). Third, if the QRS complex does not reflect two positive waves, the complex is designated as rSR' as shown in Figure 3 below.



Figure 3: QRS complex variations (Luthra, 2019)

ECG does not measure real voltages but rather fluctuations in voltage based on an isoelectric reference point. Typical recording parameters for ECGs include a speed of 25 mm/s, a vertical calibration of 1 mV/cm, and a paper medium. According to the convention, the first wave of the ECG is the P wave. It illustrates the depolarization wave that moves from the SA node all the way across the atria; in most cases, it lasts between 0.08 and 0.1 seconds. Because of this, it is obscured by ventricular depolarization and has a very small amplitude; as a result, there is no wave that clearly demonstrates atrial repolarization on the ECG (Klabunde, 2012). After the P wave, there is a brief isoelectric (zero voltage) period. At this

time, the atrial cells become depolarized, and the impulse passes via the AV node, which is characterized by a significant slowing in conduction velocity (Klabunde, 2012).

In most cases, the PR interval, also known as the period that elapses between the beginning of the P wave and the beginning of the QRS complex, lasts somewhere in the range between 0.12 and 0.20 seconds. The period that has passed from the beginning of ventricular depolarization and the beginning of atrial depolarization is referred to as this gap. A conduction deficit occurs in the AV node most of the time if the PR interval is greater than 0.2 seconds. A sign of ventricular depolarization is the QRS complex in the ECG. In most cases, the QRS complex lasts between 0.06-0.10 seconds, which indicates that ventricular depolarization takes place rather fast. If the QRS complex lasts for more than 0.1 seconds, it might interfere with the conduction that takes place within the ventricles (Klabunde, 2012)..

Impairment may occur as a consequence of anomalies or improper conduction, or it may develop as a result of an ectopic ventricular pacemaker generating ventricular depolarization. Both of these scenarios have the potential to cause ventricular dysfunction. These ectopic foci nearly typically cause impulses to travel through slower paths within the heart, which lengthens the duration for depolarization as well as the QRS complex. The isoelectric period, also known as the ST segment that comes after the QRS complex, is responsible for depolarizing the whole ventricle and, more or less, corresponds to the plateau phase of the ventricular action potential. Since the ST segment may become depressed or elevated, which suggests nonuniform membrane potentials in ventricular cells, it is an essential part of the diagnostic process for determining whether or not ventricular ischemia has occurred. The T wave, which indicates repolarization of the ventricle, is present for a longer period of time than the depolarization (Klabunde, 2012)..

During the QT interval, there is both depolarization and repolarization of the ventricles of the heart. This period of time is approximately equivalent to the length of ventricular activity potentials. The QT interval may range anywhere from 0.2 to 0.4 seconds, depending on the patient's HR. The QT interval becomes reduced when HRs are high because this causes ventricular action potentials to be shorter, which in turn causes the QT interval to get shortened. It is quite important to determine whether or not a certain QT interval is excessively long. This is due to the fact that prolonged QT intervals may suggest a predisposition to particular types of arrhythmias. In clinical settings, the QT interval is often expressed as a QT interval that has been corrected. QTc is calculated by dividing the QT interval by the square root of the RR interval, which is the interval that represents the time between ventricular depolarizations (Klabunde, 2012). It is now possible to compute the QT interval without taking the HR into account. In most cases, the QTc intervals are set to 0.44 seconds, as shown in Figure 4.

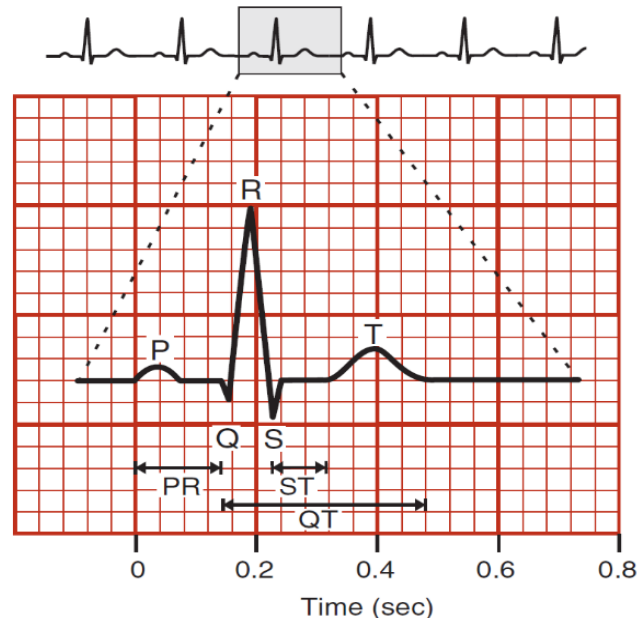


Figure 4: ECG elements. The P wave, QRS complex, and T wave reflect atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively. PR interval is time necessary for depolarization wave to cross atria and AV node; QT interval reflects ventricular depolarization and repolarization; ST segment is isoelectric period when whole ventricle is depolarized. 1mm squares.

2.3 Generating ECG signals.

There are a few different electrode designs that have already been established for use in medical diagnostic applications. These layouts provide the use of standardized methods, which ensures productivity, data quality, comparability, and reproducibility. Although there are many other ways to take an ECG, both the standard 12-lead and the modified orthogonal ECG configurations are widely used.

2.3.1 Standard 12-lead configuration

A 12-lead ECG is an approach that is used in medical treatment the most often nowadays. When seen from a certain perspective, lead serves as a sign of the electrical activity of the heart. With no more than 10 electrodes, this method generates ECG data from a total of 12 leads, offering 12 vantage points on heart activity over 2 electrical planes (vertical and horizontal) (Sohn, Yang, Lee, Ku, & Kim, 2020). These 12 leads may be broken down into three distinct groups: three unipolar and bipolar limb leads and six unipolar precordial leads.

There are three individuals in the lead position for each of these three categories. The information that is provided by the limb leads is in the vertical plane, while the information that is provided by the accountable for a substantial percentage is in the horizontal plane. Willem Einthoven, who was awarded the Nobel Prize in physiology in 1926 for his contributions to electrocardiography, is responsible for leaving behind the bipolar limb leads as a legacy. An equilateral triangle, also known as Einthoven's triangle, will serve as the basis for the coordinate system that will be used in order to record the dipole projection of the heart. Each electrode is placed on the left arm (LA), right arm (RA), and left leg (LL), which together make up the vertices of the diagram. The right leg, sometimes referred to as RL, serves as a reference point for calculating the amount of electrical current that is being pulled. The RL electrode is recognized as a surface of the electrode that helps reduce the amount of ECG artifacts as a consequence of this.

However, it does not contribute to the interpretation of ECG; it is not considered one of the 12 leads. This is because the RL electrode does not show up in ECG measurements. Lead I's RA and LA, Lead II's RA and LL, and Lead III's LA and LL are all monitored for possible variances in order to capture signals in this configuration. In addition, these three electrodes are used to collect both bipolar limb leads and monopolar limb leads, commonly known as superior limb leads. The cardiac dipole in the direction of each limb electrode is measured, and this information is reflected in the limb leads that are collected in the same manner as enhanced limb leads. This creates three distinct signals, one in the direction of LA (aVL), one in the direction of RA (aVR), and one in the direction of LL (aVF) (Pinto, 2017). The six monopolar precordial leads, or transverse leads (V1 to V6), are located in the chest and give information on the heart's horizontal plane. These leads are labeled V1 through V6. Unipolar precordial leads use just a positive electrode, similar to unipolar limb leads. One end of these six wires, the negative pole, is located near the center of the heart.

2.3.2 Corrected Orthogonal Configuration (Frank Leads)

The Frank electrode system measures the cardiac dipole's three-dimensional reach by collecting data from three channels, one for each orthogonal axis. Since Frank developed this mechanism, it is conceivable (x, y, and z, respectively). Using the limb leads, one may not only measure potential differences in the frontal plane but also in the anteroposterior direction of the body (Frank, 1956). With the Frank Leads technique, seven distinct electrodes—labeled I, E, C, A, M, and F—are used. Five of them are coplanar with the axis of rotation (A, E, M, C). I and A are located beneath the respective axillae, E and M are placed at the juncture of the sagittal and axial planes, with E anterior to the subject and M posterior to the subject, and C is located in the middle of A and E. In a nutshell, H should rest at the nape of the neck, and F should rest on the outside of the left ankle (Pinto, 2017). The three voltages V_x , V_y , and V_z represent the results of

the circuit's operation. These results are related in a way that is proportional to the p_x , p_y , and p_z dipole components (Frank, 1956).

2.4 Adjusting the Settings for Acquiring Emotional Recognition

Only signals analogous to the bipolar limb leads are used when obtaining an ECG for the study of emotion recognition. This shows that the collection of data is less difficult than that of standard medical acquisitions. This lead option may be made due to the fact that limb leads, particularly Lead I, often provide the user with a higher level of comfort. In this sense, they are the ones that are most relevant and commonly employed in applications that deal with true emotion identification. In the experiment described in (Abadi, Subarmanian, Kia, Avesani, Patras & Sebe 2015), three sensors were attached to the subject: a reference electrode was placed in the ulna bone (in the arm) and two electrodes were attached to the wrist.

This setup allows for accurate monitoring of the vital signs, including the commonly measured HR and the emotionally significant HRV. In (Katsigiannis & Ramzan, 2018), an ECG was obtained by employing a SHIMMERTM, which is a wireless sensor that can acquire vectors for both Lead II (RA to LL) and Lead III (LA to LL), however, they only employed Lead II for future research. In regard to the ECG was captured with the use of a SHIMMER 2R and three electrodes.

As a point of reference, the third one was inserted at the ankle, while the first two were positioned at the right and left arms. This configuration allowed for the accurate detection of HR as well as the whole QRS complex. In some tests, the electrodes are positioned below the collarbone as opposed to being placed on the wrists or arms (Soleymani, Lichtenauer, Pun, & Pantic, 2011). This is done in lieu of the more traditional methods. However, while doing research on emotion detection, experimental designs need to take emotion elicitation into consideration as well. This is because the ECG signal should show some degree of emotion variance. There are two distinct kinds of induction studies that may

be regarded within the realm of emotion recognition: passive and active. as opposed to a passive experiment, when people are simply exposed to a variety of emotional stimuli, participants in an active experiment must actively participate in the process of the experiment (Agrafioti, 2011).

The subject may provide a self/explicit evaluation (sometimes called an internal annotation) of the emotion they are feeling, or the annotator may make an implicit assessment based on the subject's behavior (also known as an external annotation), when the subject's emotional state is assessed from the outside by closely monitoring their actions or other outward manifestations of their state of mind. Nevertheless, the results of any of these two approaches are subject to some degree of ambiguity since they rely on the observer or the evaluator. To guarantee accurate labelling of the data, several research projects use a hybrid approach that combines the two different methods. For the purpose of emotion elicitation, passive studies may make use of films, still images, or even music recordings as an external stimulus. An example of this is the International Affective Image System (IAPS), which is a picture collection that offers rates of affect for a variety of different emotional states (Agrafioti, 2011).

This particular example is used rather often. The participants are often requested to choose a comfortable place to sit while the stimuli are being shown, and the evaluation of their emotional responses may take place either during or after the presentation of the stimuli. In the case of the active experiments, the stimulus had a dynamic interaction with the participants who volunteered their time. A popular commercial video game was used in (Agrafioti, 2011), which progressively became more challenging while also drawing the user deeper and deeper into the experience. After becoming familiar with it, the ECG and a recording of the subject's facial expressions will begin simultaneously. In conclusion, the participants were asked to play the game while watching the movie and describe how they felt each time they were presented with a new challenge. In regard to

this particular experiment, the participants provided reports on their levels of arousal. However, these active experiments are relatively new in the area of emotion recognition; hence, the majority of researchers choose to use elicitation strategies that include passive experiments, such as photographs, films, or music.

Part Two: Literature Review

2.5 Emotion Recognition from ECG Signals Using Deep Learning Techniques

Demonstration learning approaches that involve multiple layers of representation are referred to as deep-learning techniques. These techniques are based on the construction of fundamental but non-linear modules that iteratively transform the representation at one level, starting from the raw input, into a more abstract representation at a higher level (LeCun, Bengio, & Hinton, 2015). These modules operate by transforming the representation at the lower level, which starts with the raw input, and then altering the representation as it works its way up through the layers. By including a sufficient number of these transformations in the learning process, it becomes possible to learn very complex functions. The advantage of using deep-learning techniques is that they bring forward the critical features of the input that are essential for classification while minimizing the differences in categorization that are not relevant (LeCun, Bengio, & Hinton, 2015).

A significant property of DL is that the layers of features are discovered within the data using a general-purpose learning strategy rather than being developed by human engineers. This is a crucial aspect of DL, which has enabled the technique to make significant strides in solving challenging problems that have remained unsolved in the field of AI for a long time. DL has proven to be especially effective in uncovering complex structures in high-dimensional data, which has

made it valuable in a range of academic fields, as well as industry and government.

Another advantage of DL is that it requires minimal hand engineering and can quickly adapt to advancements in computer technology and the availability of data. As a result, it is expected that DL will continue to experience significant success in the years to come. The development of learning methodologies and topologies for DNNs is underway, which will only accelerate the advancement of DL.

DL techniques are powerful tools for uncovering complex structures in high-dimensional data. These techniques are based on the construction of non-linear modules that iteratively transform the representation at one level into a more abstract representation at a higher level. DL has made significant strides in solving challenging problems in the field of AI and is expected to continue to be successful in the future. As the development of learning methodologies and topologies for DNNs continues, the power of DL will only increase.

The results of the classification show that the proposed DL technique can be used for ECG classification even when the training data is unstructured and imbalanced and can be determined as a 1D time series with a standard time length for portable single-lead ECG devices (Pyakillya, Kazachenko, & Mikhailovsky, 2017). This is demonstrated by the fact that the results of the classification show that the proposed DL technique can be used for ECG classification. This is shown by the categorization findings, which point in this direction. In addition, the power of the model to extract features is highly significant when there are no. The proposed DL technique has shown promising results in the classification of ECG data, even when the training data is unstructured and imbalanced. The data is represented as a 1D time series with a standard time length for portable single lead ECG devices (Pyakillya, Kazachenko, & Mikhailovsky, 2017). The DL technique has demonstrated its ability to extract features without the need for

medical experts or related subject specialists, which is particularly useful in situations where such professionals are not readily available.

However, this study has some limitations when compared to previous DL systems, including issues with computing complexity, suggested design, and optimization approaches (Pyakillya, Kazachenko, & Mikhailovsky, 2017). RNNs, such as LSTM networks, are attractive options for ECG classification due to their ability to handle historical data, such as time series. Therefore, RNNs could be a promising candidate for ECG classification work. Nevertheless, the most significant challenge that remains to be addressed is determining the appropriate architecture for the dataset, including the number of neurons, layers, and optimization algorithms to be used (Pyakillya, Kazachenko, & Mikhailovsky, 2017).

The proposed DL technique has shown potential for ECG classification even with unstructured and imbalanced data. The ability to extract features without the need for medical experts or related subject specialists is particularly valuable. However, further research is needed to optimize the architecture of the DL model for ECG classification, including the use of RNNs and other advanced techniques.

In the research (Etemad & Sarkar, 2020), they used a number of different datasets, some of which were referred to as Amigos, Dreamer, WESAD, and SWELL. The authors proposed a model that is self-supervised and comprises of two stages (Etemad & Sarkar, 2020). The first step is to get familiar with the ECG's graphical depiction. The second step is in charge of putting the network that is able to extract ECG representation to use in order to train another network that is able to recognize emotions. The learning ECG representation that is presented here is one that is self-supervised and multitasking. CNN phase has shown the ability to recognize a variety of changes that were done to input ECG signals without making use of any of the labels that were defined by humans and provided along with the data. The network that is being shown has three

convolutional blocks serving as the common layers and seven branches, each of which has two dense layers, performing the roles of the task-specific layers (Etemad & Sarkar, 2020). The AF RELU is used inside the convolutional blocks. After that, the outcome was sent to two fully linked layers comprising a total of 128 nodes and using RELU as the AF. In addition, there is a dropout rate of sixty percent to avoid overfitting. The sigmoid activation layer was employed for the output layer. The network is able to potentially obtain robust ECG representations by learning the collection of synthetic input-output pairs that have been described above. These representations can be produced by flattening the feature embeddings that have been obtained from the shared convolutional blocks (Etemad & Sarkar, 2020).

The second step, which focuses on the categorization of emotions, involves training the model with the actual emotion labels serving as the y variable (Etemad & Sarkar, 2020). Layers with 512 fully connected hidden nodes follow convolution layers identical to those used while learning ECG representation in the network for emotion recognition. Instead of training on unlabelled data, the fully linked layers of the emotion identification network are used for emotion recognition (Etemad & Sarkar, 2020), which means that the training is carried out under full supervision. It is important to stress that the fully connected layers for each dataset that was utilized in this study have been fine-tuned; this is something that should be stressed. Two dense layers are used for the SWELL and WESAD datasets; however, for the AMIGOS and DREAMER datasets, three dense layers are used with L2 regularisation (0:0001) and 40% and 20% dropout, respectively.

The final result is produced by a layer that is either a sigmoid binary classification or a SoftMax layer. When applied to the AMIGOS dataset, the suggested model achieved an accuracy of 88.9% for arousal and 87.5% for valence. The accuracy for the dreamer dataset was 85.9% for arousal and 85% for valence. The SWELL

accuracy for 93.3% for stress, 96.7% for arousal, and 97.3% for valence. The WESAD accuracy was 96.9% (Etemad & Sarkar, 2020).

Data augmentation for human emotion identification via a CNN based on an ECG signal (Nita, Bitam, Heidet, & Mellouk, 2022). It is recommended by (Nita, Bitam, Heidet, & Mellouk, 2022) that the randomise, concatenate, and resample realistic ECG episodes method be used in order to add a considerable number of accurate ECG samples to the ECG dataset. (Nita, Bitam, Heidet, & Mellouk, 2022) This method may be found in the following sentence. In order to resample more realistic ECG events, this step would need to be taken. As a direct result of this, a one-of-a-kind CNN classifier consisting of seven layers has been provided in order to categorize human emotions in accordance with the levels of valence, arousal, and dominance. Proposed Data Augmentation CNN method was put to the test on the standard-setting DREAMER database. The results showed an accuracy rate of 95.16% for recognizing valence, 85.56% for detecting arousal, and 77.54% for detecting dominance (Nita, Bitam, Heidet, & Mellouk, 2022).

DNNs are modeled in two different ways, according to the authors of (Dar , Akram , Khawaja, & Pujari , 2020). The EEG, was the first model input (Dar , Akram , Khawaja, & Pujari , 2020). The ECG, and the GSR, were the other model inputs. Both the CNN and LSTM architectures were used in the study that was carried out by (Dar , Akram , Khawaja, & Pujari , 2020) using the Dreamer and AMIGOS datasets. Both ECG and GSR modalities have the potential to be used inside a DNN framework as series inputs. By scaling it down to 81 by 128 pixels, an EEG signal that has 14 channels may be utilized as a picture input to a DNN architecture. This is possible when the signal is reduced in size. Due to this, two distinct DNN designs for neural network classification were developed: one for EEG, which would be useful for picture classification, and another for ECG and GSR, which would be useful for signal or time series data classification. Both of these designs were developed as a result of the aforementioned reason. As a direct

consequence of this, both of these designs came into existence. The design of the EEG is predominately based on 1D convolutional layers and LSTM, whereas the architecture of the ECG and GSR is predominately based on a mix of LSTM and 2D convolutional layers. This is due to the fact that it is predicted that LSTM will fully leverage the potential of series or time-series data (Dar , Akram , Khawaja, & Pujari , 2020). The construction of an EEG is mostly based on one-dimensional convolutional layers. The accuracy of the model suggested by (Dar , Akram , Khawaja, & Pujari , 2020) towards the AMIGOS dataset for the ECG left sensor was 96.96% and for the right sensor was 98.73%. In term of the Dreamer dataset, the accuracy was 89.25% for the combined ECG signal left sensor and right sensor.

Bayesian neural networks were employed, together with the AMIGOS and Dreamer datasets, in the study that was done by (Harper & Southern, 2020). Two independent streams of information are always flowing. Each stream is made up of four convolutional layers that are layered on top of one another to extract local patterns from the time series. After each convolutional layer, a dropout and a ReLU AF are used. Over-fitting is then reduced by applying a global average pooling layer to the model. The second one uses a dropout-followed bidirectional LSTM. Input sequence history and future are both modeled in this way. A valence regression estimate is obtained by concatenating the results of the two streams and then passing them through a thick layer. By using LSTM and CNN, the model achieved 90% accuracy on the AMIGOS dataset but only 86% accuracy on the Dreamer dataset using the bayes $\alpha=0.9$.

This is the essential problem that still needs to be solved. In the study that (Yoo, Lee, Par, Kim & Jeong 2005) carried out, the ECG and the GSR were both used as model inputs. Fear, interesting pleasure, calm pleasure, and sorrowful pleasure were the four emotions that (Yoo, Lee, Par, Kim & Jeong 2005) advised using a neural network model to classify. The research presented a model that made use

of the multi-layer perception, also known as MLP, which consists of layered units. Each layer is made up of nodes, and in the completely connected networks that were investigated, every node was connected to every other node in the layer that came after it. The job of distributing inputs to the layers that come after the input layer is the input layer's responsibility. Input node AFs are linear, and there are no thresholds at any point in the network. Each hidden unit node and output node also has a threshold tied to it, in addition to the weights. The AFs of the hidden unit nodes are nonlinear, whilst the AFs of the outputs are linear. Therefore, the initial input of each signal that is received by a node in a subsequent layer is multiplied by a weight and a threshold before being subjected to either a linear or nonlinear AF (Delashmit & Manry, 2005).

The accuracy of the model was estimated to be 80.2% (Yoo, Lee, Par, Kim & Jeong 2005). The writers of this article made use of data obtained via the use of SAM which is a reliable grading system developed by Lang. The experiment that was carried out in order to acquire the data was conducted by the Medical College of Severance. The participants consisted of three men in good health and three women in good health. The measurements took thrice daily, three times every week, and for a total of three weeks. The MP150 was used in order to get the ECG and GSR readings (BIOPAC SYSTEM, USA). ECG signals are generated by the ANS and HRV. HRV plays a vital role in determining the emotion that is created by ECG signals. The sampling speeds for the ECG and GSR were 1000 Hz and 250 Hz, respectively, since HRV is especially susceptible to being affected by background noise (GSR is not as sensitive to noise).

A model is presented by (Li & Chen, 2022) that makes use of the integration of ECG and EMG feature inputs in order to determine whether or not a person is in a state of relaxation, transition, or fatigue. In this study, several different learning methods, including the BPNN (Li & Chen, 2022), are covered within the broad topic of DL. It has an accuracy of 85.80 percentage points. The BPNN model may

be broken down into two primary stages, as shown in (Li, Cheng, Shi, & Huang , 2012). The first step is called operational signal forward propagation, and it refers to the process in which the input signal is conveyed via the concealing layer on its way from the input layer to the output layer. During the forward propagation of the operational signal, the weight and offset values of the network are maintained at a constant value, and the state of the neurons in one layer does not have any effect on the state of the neurons in the layer below it. In the event that the expected output cannot be produced at the output layer, it is possible that the signal will be converted into error signal backpropagation. The following step is the backpropagation step, which is concerned with the error signal. The error signal is defined as the difference between the actual output of the network and the expected output of the network. During the backpropagation of the error signal, the error signal is transferred layer by layer from the output end to the input layer. During the process of the backpropagation of the error signal, the error feedback is responsible for managing the weight value of the network. The parameters of weight and offset are constantly being adjusted in order to make the network's actual output closer to the one that was anticipated (Li, Cheng, Shi, & Huang , 2012) .

2.6 Emotion Recognition from ECG Signals Using Machine

Learning Techniques

ML is a subfield of computer science that focuses on the understanding and creation of methods that can learn, or more specifically, strategies that make use of data to improve performance on a variety of tasks. The primary objective of this subfield of computer science is to automate as many tasks as possible. It is thought to be a component of AI. Algorithms that are built for ML are intended to make predictions or judgments without being specifically trained to do so. They do this by building a model using sample data, which is later referred to as training data once it has been processed. ML algorithms are utilized in a wide

variety of applications, including medicine, email filtering, voice recognition, and computer vision (Mitchell , 1997). These algorithms are utilized in situations where it would be difficult or impossible to develop traditional algorithms to execute the required tasks. These methods have applications in a wide number of other domains as well.

Nevertheless, statistical learning is just one component of ML. ML also includes other types of learning. The field of computational statistics, which focuses on the generation of predictions using computers, has a close connection to a subset of ML; however, statistical learning is not included in all instances of ML. This is because computational statistics is focused on the generation of predictions using computers. The study of ML might potentially gain something from the research that has been done in mathematical optimization, namely in the areas of methodology, theory, and application. One field of research known as data mining, which focuses on the exploratory examination of data via the use of unsupervised learning, is comparable to the topic at hand. The use of data and neural networks to represent the functions of the brain enables some applications of ML to create a simulation of how the brain really works. The term predictive analytics is widely used to refer to the application of ML to problems that arise in the commercial world (Mitchell , 1997).

The volunteers in the experiment that (Kim & André, 2008) prepared and listened to music while their ECG signals were monitored so that the researchers could identify how the individuals were feeling. In addition, tests of respiratory function, skin conductivity, electromyogram, and ECG were included in the study (Kim & André, 2008). In addition to this, linear discriminant analysis was a technique that was used throughout the process of the creation of ML (Kim & André, 2008). The suggested model conducted by (Kim & André, 2008) is the advanced linear discriminant analysis called pLDA; the accuracy of the proposed model was 87%. The (Kim & André, 2008) also introduced the EMDC schema that results high accuracy with 95%.

A related study conducted by (Sun & Lin, 2022) resulted in the development of a model that was able to differentiate between four distinct states of emotion: calm, enthusiastic, tense, and delighted. In the model that was developed as a result of this study, the ECG served as an input. PSO was employed as an optimization approach to maximize the added parameters. Other classifiers, such as random forest, SVM, and KNN, were also applied. The SVM model is a supervised learning approach that is used to categorize and make predictions. The primary objective here is to locate an appropriate hyper-plane that can partition the data into a variety of diverse groups. The hyperplane is found by looking at where the support vectors are located on the surface of the hyperplane. An ideal hyperplane is one that allows for the greatest possible margin of separation between points. Moreover, the structure of a Random Forest may be well described by using the metaphor of a decision tree. Combining the results of many decision trees is an effective way to improve precision. In the field of ML, there is a model called a decision tree that depicts the link between different things and the traits that they have.

The conditions under which an analysis was carried out are represented by the trunk of the tree, while the findings reached as a result of the study are represented by the branches and leaves of the tree. In Random Forest, the analysis is less structured and more open to chance than in other supervised learning approaches. In order to create decision trees, it first takes a subset of the characteristics that make up a feature set and then averages the results. So, when compared to other classifiers, the random forest has a generalization capacity that is greater. The method of ensemble learning known as random forests is superior than bagging in terms of performance. Ensemble learning is a method that handles the problem of creating accurate predictions by combining a number of different models into a single answer. A random forest is a collection of decision trees, each of which may be based on a wide range of various data sources. Random forests are used to make predictions about the future (Sun & Lin, 2022).

The Random Forest model has an accuracy of 81.70 percent overall. The KNN method is the most straightforward non-parametric classification algorithm available. This model's accuracy was 82.48%, and it assigns classes to samples that have not been seen by comparing them to those in the training set and the test set and determining how similar they are to those samples (Sun & Lin, 2022).

In other research by (Sepúlveda , Castillo, Palma , & Rodriguez-Fernandez, 2021), which also made use of the AMIGOS dataset, the authors classified the data utilizing a number of distinct methods, such as LDA, decision tree, KNN, ensemble bagged tree, SVM, and kernel naive bayes. This research also made use of the AMIGOS dataset. In a proposed study (Sepúlveda , Castillo, Palma , & Rodriguez-Fernandez, 2021), the researchers extracted characteristics from signals using a technique known as wavelet scattering. In order to extract information from the signals, this method first segments each signal into a certain number of scattering windows.

Then, the scattering windows undergo wavelet transformations so that the information may be extracted. After that, each of the scattered windows is classified on its own, and the initial segment is classified based on a voting system that gives an equal amount of importance to each of the classifications of the scattered windows (Sepúlveda , Castillo, Palma , & Rodriguez-Fernandez, 2021). After that, each of the scattered windows is categorized in accordance with its own criteria. When the wavelet scattering approach is used, the three steps of the signal alteration that take place are called wavelet convolution, wavelet modulation, and filtering (Sepúlveda , Castillo, Palma , & Rodriguez-Fernandez, 2021).

The model measurements as the following: 43.2% accuracy was achieved with the decision tree model. Naïve bayes model accuracy 19.3%. The SVM model achieved an accuracy of 51.9% overall. The LDA model accuracy was 66.1%. An accuracy of 88.9% was achieved using the ensemble bagged tree model,

which was the highest achieved accuracy in (Sepúlveda , Castillo, Palma , & Rodriguez-Fernandez, 2021) research.

According to the conclusions of the study (S, Murugappan, Wan, & Yaacob, 2014), using characteristics based on the frequency domain rather than EMD-based features offers a number of benefits that cannot be ignored. They assert that in contrast to EMD features, which give information about the local properties of the ECG wave, the DFT approach offers information about the frequency content of the signal. This is in contrast to the fact that EMD features give information about the local properties of the ECG wave. By using DFT-based characteristics of 10 intrinsic mode functions developed from the EMD approach (S, Murugappan, Wan, & Yaacob, 2014), they were able to distinguish between neutral, happy, sad, fearful, surprised, and disgusted feelings from ECG signals with an accuracy of 54%. These feelings were able to be gleaned from the ECG signals. Because of this, they were able to categorize the feelings of the individuals as being either neutral, pleased, sad, astonished, or disgusted.

Table 1: Previous studies models and results

Paper	Dataset	Model	Accuracy
(Yoo, Lee, Par, Kim & Jeong 2005)	Private	Multi-layer perception (MLP)	80.2%
(Etemad & Sarkar, 2020)	AMIGOS	CNN	Arousal 88.9% Valence 87.5%
	DREAMER		Arousal 85.9% Valence 85%
	SWELL		Stress 93.3% Arousal 96.7%% Valence 97.3%
	WESAD		96.9%
(Nita, Bitam, Heidet, & Mellouk, 2022)	DREAMER	CNN	Arousal 85.56% Valence 95.16% Dominance 77.54%
(Li & Chen, 2022)	Private	back propagation neural network (BPNN)	85.80%
(Dar , Akram , Khawaja, & Pujari , 2020)	AMIGOS	CNN and LSTM	ECGL 96.96% ECGR 98.73%
	DREAMER		89.25%
(Harper & Southern, 2020)	AMIGOS	LSTM and CNN with	90%
	DREAMER	Bayesian neural networks	86%
(Kim & André, 2008)	Private	pLDA	87%
		(EMDC) schema	95%
(Sun & Lin, 2022)	Private	Random Forest	81.7%
		KNN	82.48%
(Sepúlveda , Castillo, Palma ,	AMIGOS	Decision Tree	43.2%
		Naïve bayes	19.3%
		LDA	66%

& Rodriguez-Fernandez, 2021)		SVM	51.9%
		Ensemble	88.9%
(S, Murugappan, Wan, & Yaacob, 2014)	Private	EMD	54%

2.7 Emotion Recognition Using Deep Learning and machine learning Techniques

People have to depend on their own interpretations of the sentiments that others are expressing in daily life by studying the facial expressions and general body language of the other person. The latter is of utmost significance due to the fact that it is the primary mode via which nonverbal communication is carried out. It's possible to tell a lot about a person just by looking at their face, including their age, gender, identity, background, and even how they are feeling at any particular time. It should not come as a surprise as a result of this that a large number of behavioral scientists are interested in how the expressions on a person's face may transmit a large amount of information (Cohn & Torre, 2015). As a direct result of this, it should not come as a surprise that this is the case.

Since the early 1970s, computer scientists have been employing a person's face as a biometric modality. This is due to the fact that it is the approach that is now accessible that is the most intuitive. This is due to the fact that the face is the most reliable predictor of the identification of a person. The FACS, is today regarded to be one of the most effective ways for distinguishing facial expressions of emotion (Cohn, Ambadar, & Ekman, 2007). This technique analyses facial expressions by focusing on the independent movements made by each of the muscles in the face. Action Units (AUs) are the smallest motions that can be recognized and detected, and their name comes from the fact that they are the

smallest movements that can be identified. Moreover, the ability to recognize micro-expressions is provided by this technology.

(Cohn, Ambadar, & Ekman, 2007) served as a source of motivation for a great number of academics working in the academic world, who have now begun investigating facial expressions by using images in addition to films. In the realm of facial expression recognition, various different image-processing approaches have progressively found their way into use over the course of time. Pre-processing, feature extraction, and classification are the three primary stages that are often included in modern approaches to FER. It is possible to enhance the functionality of FER by carrying out a preparatory procedure known as pre-processing. It is possible to carry out this step. It is common practice to do this step before the feature extraction stage in order to facilitate not just the feature extraction process but also the classification step that will follow.

(Bashyal & Venayagamoorthy, 2008) Enhancing the picture's clarity and contrast, making modifications to the size of the image, and scaling it is some of the most popular procedures used in image pre-processing. Following that, certain qualities are retrieved and entered into a system that is concerned with the process of categorization. The last but not least component of this experiment is the output, which is one of the previously chosen labels or types of emotions. Nevertheless, FER approaches may be divided into two separate categories, depending on whether the features are created manually or automatically via the use of a DNN. The manual production of features is one of these types. Depending on the particulars may further classify the handmade aspects as feature-based or region-based categories, respectively. This would depend on the circumstances. Feature-based identification refers to the process of determining an individual's identity by analyzing distinct physical characteristics of their face, such as the eyes or the wrinkles at the corners of the mouth. On the other hand, detection that is region-based homes down on certain parts of the face in order to identify shifts

in expression. This is done in order to distinguish between different emotions. For example, one technique for doing this analysis is to concentrate on the area surrounding the eyes and eyebrows, in addition to the lips, while ignoring the rest of the face. The identification of face structures, the extraction of facial characteristics, and the categorization of facial emotions are the three main phases involved in traditional FER approaches, as shown in the illustration (Ko, 2018).

On the other hand, DL methods present themselves as intriguing approaches that can be used to extract features automatically without any prior effort or even awareness of the process of automatically extracting particular and significant characteristics. This makes them an attractive alternative to traditional ML techniques. The use of DL strategies has gained popularity in recent years. Several various kinds of methods, such as CNN (Zhang, Jolfaei, & Alazab, 2019), 3DCNN (Pranav, Kamal, Chandran, & Supriya, 2020), and even RNNs (Li & Deng, 2022) and (Mellouk & Handouzi, 2020), are described in the scholarly literature. Moreover, there are RNNs (Li & Deng, 2022) and (Mellouk & Handouzi, 2020) and 3DCNN (Zhang, Jolfaei, & Alazab, 2019) and (Pranav, Kamal, Chandran, & Supriya, 2020). Having said that, methods that are based on DL might eliminate the need for effective pre-processing and feature extraction in order to get desirable outcomes. Nevertheless, these technologies need a significant amount of processing power, a massive dataset, or other ways like data augmentation in order to function well. In addition to this, the amount of time necessary to use these technologies is significantly increased. In conclusion, facial expressions continue to be one of the most often employed modalities due to the great appeal of these emotions.

Speaking to one another and interacting with one another is the most natural and instinctive method for people to communicate with one another and do so. In addition, a significant deal of information about how individuals are feeling may be gained from their speech, not just from the words that they say but also from

the tone of their voice, tremble, or amount of assertiveness in the way that they make their statements. People's voices have a tendency to shake more when they are concerned, and when they are happy, they have a tendency to speak more loudly and with a broader variety of intonation. As a direct consequence of these findings, a number of researchers made the decision to investigate the ways in which changes in patterns of speech may be associated with differences in assumed states of mind. Since the 1950s, several research projects on voice recognition have been carried out, and one of their primary focuses has been the translation of human speech into written language. On the other hand, there has been a very limited amount of research carried out on the subject and the outcomes that have been accomplished in this area (Nicholson, Takahashi, & Nakatsu, 2000). SER is the process of determining a person's emotional state by studying the content of their speech.

Using audio or transcriptions in combination with text processing is one method that may be used to accomplish this objective successfully. In spite of this, it is not a simple process because of the variability brought about by the large variety of speakers, speaking styles, and speaking speeds (Banse & Scherer, 1996). The first phase, known as feature extraction, is analogous to other classification tasks, such as FER. Finding appropriate characteristics that may adequately define emotions is the goal of this step of the process. These characteristics may be local, in which case they may be determined by chopping the signal into extremely tiny pieces and analyzing each one in a static manner; alternatively, they may be global, in which case they cannot be discovered in this manner.

Following that, the pitch and the energy can be retrieved from each frame, and then those two factors may be utilized to calculate the various changes in mood. In order to produce global features, the data representing each of the speech attributes are collected and used. Even though there is debate in the academic community over which kind of feature is superior, the vast majority of industry

professionals are in agreement that global features provide the greatest benefits in terms of the amount of time saved and the quality of the categorization (El Ayadi, Kamel, & Karray, 2011).

The TEO-based category of speech characteristics includes the subsets of speech characteristics known as spectral speech characteristics, continuous speech characteristics, and qualitative speech characteristics. DL techniques are also utilized to identify the emotions conveyed in a speaker's voice, hence eliminating the requirement for human feature extraction in the analysis process. SVMs, and ANN are two of the classifiers that are used the majority of the time. On the other hand, the HMM, is the classifier that is used the vast majority of the time in speech emotion classification and identification (Nwe, Foo , & De Silva, 2003) (Schuller, Rigoll, & Lang, 2003). In spite of the fact that in the past, some positive findings have been achieved, this topic continues to be intriguing, and there is a great deal of potential for progress that may be attained by tackling obstacles such as speaker-dependent categorization. Despite this, there are still certain gaps in the ability to recognize emotions via spoken language. While it is not as simple to change and modify as facial emotions, some components of the performance, such as pace and intensity, may be modified to disguise specific emotional states. This is possible even though it is more difficult to do so than with facial expressions.

DCCA is a method for non-linear transformation, and it was utilized in the article that was described by (Liu, Qiu, Zheng, & Lu, 2019), which included multiple datasets such as SEED, DEAP, and Dreamer. The technique for non-linear transformation is called DCCA. In this particular investigation, the DCCA model was used for the analysis of EEG data as biosignals. The DCCA approach is used in this study to recognize multimodal manifestations of emotion. DCCA is a method that was proposed by (Andrew, Arora, Bilmes, & Livescu, 2013) as a strategy for learning complex nonlinear transformations of two different

viewpoints on data, which would ultimately result in highly linearly linked representations. DCCA is a strategy for learning complex nonlinear transformations of two different viewpoints on data. Learning the parameters of both transformations at the same time is necessary in order to get the greatest possible improvement in the regularised total correlation. Within the nonlinear domain, it is an extension of the LCCA approach. For the purpose of learning correlated nonlinear transformations, it is a nonparametric approach that is analogous to KCCA.

DCCA, in contrast to KCCA, does not require an inner product, and it possesses the benefits of a parametric technique in the sense that the training process scales well enough with the size of the data, and the training data do not need to be referred to when calculating representations of unseen instances. KCCA, on the other hand, does require an inner product. DCCA also possesses the benefits of a nonparametric technique in the sense that it scales well enough with the size of the both of these are significant improvements over the KCCA approach. This work makes use of a variety of datasets, and each of those datasets has a model with its own unique set of hidden layers and output units (Andrew, Arora, Bilmes, & Livescu, 2013). The researchers created a model for the SEED dataset that included 6 hidden layers and 20 output units. The study found that this model had an accuracy rating of 94.58% on average. The building model with 7 hidden layers and 20 output units for the SEED-IV dataset had a mean accuracy of 87.45%. This model had been given these parameters. Also, they employed the SEED-V dataset. The model that was used for this dataset has 2 hidden layers and 12 output units, and it had an accuracy mean of 83.08%. One of the datasets that were used was DEAP, and the model that was developed included seven hidden layers and twenty output units. The overall accuracy was a mean of 88.51 percent (Liu, Qiu, Zheng, & Lu, 2019).

They used the Dreamer dataset, and the model that was utilized for this dataset consists of 2 hidden layers in addition to 5 output units (Liu, Qiu, Zheng, & Lu, 2019). 88.99% on the arousal scale, 90.57% on the valence scale, and 90.67% on the dominance scale for arousal.

(Liu, Qiu, Zheng, & Lu, 2019) the SEED Dataset was used by the authors of the aforementioned research. The researchers chose 15 Mandarin video clips from a collection of materials to use as stimuli for their study. These films represented three different emotions: joyful, neutral, and sad. Before beginning any section of the experiment, the participants were given an explanation of the methodologies that would be used throughout the experiment. Over the course of the tests, the participants were given the assignment to watch 15 movie clips and write down how they felt about each one. After seeing a short film clip, the participants each had 45 seconds to provide their feedback, followed by 15 seconds of respite. We use the identical subset of the SEED dataset that we used in our prior study for the comparative research that we are doing. Both EEG and eye movement signals are included in the SEED dataset. The data for the EEG were collected at a sampling rate of 1000 Hz using an ESI NeuroScan system, which was connected to a 62-channel electrode cap. For the purpose of gathering eye movement signals, SMI eye-tracking glasses were used.

In addition, the SEED-IV dataset was one of the datasets that were utilized for the research (Liu, Qiu, Zheng, & Lu, 2019). The methodology was quite similar to what was done for the SEED dataset, and the stimulation materials consisted of seventy-two film clips. The dataset comprises emotional EEG signals, as well as eye movement, signals from four unique emotions: joyous, sad, neutral, and fear. The trials involved a total of 15 subjects (7 male and 8 female) (7 male and 8 female). Each participant went through a total of three sessions spread out over many days, with each session including a total of 24 trials. In each of the experiments, the subject saw a different portion of the movie. In addition, SEED-

V was used, which includes EEG and eye movement information for each of the five distinct states of mind (happy, sad, neutral, fear, and disgust). Each participant went through the procedure three times, and there were a total of sixteen participants (6 men and 10 women) chosen to take part in the experiment. The participants were told to see 15 movie excerpts during the trial (3 clips for each emotion) (3 clips for each emotion).

The SEED dataset, the SEED-IV dataset, and the SEED-V dataset all made use of the same pieces of equipment. The SEEDV dataset that was used in this investigation will be made available to the academic community when it has been included in SEED and become part of its subset. One of the datasets is called DEAP, and it is a multi-modal collection of data that is used to research human emotions. Throughout the course of the experiment, a total of 32 participants had their EEGs and other physiological data recorded while watching forty music videos that each lasted one minute. The participants gave each film a score based on how it made them feel in terms of arousal, valence, like/dislike, dominance, and familiarity.

According to the findings of research (Ayata, Yaslan, & Kamasak, 2017). The EMD approach disassembles the signal into a series of oscillating signals in a way that is comparable to the previous example. There is a connection between the characteristics of this technique and a group of fragmented components that are linked to a GSR wave. There is a linkage between these two sets of characteristics. In contrast to the two approaches discussed above, the inner beat methodology performs an analysis of the raw GSR wave in the temporal domain. Also, in contrast to the frequency domain-based characteristics gained by the EMD and HRV techniques, the DFT method offers an overview of the frequency domain of the original GSR wave. This is because the DFT method uses DFTs. Because of this result, the DFT technique is also known as the DFT method. The characteristics that are generated as a result of using any of these methods

represent a diverse collection of traits that correspond to various domains and areas of the ECG wave. Each of these methods carries with it a distinct set of benefits, and the characteristics that are generated as a result of using any of these methods can be generated by using any of these methods. In spite of this, the majority of GSR-based feature extraction algorithms found in the published research have emotion detection abilities of about 55%. These algorithms are used for a variety of classification tasks. This proportion varies based on the particular approach that is used. It was indicated that a research study authored by (Ayata, Yaslan, & Kamasak, 2017) is an example of one that used GSR signals in the process of emotion detection. In this specific piece of study, the authors made use of a decision tree, a KNN, a random forest, and a SVM (Ayata, Yaslan, & Kamasak, 2017). Each of these four classifiers has its own unique set of advantages and disadvantages. In addition, it was stated in (Ayata, Yaslan, & Kamasak, 2017) that the authors make use of EMD, which contributes to an improvement in the degree of accuracy.

2.8 ECG Signals Dataset

ECG data from 17 subjects are included in the dataset that was collected for WESAD. At a sampling rate of 700 Hz, ECG information was collected using RespiBAN Professional (Schmidt, Reiss, Duerichen, Marberger, & Laerhoven, 2018) sensors. The purpose of the research was to investigate four distinct emotional states, namely, neutral, amused, stressed, and meditative. In order to carry out this study, four different types of test scenarios were designed. In the beginning, participants were asked to engage in mundane tasks such as reading a magazine and sitting or standing at a table for twenty minutes so that objective data could be collected. Participants saw a total of eleven comedic films with a combined running time of 392 seconds during the pleasure activity. After that, as part of the stress scenario, the participants had to perform 10 minutes of activities that included both public speaking and mathematics. At the end of the class,

everyone took part in a guided meditation session that lasted for seven minutes. WESAD dataset was used by (Etemad & Sarkar, 2020).

PANAS method (Watson, Clark, & Tellegen, 1988) was used at the end of each trial to gather the ground truth labels for the affect states. In addition to that, the SWELL dataset was used (Etemad & Sarkar, 2020) in the process of developing the model that was recommended. This dataset was produced with the objective of gaining a deeper comprehension of the mental stress or emotional characteristics experienced by employees in a typical office environment under a variety of conditions (Koldijk, Sappelli, Verberne, Neerincx, & Kraaij, 2014). The usual, the time crunch, and the interruptions were the three types of events that were manufactured. Under typical conditions, participants were given a maximum of forty-five minutes to engage in a variety of tasks, including the creation of reports, presentations, and other similar endeavors. Under the time pressure condition, the maximum amount of time that may be allowed was reduced to 30 minutes in order to generate pressure. During the session that was designated as an interruption, participants were periodically contacted by email and text message and given the task of responding to these messages. At a sampling rate of 2048 Hz, ECG data from 25 participants were collected with the use of a TMSI MOBI device (Etemad & Sarkar, 2020) that had self-adhesive electrodes. The last step was collecting the participants' self-reported affect ratings at the end of each scenario. These evaluations were on a scale that ranged from 1 to 9.

In addition, the AMIGOS dataset was included in the model that was developed (Etemad & Sarkar, 2020) and the model used by (Dar , Akram , Khawaja, & Pujari , 2020). Also, (Harper & Southern, 2020) used the AMIGOS dataset and (Sepúlveda , Castillo, Palma , & Rodriguez-Fernandez, 2021). The AMIGOS dataset (Miranda-Correa, Abadi, Sebe, & Patras, 2018) contains responses from forty different people who were asked to interact with multimedia content in two

different settings: either by themselves or in a group of four other people. The goal of the study was to investigate factors such as personality, mood, and emotional responses to the content. Under the context of this study, participants were shown both brief and lengthy emotional video clips with the purpose of eliciting affective reactions. The shorter video snippets clock in at a duration of just 250 seconds, but the longer video clips clock in at more than 14 minutes. Data from an ECG was collected using shimmer sensors (Mellouk & Handouzi, 2020) at a sampling frequency of 256 Hz. After each trial, each participant was shown a total of 16 video clips, and their self-assessed arousal and valence ratings on a scale ranging from 1 to 9 were recorded.

Throughout the process of data collection, the electrodes were not removed from the body unless in very unusual circumstances (eating a meal, using the restroom, etc). As a consequence of this, data collectors were able to almost continually monitor significant signals under any conditions, even routine ones. Alterations in their environment had no impact on the participants since they were kept in the same setting throughout the stimulation and relaxation phases of the experiment. Experiment materials included video excerpts from various online platforms. Some were designed to provoke positive feelings, while others were used to provoke unfavorable responses from the audience. Due to the variety of preferences among us, we were unable to decide which film would best illustrate the issues. These participants were provided with video clips of television episodes that they typically watch in the comfort of their own homes. After seeing each video clip, the participants were given a questionnaire to complete in which they were asked to provide an in-depth assessment of how they felt over the course of the film (Yoo, Lee, Par, Kim & Jeong 2005).

This study included twenty physically healthy participants ranging in age from 22 to 26 years old, with 8 males and 12 females included. Throughout the course of the research, both Smart Sensors and Trigno Wireless Systems were used,

and triaxial accelerometers were included in the EMG sensors. The EMG data were collected using EMGworks4 software, which allowed for the establishment of 16 EMG sensors and 48 accelerometer analogue channels as part of a collection and analysis program for motion capture and other third-party data collecting systems. Integration with a wide variety of measuring systems was made easier by the full trigger function, which allowed the sensor to react fast to skin interference. The researchers examined 20 different physical exams, each performed by participants who were in excellent physical shape, free from obvious disease, and had no lower limb muscle or knee problems. They were also required to be full-time majors in Physical Education, with aerobics as the primary focus of the program. Prior to beginning the study, the experimental procedure was discussed with the participants who volunteered their time, and everyone who took part in the experiment provided written approval. In accordance with human anatomical knowledge, simultaneous recordings of ECG and surface EMG signals were carried out at the anterior tibialis and semitendinosus muscles at a sampling frequency of 2 kHz. Ch-2 and Ch-3 were surface EMG sensors in the right semitendinosus and left anterior tibialis muscles, respectively, whereas Ch-1 was an ECG sensor. Two groups, each consisting of 10 people, were created, and Pilates movement training was arranged.

Workouts were to last for a minimum of 15 minutes, and as part of the research, ten different people each went through one training cycle, resulting in 40 different examinations, with each one taking 150 minutes. Upon completion of the exam, the first group went straight to sleep, but the second group kept going. Throughout each training cycle, the participants returned to a calm standing position every 30 seconds, assessed their exhaustion, and assigned a number from one to three for their level of fatigue: relaxed equals one, transition equals zero, and fatigued equals one. The surface EMG, the ECG, and the total amount of training

time were all recorded, resulting in a total of 30 surface EMG and ECG groups formed by each group trial. There were a total of 90 surface EMG and ECG datasets, and each set was classified into one of three categories of tiredness. When the experiment was complete, a total of 3600 surface EMG and ECG data points were gathered. (Li & Chen, 2022).

The dataset was gathered with the help of an iRealcare ECG sensor. The study used the ECG sensor that was provided by IRealcare in order to monitor the patient's HRs as a function of their emotional states. After transmitting the data to a mobile app through BLE, the iRealcare ECG sensor then uploads the information to the cloud BLE. It is possible to obtain the unprocessed ECG signals from the cloud. The signals of individuals' ECGs were captured while they were in one of four distinct states of mind: calmness, excitement, happiness, or stress. Every other feeling, with the only exception of calmness, is brought on by events that occur in one's immediate environment. An ECG was taken while the subject was in this state of tranquility since health is best assessed when the subject is at rest. The ECGs of participants were monitored as they worked out or watched comedies or suspense films in order to record the signals that indicate excitement, happiness, and stress. Given the potential for fast shifts in sentiment, the duration of the record should not be excessively drawn out. The longer ECG signals are gathered, the greater the likelihood that they may include noise. As a result, they maintain a journal of their feelings for a period of five minutes. A sample is taken of either the ECG or the heartbeat at a rate of 128 hertz. Given that each individual is one of a kind, the author has chosen five people to take part. There are a total of twenty data sets pertaining to the four emotions, with each emotion being represented by five different sets (Sun & Lin, 2022).

2.9 Conclusion

In the previous section, each proposed research and study that demonstrates emotion recognition uses ECG signals and other emotion recognition methods, such as emotion recognition using facial expression, speech, and other biosignals.

The introduced studies have different methods for achieving emotion detection. All the discussed articles did not suggest a model that combines BiLSTM and CNN, but some studies provide a model that uses the combination of LSTM and CNN, as stated by (Dar , Akram , Khawaja, & Pujari , 2020) and (Harper & Southern, 2020) . Both of them used the AMIGOS and Dreamer datasets, but the model conducted by (Dar , Akram , Khawaja, & Pujari , 2020) has a higher accuracy than (Harper & Southern, 2020).

The studies discussed used different approaches of DL and ML; some of the studies used will known datasets, and others used their own data collected by the researchers themselves. The other proposed articles apply different pre-processing methods to be able to the developed model. As recognized in the introduced papers applying LSTM has a positive impact on the accuracy of detecting emotions, as illustrated in Table 1.

After considering the various issues addressed by prior state-of-the-art investigations, we conclude that a new DL approach adopting a combination of BiLSTM and CNN is required for ECG emotion detection.

Chapter 3

Methodology

3.1 Introduction

3.2 The Proposed Model

3.3 Pre-processing

3.4 Convolutional Neural Network

3.5 Bidirectional LSTM

3.6 Fully Connected Layer

3.7 Activation Functions

3.8 Conclusion

Chapter 3: Methodology

3.1 Introduction

The proposed model is a combination of a CNN, bidirectional LSTM, and fully connected layers. In the first section of this chapter, the architecture of the proposed model was demonstrated, as shown in Figure 5. The second part provides the dataset pre-processing methods explained. In addition, the CNN that is part of the suggested model is discussed, and a brief description of the BiLSTM architecture. Moreover, the fully connected layer is discussed. At the end, the used AF is explained briefly.

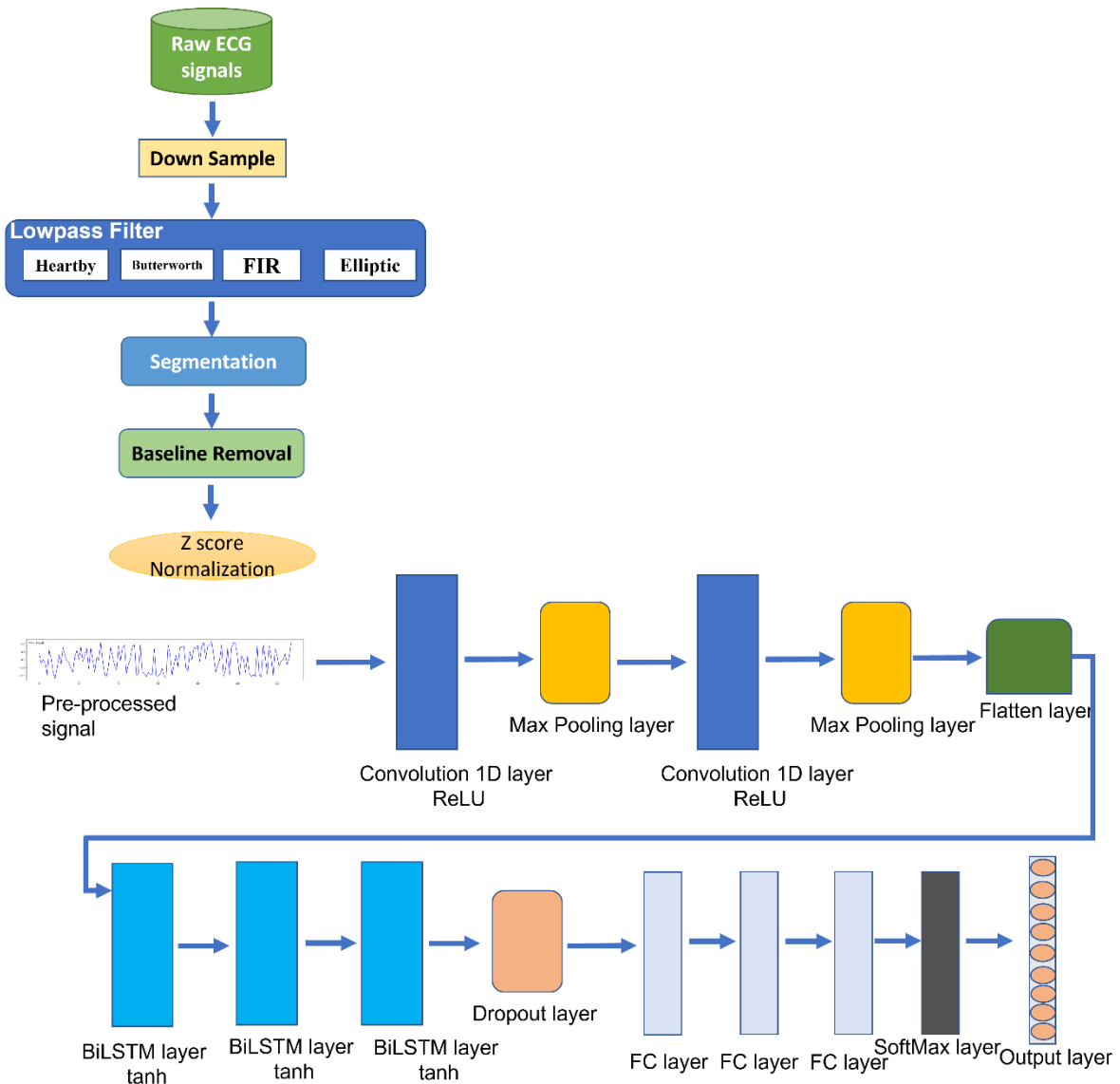


Figure 5: The proposed model architecture

3.2 The Proposed Model

The model proposed in this thesis is a CNN BiLSTM model, as shown in Figure 15. One second signal is the data entry to the model. The sequence of the data is passed to a 1D-CNN layer with a filter size of 150, a kernel size of 3, and an input shape of (128,1). Kernel size of 3 allows the model to capture local patterns in the signal, which may be important in differentiating between different emotions. Using a larger kernel size may result in the model missing out on these local patterns. On the other hand, a smaller kernel size may result in the model being too sensitive to noise and minor fluctuations in the signal. Also, a kernel size of 3 is commonly used in 1D-CNNs for time-series data, as it strikes a balance between capturing local patterns and reducing the number of parameters in the model. Using a larger kernel size would result in a larger number of parameters and a more complex model, which may lead to overfitting and reduced generalization performance.

The AF in the first 1D-CNN layer is ReLU. A max pooling layer with a size of 2 follows to reduce the number of parameters. Then the second 1D-CNN layer with a filter size of 100, a kernel size of 3, and ReLU as the AF. Then, max pooling with a size of 2 is used to extract temporal features. The features extracted are flattened to be used in the BiLSTM layer. The order dependency between extracted temporal properties is learned by this BiLSTM layer, making it suitable for the categorization of time-series data. The first BiLSTM layer after the flattening layer has a filter size of 100 and uses tanh as the AF. Then, two BiLSTM layers are followed with a filter size of 100 and the AF tanh. A dropout layer with a rate of 0.5 follows after the BiLSTM layers to reduce overfitting. Three fully connected layers are inserted after the BiLSTM layers and the dropout. The three fully connected layers have a size of 200 and use ReLU as the AF. The last layer is the SoftMax layer to classify the 9 classes of emotions. The optimizer used in this model is Adam. The total number of learnable

parameters is 600 in the first 1D-CNN layer, and 45,100 is the number of parameters after the second 1D-CNN layer. In the first BiLSTM layer, the total number of parameters is 160,800. The second and third BiLSTM layers have a total of 240,800 parameters. In the three fully connected layers, the number of parameters is 40,200.

3.3 Pre-processing

To ensure accurate and reliable results, it is essential to properly prepare and organize the dataset records before using them in the created model. The pre-processing stage plays a critical role in enhancing the quality of the signal and improving the model performance. To achieve this, several filters are applied to the signals to remove any noise from the ECG signal. Moreover, to increase the amount of data available for training, ECG augmentation techniques are applied in this thesis. These techniques involve creating variations of the original ECG signals, such as changing the amplitude or frequency, while preserving the underlying physiological information. Additionally, various noise removal filters are adopted to enhance the quality of the ECG signals used in the dataset. These filters help to eliminate any unwanted artifacts or interference that may be present in the signal, thereby improving the accuracy and reliability of the model. Overall, the pre-processing stage is crucial in ensuring the quality and reliability of the dataset records used to train the ECG emotion detection model.

3.3.1 Down-sample and Lowpass

Both channels of the ECG signals have been recorded at the same sample rate of 256 Hz, as shown in Figure 6. However, to generate a lower measurement for the ECG signals that can be used as input to the model, the signals were down-sampled to 128 Hz, as illustrated in Figure 7. Furthermore, to increase the amount of data available for training, data augmentation techniques were applied after the down-sampling stage. Specifically, various lowpass filters were adapted to the

signals to remove high-frequency noise and enhance the quality of the ECG signals. The primary function of the lowpass filter is noise removal, which is a crucial data-cleaning process. The first lowpass filter used was the Heartpy library lowpass filter (Gent, Farah, Nes, & Arem, 2019) with a cutoff frequency of 60 Hz to remove high-frequency noise from the ECG signals. The second lowpass filter applied was the Butterworth lowpass filter (Butterworth, 1930) with a cutoff frequency of 60 Hz. Additionally, an FIR lowpass filter (Bhaskar & Uplane, 2016) with a cutoff frequency of 60 Hz was adopted, along with an elliptic lowpass filter (Thompson, 2014) with a cutoff frequency of 60 Hz. After applying the lowpass filters, the signals were segmented into 1-second intervals, as shown in Figure 8. This segmentation step helps to ensure that the ECG signals are analyzed at a specific time interval, which is necessary for the accurate detection of emotions. Although there are previous studies that applied down-sample before lowpass filters to ECG signals such (Dar , Akram , Khawaja, & Pujari , 2020) and (Etemad & Sarkar, 2020). Down-sampling the signals reduces the computational complexity of the filtering process, which can make it faster and more efficient. Also, helps to reduce the amount of high-frequency noise in the signals, which can improve the effectiveness of the lowpass filters. By reducing the sample rate, the signals are effectively low-pass filtered, which removes high-frequency components that can interfere with the filtering process. This can result in better noise reduction and a higher signal-to-noise ratio. Overall, the application of lowpass filters and segmentation techniques are crucial for enhancing the quality and reliability of the ECG signals used in the dataset and improving the accuracy and performance of the ECG emotion detection model.

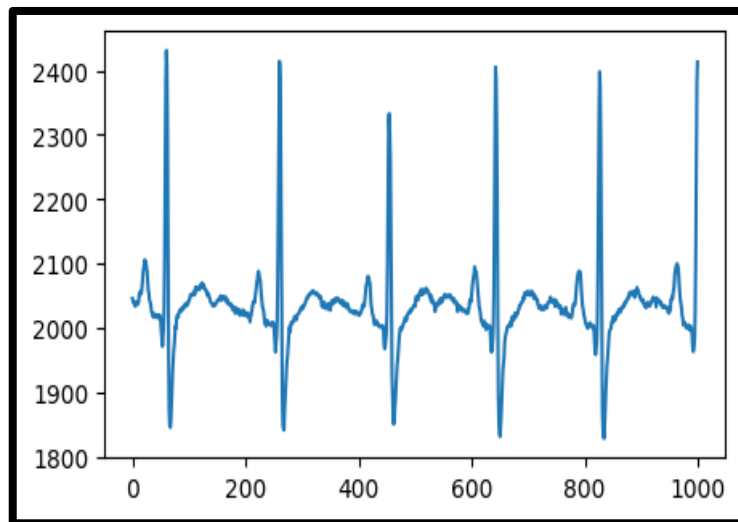


Figure 6: Raw ECG signal with 256 Hz

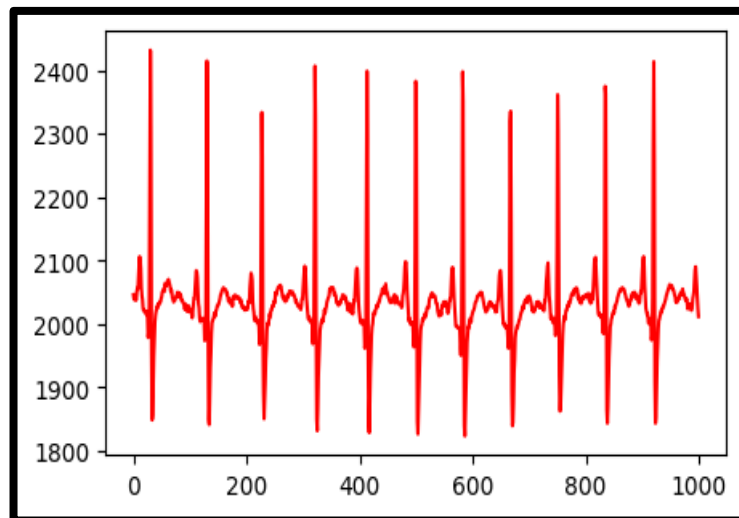


Figure 7: ECG signal after down sampled to 128 Hz.

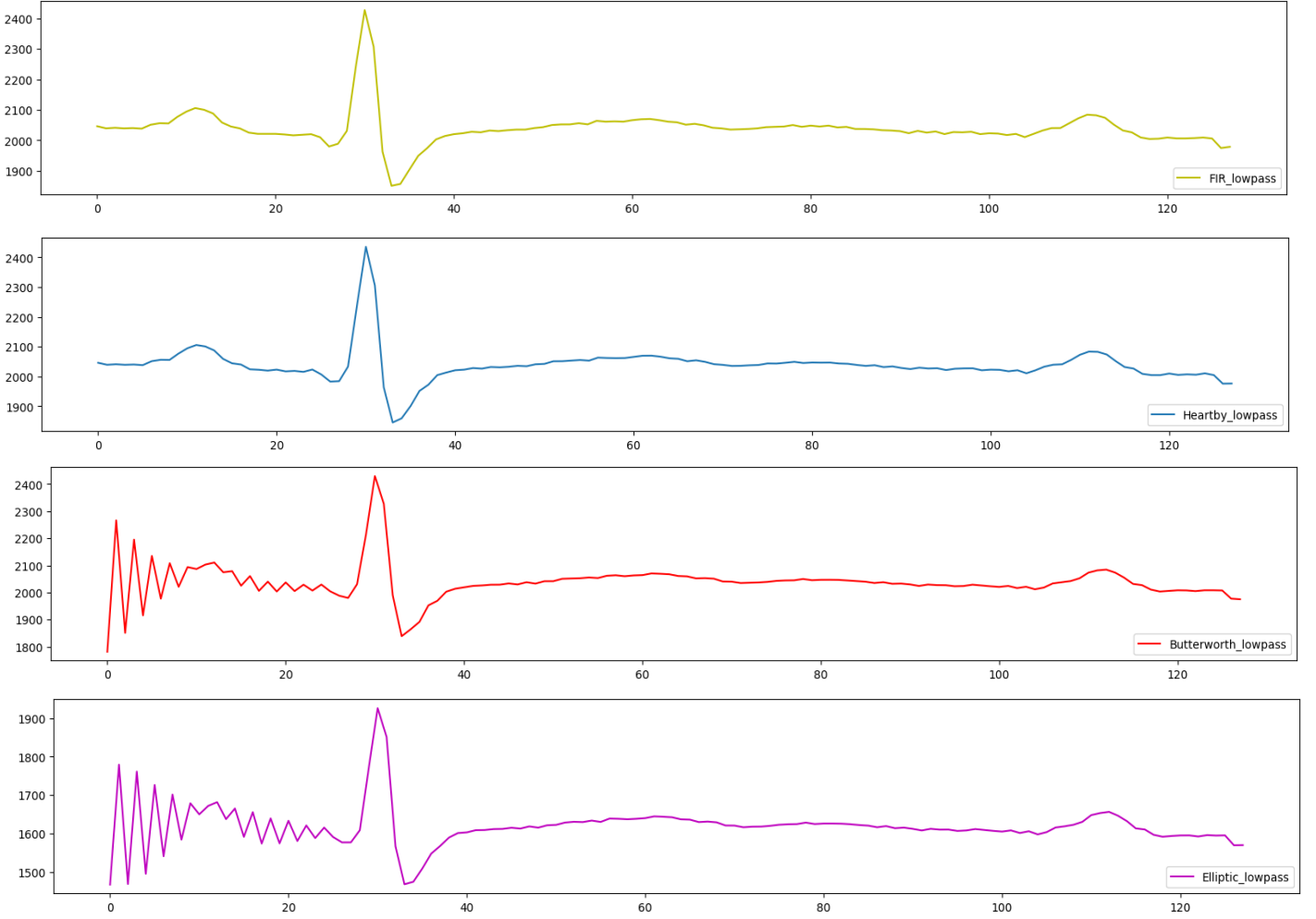


Figure 8: Signals after applying lowpass filters and segmented to 1s

3.3.2 Baseline Removal

The Dreamer dataset consists of both stimulus and baseline signals. The baseline signals are recorded without any stimulus, and their length is 57 seconds, during which the participants are influenced by any emotion while watching a neutral video. As a pre-processing step, these baseline signals need to be removed from the stimulus ECG signal to ensure accurate emotion detection. The baseline signals are down-sampled to 128Hz, and low-pass filters are applied to remove noise. The same low-pass filters used for the stimulus signal are also applied to the baseline signals. Then, the baseline signal is segmented into 1-second

intervals, and the average of the baseline segments is calculated for each signal using Eq. (1).

$AverageBaseline = \frac{1}{S} \sum_{s=1}^S Baseline_s$	(1)
---	-----

Eq. (2) is a mathematical formula used to remove the baseline effect from the ECG signal segments. It involves subtracting the average of the baseline signal from each ECG signal segment. This step is crucial to ensure accurate emotion detection and improve the performance of the ECG emotion detection model. By removing the baseline effect, the emotional response of the participant to the stimulus can be more accurately captured and analyzed. The resulting signal segments represent the emotional response of the participant to the stimulus without any confounding effects from the baseline activity. Overall, Eq. (2) is an important step in the pre-processing stage of the ECG signals and plays a critical role in ensuring the accuracy and reliability of the ECG emotion detection model.

$SignalBl = Signal_s - AverageBaseline$	(2)
---	-----

The pre-processing steps discussed are crucial for removing the baseline effect from both ECG channels and enhancing the emotional effect of the corresponding stimuli. For the DREAMER dataset, the baseline signal was computed from the mean of all 57 segments of baseline activity. This ensures that the baseline signal is representative of the overall activity during the baseline period.

To provide a visual representation of the baseline signal, Figure 9 demonstrates ten baseline segments for the left channel and the average baseline segment. The segments illustrate the variability and noise present in the baseline activity, which can interfere with the accurate detection of emotions. However, by calculating

the average of all baseline segments, a more robust representation of the baseline activity can be obtained, which can then be used to remove the baseline effect from the ECG signal segments.

After removing the baseline signal, the resulting ECG signal segments represent the emotional response of the participant to the stimulus without any confounding effects from the baseline activity. Figure 10 illustrates the left channel signal after removing the baseline signal. It is evident that the baseline signal has been effectively removed, and the resulting signal segments capture the emotional response of the participant to the corresponding stimuli accurately. Overall, these pre-processing steps are crucial for enhancing the accuracy and reliability of the ECG emotion detection model and ensuring the accurate detection of emotions in the DREAMER dataset.

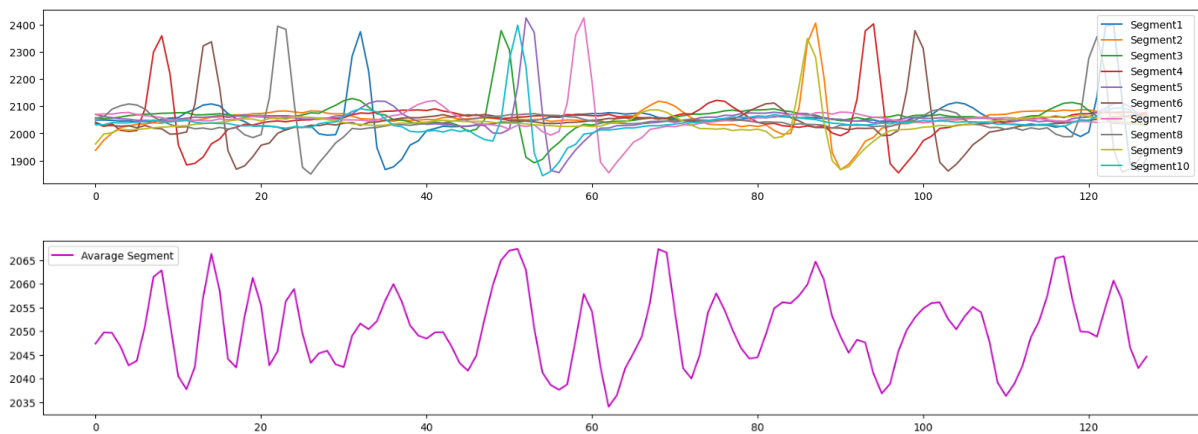


Figure 10: Baseline Signal segments and Average

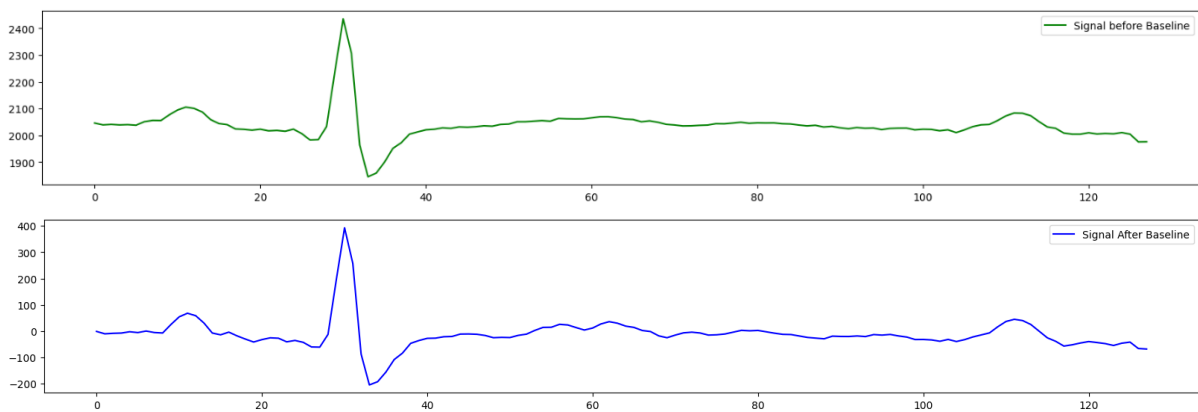


Figure 9: Signal before and after remove baseline

3.3.3 Z-Score Normalization

After removing the baseline signal from the ECG signal segments, the next step in the pre-processing stage is to normalize the signals using Z-score normalization. This step is critical for adjusting the signals to be used as model inputs and ensuring that they are in a standardized format.

Z-score normalization involves converting the signals to a standard scale with a mean of zero and a standard deviation of one, using Eq. (3). This normalization process allows for the comparison of signals with different scales and ensures that each signal contributes equally to the neural network model. Once the ECG signals have been normalized, they are prepared to be used as a set of inputs to the neural network, as shown in Figure 11. This ensures that the signals are in a suitable format for the model to process and enables accurate detection of emotions.

Overall, the Z-score normalization step is a critical aspect of the pre-processing stage, enabling the signals to be standardized and ensuring that they can be used as inputs to the neural network model. This normalization process plays a crucial role in enhancing the accuracy and reliability of the ECG emotion detection model.

$Zscore(x) = (x - \mu)/\sigma$	(3)
--------------------------------	-------

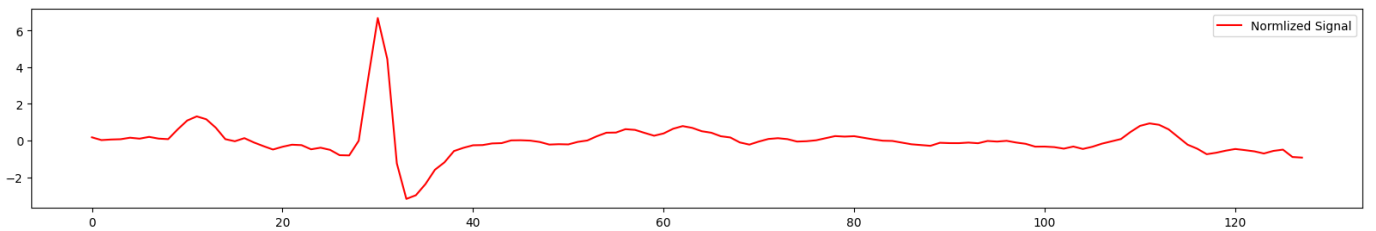


Figure 11: Signal after applying baseline removal and Z-score normalization.

3.4 Convolutional Neural Network

There are a variety of DL models, some of which include the supervised neural network, CNN, and RNN, while unsupervised models include the self-organizing map and autoencoders. A large number of individuals are now using a wide variety of models throughout the majority of sectors, particularly during the most recent development of automation technology.

This thesis proposed the first part of the model, which is 1DCNN supervised model, which is used to achieve the thesis objective of enhancing emotion recognition using the ECG signals. The CNN is a DL model that is capable of determining the communications between the inputs and outputs. The CNN consists of various layers, and these layers may have pooling layers between them. The CNN layers have mathematical techniques to examine the model inputs such pre-processed signals.

Applying CNN has many advantages, as it can determine the fundamental data features. The limited number of connections, as well as the possibility of exchanging parameters locally, are two of the benefits. The term ‘sparsity’ refers to the fact that the output of each layer only relies on a small number of inputs rather than the whole dataset. When all of the neurons in a feature map have the same weights, it is possible to achieve very efficient calculations thanks to the sharing of the parameters.

The CNN layers can consist of pooling layers. The pooling layer has various types, but the most commonly used one is the max pooling layer used in the proposed model. The CNN layer has a kernel that moves to extract the data features. The pooling layer is added before the CNN layer. The purpose of the max pooling layer is to eliminate the features that will be transferred to the next convolution layers, as shown in Figure 12.

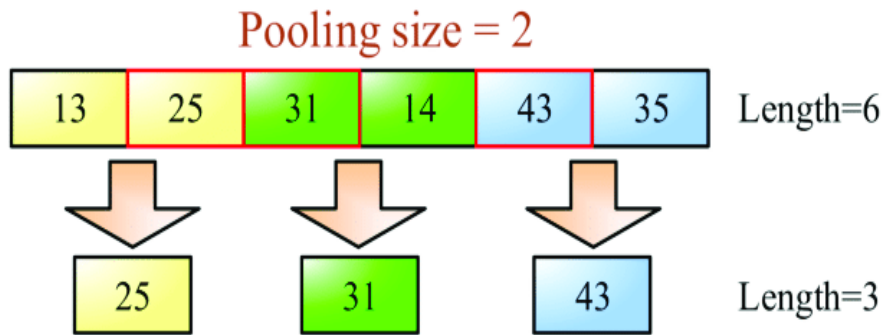


Figure 12: Example of Max pooling (Kuo & Huang, 2018)

3.5 Bidirectional LSTM

LSTM is a refined variant of regular RNNs that excels in representing long-term dependencies in time series datasets. LSTM achieves this by using a specialized memory cell that is purpose-built for this task. Unlike traditional RNNs, LSTM provides an appropriate solution to the issue of vanishing gradients, which can be problematic in conventional RNNs.

LSTM uses multiple memory cells that are capable of capturing the long-term relationships present in sequential datasets. These memory cells replace the recurrent units' structure in the network. In LSTM, current recurrent neuron states are dependent on both the current input and the previous recurrent neuron states. An LSTM cell consists of four gating units: the input gate, output gate, forgetting gate, and self-recurrent memory units. These gating units control the interactions between different memory cells. For example, the forgetting gate in the hidden cell decides whether or not state information from the preceding time steps will be remembered or forgotten. The input gate determines what pattern of input vectors should be fed into the memory cell state, while the output gate decides whether or not to modify the states of other memory cells.

However, the typical LSTM has the potential to lose essential feature information during the training phase because it only assesses input vectors along one specific direction. As a result, the sequence information cannot be accurately analyzed in its entirety.

LSTM is a refined variant of RNNs that excels in representing long-term dependencies in time series datasets. It uses specialized memory cells and gating units to capture and control the interactions between different memory cells. However, the fact that it only analyzes input vectors along one specific direction during the training phase can lead to the loss of essential feature information.

To address the issue of losing essential feature information during the training phase, a new structure was developed that could capture time series data representations in both forward and backward orientations. This structure is known as the BiLSTM model. The concept of allowing neural networks to store sequence information in both directions, either backward or forward, is the hallmark of BiLSTM. This technology is represented in Figure 14.

BiLSTM differs from the regular LSTM model in that it processes the input in both directions simultaneously. This bidirectional mode allows BiLSTM to store information relating to both the past and the future at any given time step. As a result, BiLSTM can capture a larger context of information and is better able to preserve essential feature information than a regular LSTM. This is the key differentiator between the two types of LSTMs.

In essence, BiLSTM can be thought of as two LSTMs working in tandem, one processing the input in the forward direction and the other in the backward direction. This enables BiLSTM to capture the temporal dependencies in both directions, making it an ideal choice for tasks like speech recognition, where the context of both past and future information is essential.

In summary, BiLSTM is a technology that allows neural networks to store sequence information in both directions, thereby capturing a larger context of information and preserving essential feature information. The ability to process input in both directions sets BiLSTM apart from regular LSTM models and

makes it an ideal choice for tasks that require the context of both past and future information.

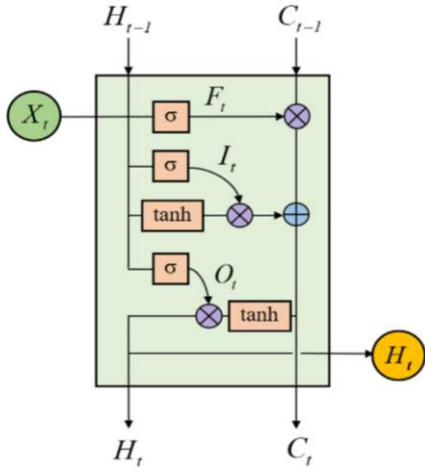


Figure 13: The classical LSTM architecture (Wang, Zhang, Liang, & Liu, 2023)

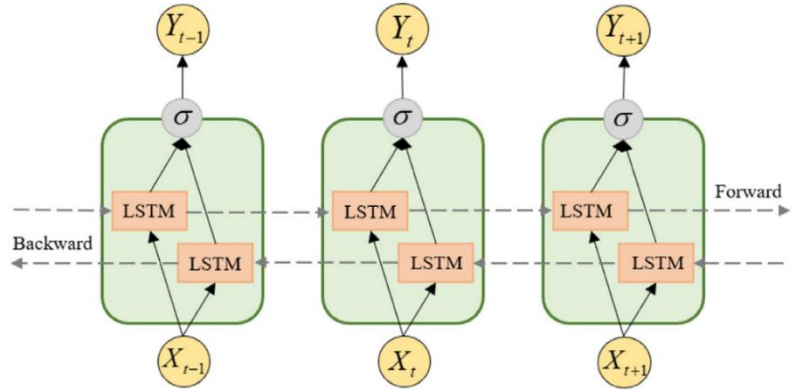


Figure 14: The BiLSTM architecture (Wang, Zhang, Liang, & Liu, 2023)

3.6 Fully Connected Layer

After the CNN and BiLSTM layers have been applied, the fully connected layers come into play. The results from the previous layers are processed in the fully connected layers, where the extracted features are combined to complete the connections between nodes. It is worth noting that the fully connected layers contain the majority of a DNN's parameters, making them a crucial component in achieving optimal performance.

Despite the widespread use of hyperparameters, such as the number of fully connected layers and the number of neurons within them, there has been little research on how to improve performance using these variables. This may be due to the enormous number of trainable parameters in the fully connected layers, which can lead to overfitting. Overfitting occurs when a model becomes too

complex and starts to fit the training data too closely, resulting in poor performance on new, unseen data.

To avoid overfitting, it is essential to strike a balance between the number of trainable parameters and the model's ability to capture the underlying patterns in the data. Regularization techniques, such as dropout and weight decay, can be used to prevent overfitting by reducing the complexity of the model and discouraging the co-adaptation of neurons.

In summary, the fully connected layers are a crucial component of DNNs, as they contain the majority of trainable parameters and are responsible for fitting sophisticated nonlinear discriminant functions in the feature space. However, overfitting can occur due to a large number of parameters, making it important to strike a balance between model complexity and performance. Regularization techniques can be used to prevent overfitting and improve generalization performance. Further research is needed to explore how hyperparameters can be optimized to further improve the performance of DNNs.

3.7 Activation functions

AFs play a critical role in neural networks by generating input and bias weights and determining whether a neuron will fire. After applying gradient processing, such as gradient descent, to the data, the modified data and the neural network's parameters are outputted. These AFs are commonly referred to as "transfer functions." The linearity or non-linearity of an AF depends on the function being activated. Proper selection of the AF can improve the performance of neural network computing, as demonstrated in object recognition and classification domains.

In a linear model, an affine transformation converts a function of the input to a function of output, such as in hidden layers that come before the class score prediction for each label. However, AFs are needed because linear

mappings produced by neural networks cannot perform pattern recognition. AFs convert linear outputs into non-linear outputs, allowing for further processing, such as pattern recognition, to occur. These functions are used to transform the outputs of linear models into non-linear outputs, aiding DNNs in learning high-order polynomials.

Non-linear AFs must be differentiable for the backpropagation process of a DNN to succeed. Hidden layers in DNNs consist mostly of convolutional and pooling layers. Convolutional layers analyze data from previous layers to identify local patterns and traits while pooling layers merge semantic properties. The output layer, which receives training information for the neural network, categorizes or predicts the network's outcomes using AFs. The AFs' network placements are determined by their functions.

In summary, AFs are crucial in neural networks for generating input and bias weights and converting linear outputs into non-linear outputs. Proper selection of the AF can improve the network's performance in various domains. Hidden layers in DNNs consist mainly of convolutional and pooling layers, while the output layer uses AFs to categorize or predict the network's outcomes.

3.7.1 Softmax Function

The AF that may be used in neural computing is known as the Softmax function. It is put to use in the process of deriving probability distribution from a vector containing real values. The result of the Softmax function is a range of numbers that may be anywhere from 0 to 1, with the overall probability being equal to 1. In order to calculate the Softmax function (Goodfellow, Bengio, & Courville, 2016), we will use the relationship. When used in multi-class models, the Softmax function will return the probabilities of each class, with the probability of the target class being the greatest of all the values it returns. The Softmax function is used in practically all of the output layers of the DL architectures (Krizhevsky,

Sutskever, & E. Hinton, 2017), (Badrinarayanan, Kendall, & Cipolla, 2017), and (Lin, Chen, & Yan, 2013), where it may be found most often. The use of the Sigmoid function in binary classification, on the other hand, is where the Softmax function shines since it is employed for multivariate classification, as shown in Eq. (4).

$f(g_i) = \frac{\exp(g_i)}{\sum_j \exp(g_j)}$	(4)
---	-----

3.7.2 ReLU Activation Function

Since its introduction in 2010, when Nair and Hinton presented it, the ReLU AF has been the most popular choice for usage in DL applications, yielding the most cutting-edge findings to this day (Nair & Hinton, 2010). Since its introduction in 2010, the ReLU AF has been the most popular choice for usage in DL applications. The ReLU is an AF that is able to learn more rapidly (LeCun, Bengio, & Hinton, 2015), and it is the function that has been shown to be the most effective and is used the most often (Ramachandran, Zoph, & Le, 2017). It offers greater performance and generalization in DL in comparison to the Sigmoid and tanh AFs (Zeiler, 2013) and (Dahl, Sainath, & Hinton, 2013). The ReLU is very close to being a linear function, and as a consequence, it preserves the characteristics of linear models that make it straightforward to optimize them via the use of gradient-descent strategies (Goodfellow, Bengio, & Courville, 2016). As illustrated in Eq. (5) and figure 15, using the ReLU AF, a threshold is applied to each input, which causes values that are larger than zero to be reset to zero. This can be seen in the context of Eq. (5). As a consequence of this, the ReLU may be stated using the following expression: Inputs with negative values are normalized using this function, which compels such values to be zero and

solves the problem of vanishing gradients that were encountered in earlier types of the AF.

Voice recognition programs (Maas, Hannun , & Ng , 2013) and item categorization programs (Krizhevsky, Sutskever, & E. Hinton, 2017) are two examples of typical applications of the ReLU function. In both applications, in DNNs, the ReLU function is used in the hidden units in combination with another AF. The primary advantage of computing with ReLUs is that they ensure quicker processing because they do not calculate exponentials or divisions, which results in an overall increase in the speed of computation (Zeiler, 2013). This is the primary benefit of using ReLUs in the calculation. This is the primary benefit that comes with making use of ReLUs. The ReLU has another benefit, which is that it produces sparsity in the hidden units. This feature is in addition to the fact that it compresses the values between zero and maximum. This is because of the compression that took place before. In contrast to the sigmoid function, however, the ReLU is prone to overfitting, which is a downside of this particular model. The ReLU has a problem in that it easily overfits, despite the fact that the dropout strategy has been used to lessen the impact of overfitting of ReLUs and that the corrected networks have enhanced the performances of the DNNs (Glorot, Bordes, & Bengio, 2011).

Despite these two facts, the ReLU still has a problem with overfitting. The ReLU and variants of it have been included in several DL architectures, as noted by (Nair & Hinton, 2010). This is likely owing to the ReLU's simplicity and dependability in practice. Other examples of such software include CNN designs and limited Boltzmann machines (Nair & Hinton, 2010). The ReLU suffers from the significant limitation that it may be fragile at times while it is being trained, which results in the death of a few of the gradients. This makes it difficult to get accurate results. This, in turn, causes the death of certain neurons, which makes it so that the weight updates do not activate in subsequent data points. This, in

turn, makes it difficult to learn new things since dead neurons supply no stimulation (Goodfellow, Bengio, & Courville, 2016). This is due to the fact that dead neurons provide no stimulation.

$f(g) = \max(0, g) = \begin{cases} g_i, & \text{if } g_i \geq 0 \\ 0, & \text{if } g_i < 0 \end{cases}$	(5)
---	-------

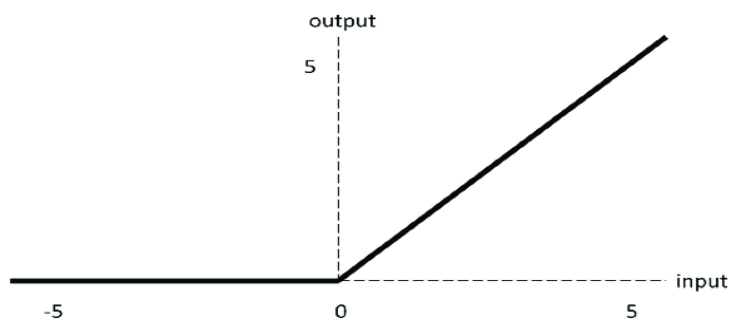


Figure 15: The ReLU activation function presentation (Sultan, Salem, & Al-Atabany, 2019)

3.7.3 Tanh Activation Function

One further kind of AF that may be used in DL is known as the hyperbolic tangent function, and there are many forms of this function that can be utilized in DL applications. The hyperbolic tangent function, often known as the tanh function, is a smoother zero-centered function whose range falls between -1 and 1 (LeCun, Bengio, & Hinton, 2015) as shown in Eq. (6). As compared to the sigmoid function, the tanh function is considered to have superior performance when it comes to the training of multi-layer neural networks (Karlik & Olgac, 2011) and (Neal, 1992). As a result, this function is now considered to be the function of choice. Unfortunately, the tanh function was not able to tackle the issue of a vanishing gradient, which was also experienced by the sigmoid function. The fact that the function generates output with a zero-centered zero point is the primary benefit it offers since it makes the process of back-propagation easier. One of the

characteristics of the tanh function is that it is unable to achieve a gradient of 1 except in the case in which the value of the input variable, denoted by x , is equal to zero. Hence, this results in the tanh function producing some dead neurons while it is being computed. The situation known as the "dead neuron" is one in which the activation weight is almost never employed due to the presence of zero gradient. Because of this shortcoming of the tanh function, more investigation into AFs was undertaken to find a solution to the issue; It inspired the creation of the ReLU AF. Most of the time, RNN have been used to build the tanh functions, typically for NLP (Dauphin, Fan, Auli, & Grangier, 2017) and ASR (Maas, Hannun, & Ng, 2013) applications.

$f(g) = \frac{\exp^g - \exp^{-g}}{\exp^g + \exp^{-g}}$	(6)
--	-----

3.8 Conclusion

In this chapter presented a DL model that utilizes both BiLSTM and CNN layers to recognize emotions using ECG signals. The thesis discussed the importance of preprocessing ECG signals and how used four different lowpass filters for data augmentation. Also explained the architecture of our model, which includes a BiLSTM layer to capture the temporal features of the ECG signals and a CNN layer to extract the spatial features. Additionally, discussed the AF used in our model, which was the ReLU, Tanh, and softmax. This AF has been shown to be effective in deep-learning models and is commonly used in CNNs and LSTMs. Overall, our study contributes to the field of affective computing by demonstrating the potential of using DL models with BiLSTM and CNN layers for emotion recognition using ECG signals. Our results show that the combination of these layers can effectively capture both temporal and spatial features in the ECG signals, leading to high accuracy in emotion recognition.

Chapter 4

Experimental setting, Result and Discussion

4.1 Introduction

4.2 Data Collection

4.3 Evaluation Metrics

4.4 Results and Decision

4.5 Conclusion

Chapter 4: Experimental setting, Result and Discussion

4.1 Introduction

In order to construct an accurate system that has the potential to perform in a variety of real-life applications, it is essential to have a collection of signals that have been well-structured. (Pinto, 2017) came up with a list of characteristics that need to be considered in order to create a reliable database. Some of these elements include the sampling frequency, the health of the individuals, the number of electrodes used, and the location of the electrodes. Both the performance of the system and the location of the electrodes may be directly influenced by the number of electrodes present in the system. (Zhang & Wei, 2006) came to the conclusion that the quality of the data would be worse if the electrodes were positioned at a greater distance from the heart.

The sampling frequency turns out to be significant as well due to the fact that lower sampling frequencies result in less information being lost during the process of converting the analog signal from the heart to its digital equivalent. ECG variability may be caused by a variety of factors, including the health of the participants (because heart disease is one of these factors) and the number of subjects. When solving a problem using ML or DL, having access to a huge amount of data and a wide variety of topics is crucial. This is why signal collections with a high subject number are so useful.

In terms of emotion recognition, having a large number of subjects will make it possible to separate those subjects into training and testing sets. This will enable the problem of subject dependence to be solved by having a set of test subjects that the algorithm has not seen before. When it comes to tasks involving emotion identification, the label reliability is also a crucial factor to consider. This is because the performance of the algorithm depends on how effectively it is trained.

On the other hand, it might be challenging to conduct a quantitative analysis of one's emotional states. In a database with good organization, the amount of records is another important element to take into consideration. The availability of a large quantity of data is essential for both ML and, more specifically, DL algorithms. Overfitting is more likely to occur when the model predicts the training data accurately but can not predict the new data. That can happen if the number of records or the amount of data is few which the model does not have enough data to train.

In addition, tiny databases, which have a lower quantity of data as well as a reduced diversity of signals, may also contribute to additional issues such as signal and subject dependence. In conclusion, some factors, like the settings for the acquisition, are also significant and need to be in conformity with the aim of the study. Both the kind of sensor and the electrodes that are used are crucial components in the process of emotion identification. Wearable sensors and a limited number of electrodes, both of which ensure the collection of quality signals, make it possible for the people engaged to acquire data in a manner that is both simpler and more pleasant.

This chapter consists of a discussion of the experimental setting and the results of the proposed model. The first section of the chapter is an explanation of the used dataset. Next, the used evaluation matrices were discussed. At the end, the results of the proposed model were explained and compared with other studies results.

4.2 Data Collection

Dreamer dataset is used in this study to detect emotions using EEG and ECG signals (Katsigiannis & Ramzan, 2018). The participants of the dreamer dataset were exposed to aural and visual stimuli in the form of film clips in order to elicit emotional responses from them and capture EEG and ECG data. A dataset of 18

film clips that were chosen and analyzed for their ability to trigger different emotions. These film clips include extracted sequences from a variety of movies that have been shown to elicit a broad spectrum of feelings in viewers. From among these 18 video clips, there were two examples of each of the following nine feelings: amusement, excitement, happiness, peacefulness, anger, disgust, fear, and surprise, as demonstrated in Table 2.

Table 2: Dataset Videos and targeted emotions

Video	Emotion	Video	Emotion
Searching for Bobby Fischer	calmness	The Fly	disgust
D.O.A.	surprise	Pride and Prejudice	calmness
The Hangover	amusement	Modern Times	amusement
The Ring	fear	Remember the Titans	happiness
300	excitement	Gentlemen's Agreement	anger
National Lampoon's Van Wilder	disgust	Psycho	fear
Wall-E	happiness	The Bourne Identity	excitement
Crash	anger	The Shawshank Redemption	sadness
My Girl	sadness	The Departed	surprise

The duration of the film snippets ranged from 65 to 393 seconds, which is thought to be sufficiently given that, according to the research carried out by psychologists (Katsigiannis & Ramzan, 2018), visual stimuli that last between one and ten minutes are capable of generating individual emotions. The duration of the film snippets ranged from 65 to 393 seconds. Despite this, it is possible for a person's emotional state to change over the course of time, especially when lengthier video stimuli are used. For the purpose of this investigation, only the

recordings that were made during the last minute and a half of each video clip were considered further.

This was done in order to protect the data recorders from being influenced by a variety of different emotions. The research was carried out in a quiet location with very specific lighting conditions in order to reduce the likelihood that the findings of the investigations would be influenced by elements from the outside world. The room was made completely black by using an electric curtain, and the video excerpts were played back on a television that had speakers integrated directly into it. The EEG was recorded at a sample rate of 128 Hz using an Emotiv EPOC system. This system makes use of 16 gold-plated contact sensors that are mounted to flexible plastic arms of a wireless headset. These sensors were placed against the head in areas that correspond to the following locations according to the International 10-20 system: AF3, F7, F3, FC5, T7, P7, O1,O2, P8, T8, FC6, F4, F8, AF4, M1, and M2; AF The mastoid sensor that was located at M1 was used as a ground reference point for the purpose of comparing the voltage of each of the other sensors.

On the other hand, the mastoid sensor that was located at M2 was used as a feed-forward reference for the purpose of reducing the amount of electrical interference from the outside. This resulted in the data from the other 14 contact sensors being collected and used in the future step of feature extraction. This was a direct consequence of the previous operation. An ECG was recorded at 256 hertz using the SHIMMER wireless sensor as the recording device. The study included 25 healthy volunteers ranging in age from 22 to 33 years old. These participants all volunteered to take part. As a result of their involvement, we were able to evaluate the emotions that were sparked by the video clips as well as gather data on their physiological responses. The volunteers were first required to read an information sheet that provided an explanation of the rating scales that

would be utilized for the evaluation of their level of effect, and then they were asked to sign a consent form before taking part in the experiment.

The information sheet described the process that would be followed during the experiment, and it was also required that the volunteers read it before taking part in the experiment. It is not something that comes readily to individuals to classify emotions according to a scale that includes valence, arousal, and dominance. In order to improve the possibility that the participants would provide correct ratings, the rating scale was defined in great detail, both verbally and via the use of examples. This was done in both of these ways. Following then, the session wouldn't get started unless the attendees showed that they had a sufficient degree of grasp of the rating scales. In addition, there was a member of the team that was in charge of conducting the experiment which was available to answer any questions that were posed by the participants. It was required of the participants that they move as little as possible over the whole of each session, which lasted close to sixty minutes. In addition to this, we asked them to submit an emotional evaluation based on the feelings that they actually felt, as opposed to the emotions that they believed the video clip was supposed to elicit from them. This was done so that we could compare their actual feelings with the emotions that they believed the video clip was supposed to elicit from them.

The second stage of the experiment was seeing a short film clip that had nothing to do with politics. a video clip that, for the sake of establishing baseline signals, is assumed to be devoid of any and all valence. It has been shown that while collecting physiological data, reducing daily dependency may be achieved by the use of neutral signals as baseline signals. As a direct result of this, the neutral clip was shown before each film clip in order to aid the individual in regaining a state that was emotionally neutral. In addition to the creation of the baseline signal, this was also carried out. After seeing each video clip, the participants were given a GUI that they were instructed to utilize in order to assess how each scene made

them feel emotionally. On a five-point scale, they reported their levels of felt arousal, which ranged from uninterested or bored to excited or alert.

They also reported their levels of felt valence, which ranged from unpleasant or stressed to happy or elated, as well as their levels of felt dominance, which ranged from helpless to empowered. The person was able to more readily analyze their degrees of arousal, valence, and dominance by using SAMs, which were used throughout the study. In addition, the data from the participants, as well as the recordings of the physiological data and the evaluations of how the participants were feeling, were preserved in a data structure. This was done in order to facilitate further analysis.

In addition, despite the fact that the researchers made a concerted effort to strictly adhere to the experimental procedure and setup, there were problems with the recordings made from two of the patients as a result of technological difficulties, which led to incomplete data. This was the case despite the fact that the researchers made an effort to strictly adhere to the experimental procedure and setup. As a direct result of this, recordings produced by 23 of the 25 volunteers were finally used in this inquiry due to the unsuitability of the data received from the two individuals that were mentioned earlier in this paragraph. As shown in Table 3, the total number of participants in this study was 23, with 14 males and 9 females making up the final sample for this investigation.

Table 3: Dataset summary

Number of Clip	18
Clip duration	65 - 393 s
Participants	23
Participants (male)	14
Participants (female)	9
Participants (age)	22 - 33
Records for Each Emotion	46

4.3 Evaluation metrics

The performance evaluation of a trained classifier is critical in measuring its generalization capacity or ability to perform well on unseen data. This evaluation is done using various evaluation metrics, which provide a quantitative measure of the quality of the classifier's predictions.

One of the most widely used evaluation metrics is accuracy. This metric measures the proportion of correctly classified instances out of the total number of instances in the test data. The accuracy of a classifier is an important indicator of its ability to generalize well to unseen data.

In addition to accuracy, other assessment metrics may also be used in the process of model selection. Model selection involves choosing the best classifier out of several different models that are trained on the same data. The aim of model selection is to identify the model with the highest potential for future performance when evaluated on unseen data.

During the training phase, different solutions or models are created, and the evaluation metrics are used as a discriminator to choose the best solution among them. The best solution is then evaluated on the test data to determine its generalization performance.

In summary, evaluation metrics play a crucial role in assessing the performance of a trained classifier on unseen data. These metrics are used to measure the accuracy of the classifier's predictions, select the best model, and discriminate among different solutions during the training phase. By using appropriate evaluation metrics, it is possible to build classifiers with high generalization capacity and good overall performance. The used evaluation metrics as followed:

4.3.1 Mean Square Error Loss Function

The MSE assesses the square of the difference between the predicted output and the actual output for each observation. All observations are then summed and divided by the number of observations, as shown in Eq. (7). The MSE is used to determine how well the predictions or projections reflect the current state of the market. The smaller the MSE, the closer the prediction matches the actual values. This statistic is used to evaluate models, and a lower number implies a better fit for the data being analyzed by the regression model.

$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$	(7)
--	-------

4.3.2 Categorical Cross Entropy Loss Function

A loss function is a kind of optimization function that is used in the process of training a classification model. This type of model categorizes the data by estimating the likelihood that the data belongs to one of two classes. When there are two or more output labels, the loss function is used in multi-class classification models as a data analysis tool. The output label has a one-hot category encoding value in the form of 0 and 1, which has been allocated to it. The categorical cross entropy formed as shown in Eq. (8).

$Loss = -\frac{1}{n} \sum_{k=1}^k \sum_{m=1}^m y_m^k \times \log (\hat{y}_m^k)$	(8)
---	-----

By combining MSE with cross-entropy loss function in a multi-class classification task, the model can learn to optimize for both accuracy and confidence in its predictions. The cross-entropy loss function penalizes the model for making incorrect predictions, while the MSE loss function encourages the model to make confident predictions.

4.3.3 Precision

Precision is one of the key performance indicators for a model. Its main purpose is to determine the proportion of true events among the positive predictions made by the model. Precision measures the quality of the positive predictions produced by the model. The term precision is used to describe the ratio of the number of true positives to the total number of positive forecasts, as shown in Eq. (9). True positives refer to the number of correct predictions made by the model, while positive forecasts represent the total number of predictions made by the model that are positive. A high precision score indicates that the model is making accurate positive predictions, while a low precision score suggests that the model is producing a large number of false positive predictions. When evaluating the performance of a model, it is essential to consider both precision and other performance indicators, such as recall and F1 score, to gain a comprehensive understanding of the model's strengths and weaknesses.

$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$	(9)
--	-----

4.3.4 ROC Curve and AUC Curve

A ROC curve is a graphical representation of the true positive rate (TPR) against the false positive rate (FPR) for different cut-off points of a parameter in a binary classification model. Every point on the ROC curve corresponds to a sensitivity and specificity pair that is associated with a different decision threshold.

The AUC is a metric that measures the overall performance of the model across all possible threshold values. The AUC curve measures the area under the ROC curve, and it is evaluated by comparing the ratios of true positive and false positive rates. If the ratio of true positives to negatives is high, the AUC curve value will be closer to 1, indicating better model performance.

The ROC curve and AUC are powerful tools for evaluating the performance of classification models. They provide a comprehensive understanding of the model's performance across different threshold values and help identify the strengths and weaknesses of the model.

4.3.5 F1 Score

The F1 score is an evaluation metric that measures a model's prediction ability by considering its class-wise performance rather than overall performance, like accuracy. It integrates precision and recall, which are two competing variables that measure the quality of positive predictions and the ability to correctly identify true positives, respectively. The F1 score is expressed mathematically by the harmonic mean of precision and recall in Eq. (10), and a higher score indicates better performance. The F1 score is particularly useful in situations where precision and recall are equally important, such as fraud detection. It can also be extended to multi-class classification problems by calculating the F1 score for each class separately and then taking a weighted average. The F1 score is a widely used evaluation metric in ML and statistical analysis, providing a

balanced measure of a model's precision and recall and a useful tool for evaluating classification model performance.

$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}$	(10)
---	------

4.3.6 Confusion Matrix

The confusion matrix is a valuable tool in evaluating classification model performance. It presents the model's performance in a tabular form by comparing predicted labels with actual labels and is divided into four sections: true positives, false positives, true negatives, and false negatives. These sections help identify the model's strengths and weaknesses. While the matrix itself is easy to understand, the terminology used with it can be confusing. Understanding the terminology is important for gaining insights into model performance. The confusion matrix is sometimes called an error matrix due to its ability to present a model's errors in matrix form.

4.3.7 Accuracy

In classification problems, accuracy is a measure that indicates the proportion of times that a model has correctly predicted the outcome of the issue. In the field of DL, an accuracy score is a kind of assessment metric that compares the number of accurate predictions generated by a model to the total number of predictions generated by the model. To determine it, we first take the total number of forecasts and divide it by the number of predictions that came true.

4.4 Results and Decision

My thesis represents a novel approach to emotion recognition using ECG signals, distinct from previous studies in several ways. Firstly, while other studies measured emotions on a valence-arousal scale, the proposed methodology involves a different approach to classifying ECG signals. Specifically, each ECG signal is classified according to the emotions it represents and subsequently categorized into 9 distinct classes, as detailed in Table 4. This approach stands in contrast to previous studies, which typically classified ECG signals based on their valence and arousal, resulting in four classes: high valence-high arousal, high valence-low arousal, low valence-high arousal, and low valence-low arousal, as illustrated in Figure 16. Overall, this thesis offers a novel and innovative perspective on emotion recognition using ECG signals, representing a significant contribution to the field.

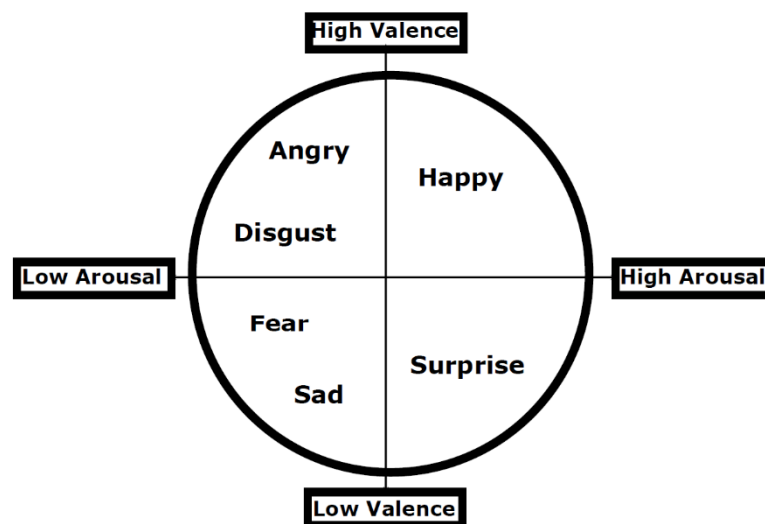


Figure 16: Valence and Arousal emotion distribution (Dar , Akram , Khawaja, & Pujari , 2020)

Table 4: Emotions Category

ECG signal	calmness	surprise	amusement	fear	excitement	disgust	happiness	anger	sadness
Detected signal(1)	1	0	0	0	0	0	0	0	0
Detected signal(2)	0	1	0	0	0	0	0	0	0
Detected signal(3)	0	0	1	0	0	0	0	0	0

The second major difference between the proposed thesis and previous research lies in the method of ECG signal detection. Unlike other studies, which utilized separate sensors for each arm and treated the resulting entities as distinct inputs to the model, this thesis employed two sensors - one on the right arm and the other on the left arm - and combined their outputs before feeding them into the model. Additionally, the created model was evaluated on each sensor's ECG signals individually. The pre-processing phase also stands out from previous research, as this thesis introduced data augmentation techniques to increase the amount of data available in the dreamer dataset. Various filters were applied not only to the ECG signal stimulus but also to baseline signals, resulting in a more robust and comprehensive pre-processing approach. In contrast, previous studies typically applied only one filter to the ECG signal stimulus and did not filter the baseline signals. Overall, these differences set the proposed thesis apart from previous research and contribute to its innovative and ground-breaking approach to emotion recognition using ECG signals.

4.4.1 Model Results

In this section, present the results of our model evaluation, which is an essential component of assessing the model's performance and applicability. The proposed model began by examining the performance of left and right sensor ECG signals

separately, as well as on merged sensor data. Table 5 provides a comprehensive overview of the model's accuracy, precision, and f1 score for each sensor configuration, highlighting the strengths and weaknesses of each approach.

Interestingly, the results showed that the accuracy of the right sensor was slightly better than that of the left sensor, with an accuracy rate of 87.5% and 86.9%, respectively. However, the highest accuracy rate was achieved when utilizing merged sensor data, which yielded an impressive accuracy rate of 98.6%. The use of merged sensor data not only improved the model's accuracy but also increased the amount of data used in the model, resulting in a more robust and comprehensive performance.

In terms of precision, it was observed that the precision rate of the right sensor was slightly better than that of the left sensor, with rates of 88.7% and 88.2%, respectively. However, the highest precision rate was achieved when utilizing merged sensor data, with an outstanding precision rate of 98.7%. This result highlights the importance of utilizing merged sensor data to achieve the highest possible precision in emotion recognition using ECG signals.

Also, assessed the f1 score of each sensor configuration, with the left sensor achieving the lowest score of 87.2%. The right sensor achieved a slightly higher f1 score, with a difference of 0.5%. Once again, the highest f1 score was achieved when utilizing merged sensor data, with an impressive f1 score of 98.6%.

Overall, these findings demonstrate the superiority of utilizing merged sensor data in emotion recognition using ECG signals, resulting in enhanced accuracy, precision, and f1 score. Furthermore, this thesis emphasizes the importance of comprehensive pre-processing techniques and data augmentation to improve the performance of ML models in emotion recognition using ECG signals.

Table 5: Model Evaluation Results

Modality	Algorithm	Accuracy	Precision	F1 score	AUC
ECGL+ECGR	CNN-BiLSTM	98.6%	98.7%	98.6%	99.7%
ECGL	CNN-BiLSTM	86.9%	88.2%	87.2%	96.5%
ECGR	CNN-BiLSTM	87.5%	88.7%	87.7%	96.8%

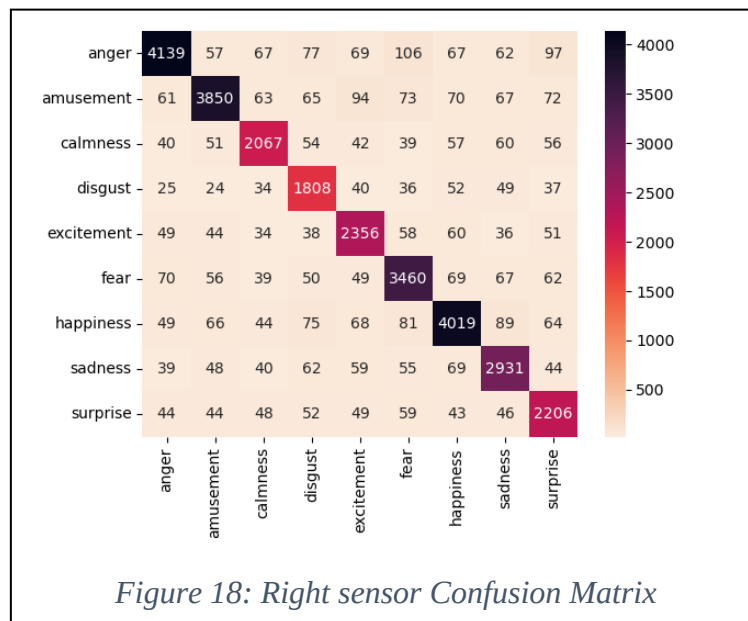
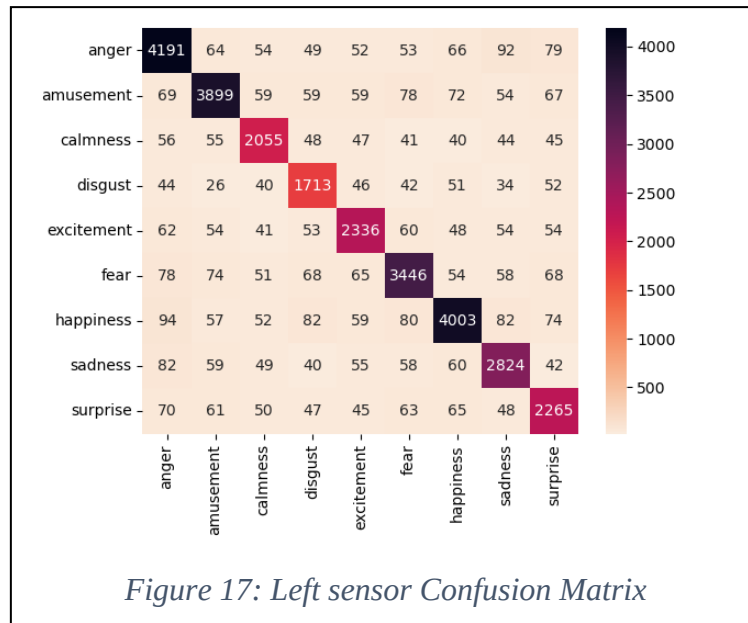
Figure 17 presents the left sensor confusion matrices, demonstrating that the model performed best in recognizing the emotions of anger, amusement, and happiness. However, the lowest recognition rate was for disgust, as indicated in the figure. Figure 18 illustrates the right sensor confusion matrices, showing similar results to the left sensor, with anger, amusement, and happiness being the best-recognized emotions and disgust being the least-recognized emotion.

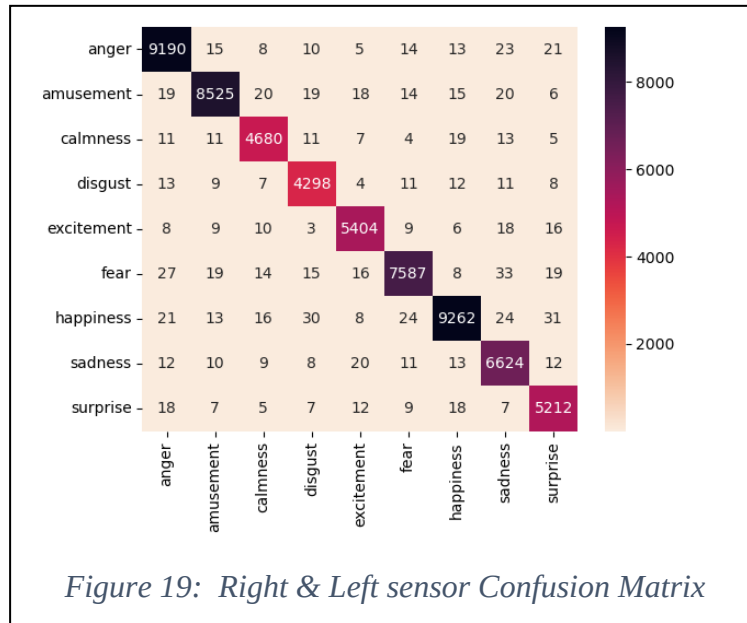
Interestingly, the left sensor outperformed the right sensor in detecting the emotions of anger, amusement, and surprise, as demonstrated in Figures 17 and 18. This discrepancy highlights the importance of evaluating each sensor configuration separately to gain a more comprehensive understanding of the model's performance.

However, the merged signal configuration stood out as having the best accuracy and emotion detection, as illustrated in Figure 19. The merged signal configuration achieved the highest accuracy rate, as well as the best detection of all the emotions evaluated. This result is particularly noteworthy, as it demonstrates the added value of utilizing combined sensor data in emotion recognition using ECG signals.

Overall, the confusion matrices provide a nuanced evaluation of the model's performance, highlighting the strengths and weaknesses of each sensor configuration and emphasizing the superiority of merged sensor data in emotion

recognition using ECG signals. These findings have significant implications for the field of emotion recognition, as they provide a valuable foundation for future research in this area.





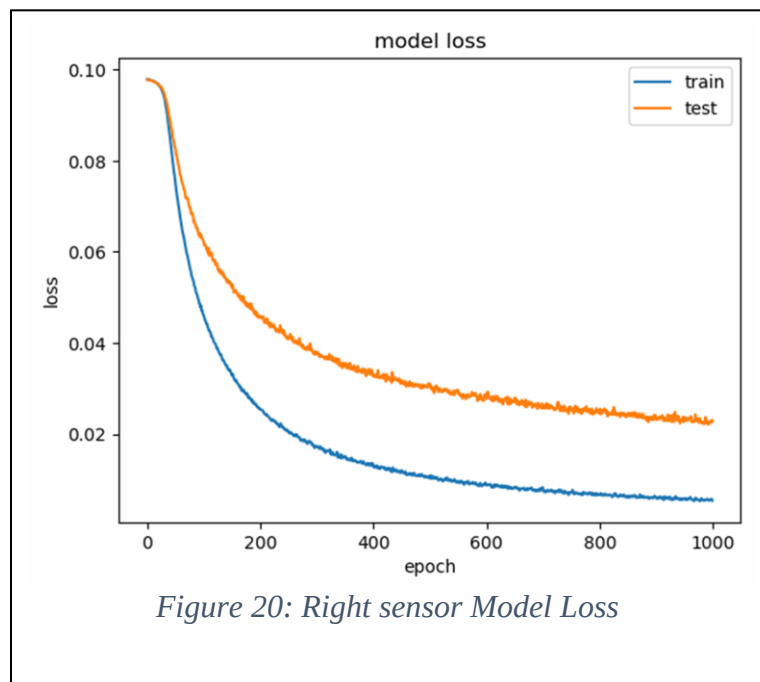
Figures 20, 21, and 22 provide a detailed overview of the loss curve for the model, which is a critical component of assessing the model's performance. Figure 20 illustrates that there is a significant gap between the test loss and the training loss, suggesting that overfitting is present in the model. Specifically, when the test loss curve is above the training loss curve, it indicates that the model is overfitting to the training data, resulting in poor performance with new or unseen data.

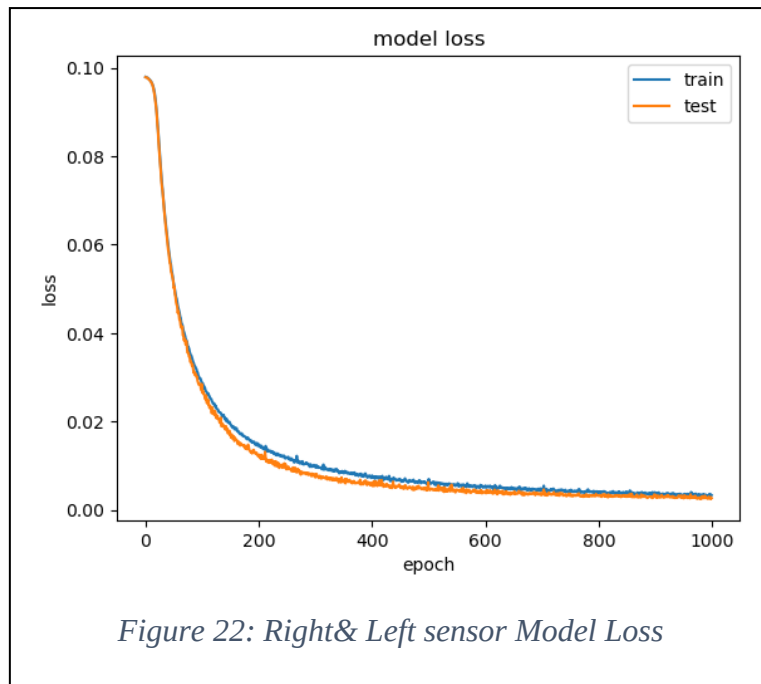
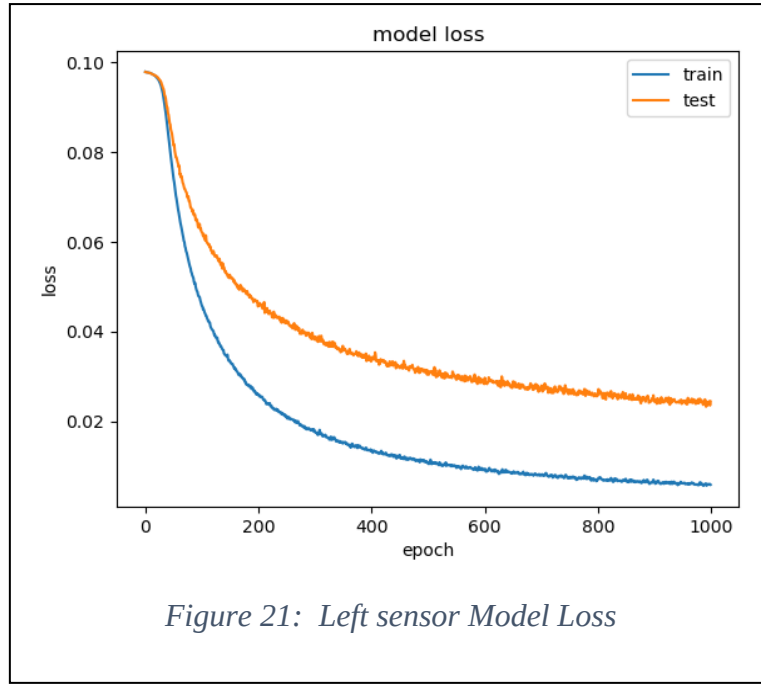
Moreover, the right sensor loss curve, as demonstrated in Figure 21, had a lower loss rate compared to the left sensor. However, it also exhibited overfitting, with a significant gap between the test loss and the training loss curves. This finding highlights the importance of evaluating each sensor configuration separately, as they may exhibit different levels of overfitting and require distinct pre-processing techniques.

Figure 22 presents the merged sensor loss curve, which showed an acceptable loss rate and a minimal gap between the test and training curves. This result indicates that the merged sensor configuration achieved a good fitting curve, with no

significant overfitting present. The merged sensor configuration's loss curve reaffirms the importance of utilizing combined sensor data in emotion recognition using ECG signals, as it resulted in a more robust and accurate model.

Overall, the loss curves provide valuable insights into the model's performance, highlighting the presence of overfitting and the importance of utilizing comprehensive pre-processing techniques to enhance the model's accuracy and precision. The merged sensor configuration's loss curve also emphasizes the added value of utilizing combined sensor data, resulting in a more robust and accurate model with minimal overfitting.

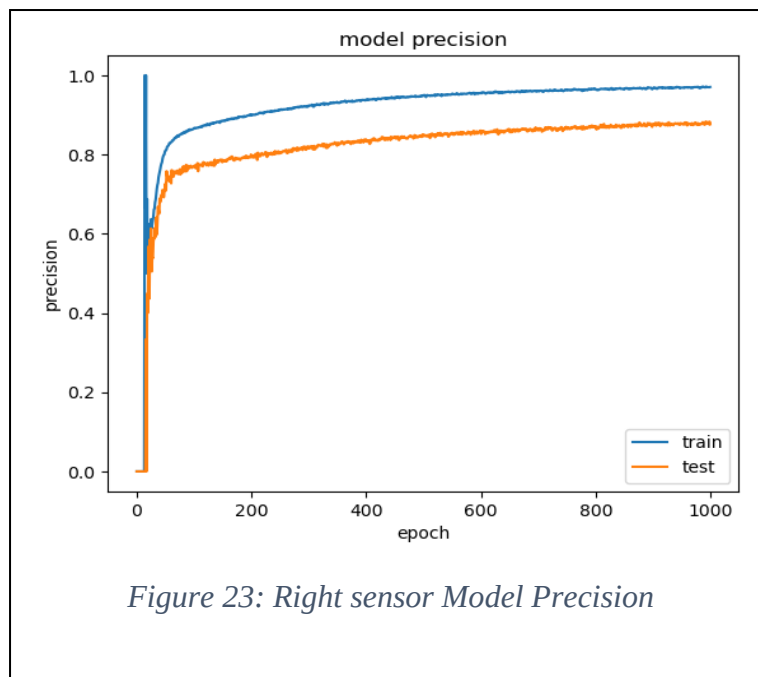


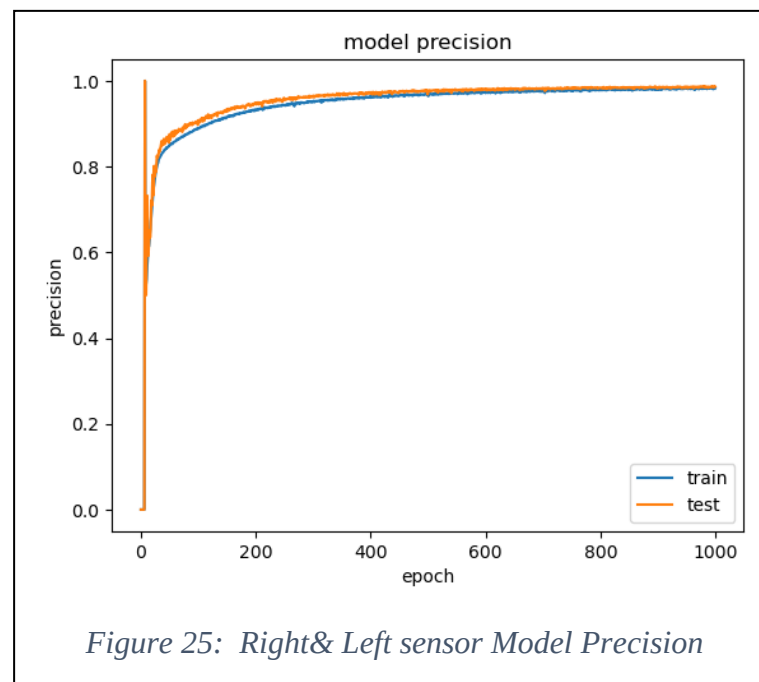
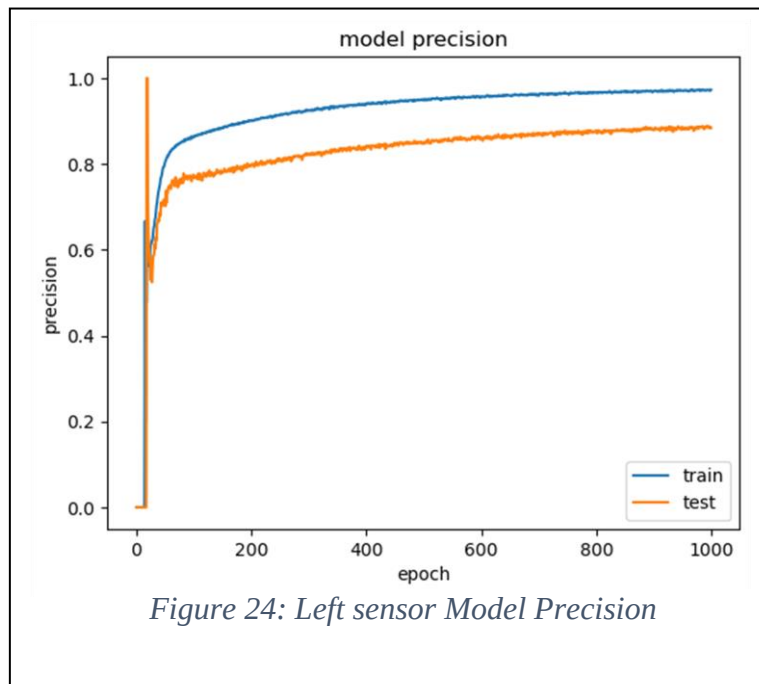


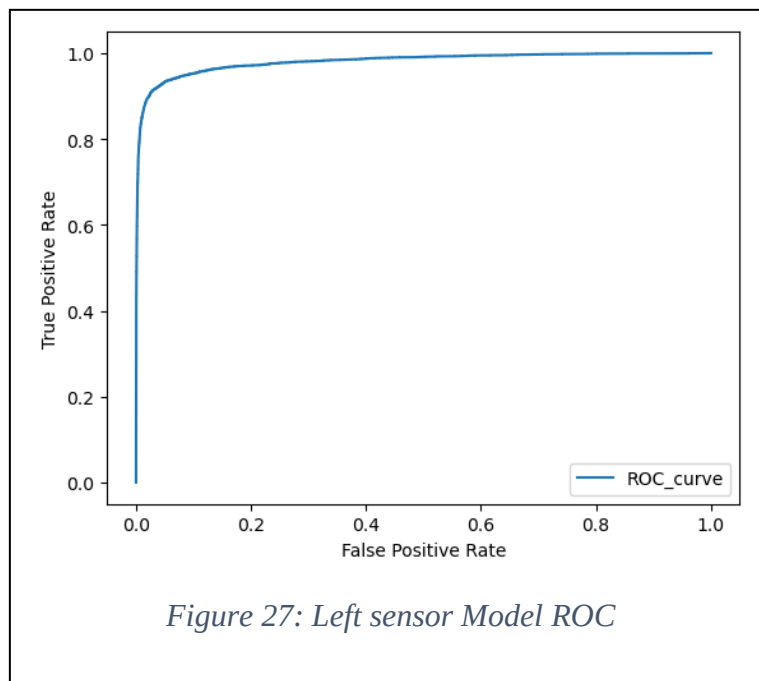
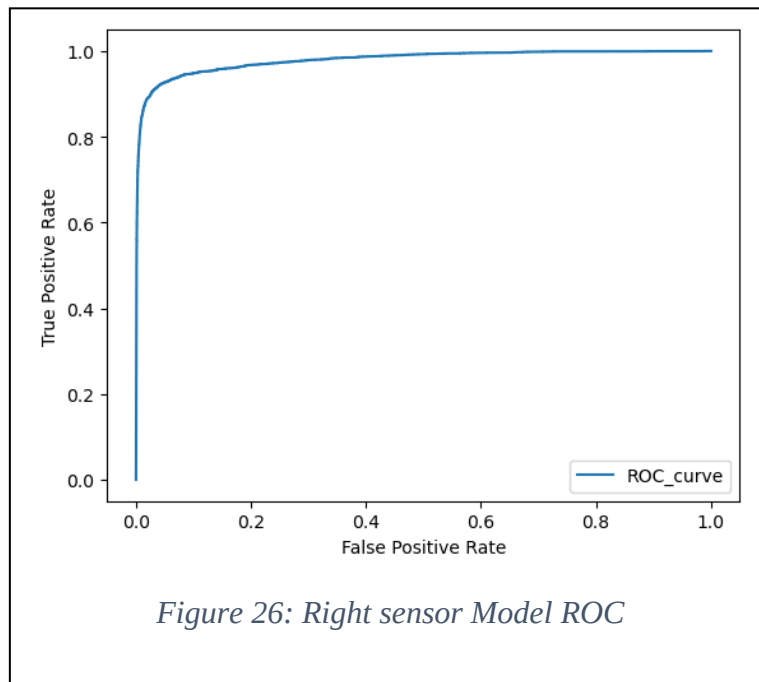
The precision curves and ROC curves are important measures of the performance of the proposed model. Figures 23, 24, and 25 illustrate the precision curves of the model, which show the relationship between the true positive rate and the false positive rate at various thresholds. The precision curves provide a visual representation of the model's ability to accurately classify emotional states. The

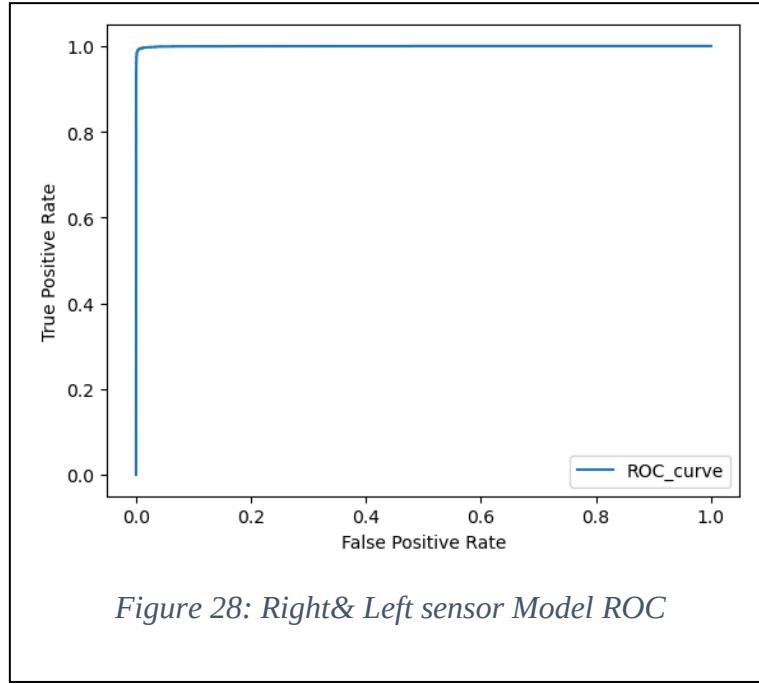
figures show that the model achieves a high level of precision across different thresholds, indicating that it is reliable and consistent in its classification.

On the other hand, figures 26, 27, and 28 demonstrate the ROC curves for the model. The ROC curves depict the trade-off between the true positive rate and the false positive rate at different classification thresholds. The area under the ROC curve (AUC) is a common metric used to evaluate the performance of a classification model. The AUC represents the model's ability to distinguish between positive and negative samples. The ROC curves for the proposed model show that it achieves an AUC of 0.99, which is an excellent performance measure. The high AUC value indicates that the model has high sensitivity and specificity in identifying emotional states.









4.4.2 Model Results vs. Other Studies

The proposed model presented in this study achieved a high accuracy rate of 98.6%, demonstrating its effectiveness in detecting emotions using ECG signals. However, this thesis is not the only study that has utilized the Dreamer dataset to detect emotions. For example, (Etemad & Sarkar, 2020) proposed a model that also used the Dreamer dataset. Their study utilized an IIR filter with z-normalization and proposed two different models, a fully supervised CNN and a self-supervised CNN, as shown in Table 4.

Similarly, (Hammad & Monkaresi, Detecting Naturalistic Emotions Using Deep Learning Techniques Applied to ECG Signals, 2022) proposed a model that utilized both private and Dreamer datasets. Their study filtered the signals, removed the baseline and powerline, and segmented the signals based on the Dreamer dataset. They used the Deep Fully Connected Neural Network with 10-fold cross-validation, as demonstrated in Table 6. Another paper by (Hammad & Monkaresi, ECG Based Emotion Detection via Parallel - Extraction of Temporal and Spatial Features Using Convolutional Neural Network, 2022) proposed a

model that also utilized both private and Dreamer datasets. They applied various pre-processing steps such as sliding window, low pass filter, baseline removal, and segmentation based on the Dreamer dataset. The model they proposed was called PETSFCNN-GSO-DNN, as shown in Table 6.

Furthermore, (Harper & Southern, 2020) proposed a model that used both Amigos and Dreamer datasets. Their study introduced different pre-processing methods, such as extracting IBIs from the ECG signals and then normalizing them using z-score. The authors proposed a model that combined LSTM and CNN and applied a Bayesian framework with a confidence threshold set to 90%, as shown in Table 7. Lastly, (Dar , Akram , Khawaja, & Pujari , 2020) suggested an LSTM CNN model that utilized pre-processed ECG signals. The authors used down-sampling, low-pass filter, baseline removal, and Z-score normalization, as shown in Table 7.

The aforementioned studies underscore the importance of utilizing various pre-processing techniques and ML models to achieve accurate and robust emotion recognition using ECG signals. The findings of these studies provide a valuable foundation for future research and development in this field.

Table 6: Other Work Model accuracies

Algorithm	Accuracy	
	Arousal	Valence
Fully supervised CNN (Etemad & Sarkar, 2020)	70.7%	66.6%
Self-supervised CNN (Etemad & Sarkar, 2020)	85.9%	85%
Deep Fully Connected Neural Network (10-fold) (Hammad & Monkaresi, Detecting Naturalistic Emotions Using Deep Learning Techniques Applied to ECG Signals, 2022)	97.5%	93.9%
PETSFCNN-GSO-DNN (Hammad & Monkaresi, ECG Based Emotion Detection via Parallel - Extraction of Temporal and Spatial Features Using Convolutional Neural Network, 2022)	96.34%	97.56 %

Table 7: Other Work Model accuracies

Algorithm	Accuracy
LSTM+CNN (Bayes $\alpha = 0.9$) (Harper & Southern, 2020)	86%
LSTM+CNN (Dar , Akram , Khawaja, & Pujari , 2020)	90.5%
Proposed model (BiLSTM CNN)	98.6%

4.5 Conclusion

In this chapter presented a DL model for recognizing emotions using ECG signals and specifically focused on the Dreamer dataset. The chapter, Discussed the importance of using appropriate datasets for training and testing DL models and explained how the Dreamer dataset was pre-processed and used in our model.

Comparing the proposed BiLSTM and CNN model with other existing models in terms of their procedures and results. The suggested model achieved high accuracy on the Dreamer dataset, outperforming several other state-of-the-art methods. The chapter attributed this success to the combination of BiLSTM and CNN layers, as well as the use of four different lowpass filters for data augmentation.

The presented results demonstrate the effectiveness of the proposed model for emotion recognition using ECG signals, and according to the results proposed work may contribute to the field of affective computing by demonstrating the potential of DL models in this area.

Chapter 5

Conclusion and future work

Chapter 5: Conclusion and future work

5.1 Conclusion

The primary objective of this research was to design and implement a DL architecture for the detection of emotions based on ECG signals, taking into consideration data division situations that were more representative of real life and evaluating signal and subject-dependence issues. In order to accomplish this goal, it was necessary to investigate the basics that lie underlying the ECG and emotions, in addition to gaining a knowledge of their normal functioning and behavior. In addition, after becoming familiar with the fundamental theoretical information pertaining to these aspects, several prior art approaches and techniques were investigated and evaluated. This was done in order to gain an understanding of the state-of-the-art that had already been developed, as well as the scientific evolution that had already taken place in this field, as opposed to what might still be lacking.

The approach discussed has effective and promising results. Using of ML and DL are helpful in building models for emotion detection. In addition, ML methodologies used ECG signals in emotion detection introduced in the literature. Moreover, DL approaches that use ECG signals in emotion recognition proposed in the literature also emotion recognition using other methods were explored. The fact that the results of the previous research using DL architectures were able to exceed the results of the studies using ML methodologies highlighted the promise of employing DNNs in this discipline.

However, several drawbacks were experienced, such find the proper pre-processing method. In addition, the strategies for dividing the data provide random splits for the train and the test. Moreover, the majority of DL algorithms employed ECG segments rather than the complete signals, which resulted in train

and test sets having segments that belonged to the same signal also, the signals were augmented and the segment was repeated using different pre-processing methods. Since it only evaluates sets of data that pertain to the same area and are derived from the same signals, this unrealistic method of assessing model performances may lead to certain outcomes that aren't entirely accurate. In this manner, it became important to duplicate several ways in order to locate a solid starting point for the work that was created in this dissertation. In addition, it became important to evaluate these promising methods using data settings that were more stringent and demanding.

As stated, many strategies were investigated and tested, taking into consideration a variety of DL architectures and methodologies. Unfortunately, due to a lack of vital information as well as a lack of clarity about all of the phases of the suggested methodology, the results of the replications showed poor emotion detection ability. Taking into consideration the shortcomings that were discussed and the limits that were shown by the replication stage. This thesis conducted a different classification method than all previous studies, but the pre-processing method was inspired by the work done by (Etemad & Sarkar, 2020) but some additions and changes were adopted.

The applied methodology supports different enhancements in the results. The conducted procedures succeeded in improving the accuracy of detecting emotions using the ECG signals. Furthermore, the proposed model has the highest accuracy among the explored studies that use the dreamer dataset. The use of alternative data divisions to this enhanced emotion recognition network resulted in much superior performances than when applying random data splits, which revealed both signal and subject dependency. This was achieved by applying varied data divisions.

Nevertheless, in order to further enhance the network that was constructed, Multiple Instance Learning approaches were taken into consideration. The

primary purpose of these techniques was to overcome certain restrictions that were caused by the sparsely labeled datasets that were available. BiLSTM structures and CNNs were applied to recognize the nine different emotions using the augmented segments that stand for the same signals. By using the suggested methodology, results were enhanced and improved.

Both data augmentation and DL have been shown to be effective in overcoming the challenges that arise from working with very limited data sets. On the other hand, the suggested solutions were effective enough to cut down on the signal and subject-dependence issues that had already been discovered. The performance of the model seems to be promising when all actual variables are taken into consideration, including signal- and subject-independent parameters, as well as cross-database trials.

These findings bring into perspective the overall techniques that claimed promising model performances with random data splits. As these dependencies hinder them from having a true application in the actual world, these dependencies preclude these approaches from having a proper application. In light of this, in addition to looking for a technique that shows promise for emotion identification, the primary emphasis of this thesis was on presenting a combination of the BiLSTM CNN model that enhances the ability to predict emotions, as well as the differences between this thesis and previous research.

Using data augmentation because the available datasets are considered small. Even with complicated architectures that would likely make it possible to acquire correct results, models that are not adequately trained as a consequence of not having sufficient subjects nor sufficient data per subject are unable to effectively distinguish emotions from data that has not yet been seen. This is the case, even though having sufficient subjects and sufficient data per subject would likely make it possible to obtain accurate results. It is common knowledge that methods of DL call for a substantial amount of data, but there are currently no databases

on the market that contain physiological signals for the purpose of emotion recognition. Such a database would be able to guarantee the quantity of data required to produce accurate models that have the potential to be used in practical settings.

5.2 Recommendations and Future Work

In this manner, in addition to reporting these systematic and recurrent problems, this dissertation improved an available literature approach by taking into consideration, as it was intended to do so, various DL architectures. Some examples of these architectures include BiLSTM structures and CNNs. But, in addition to these, alternative strategies, such as hand-crafted methodology or even multimodal approaches, might also be taken into consideration. Nevertheless, since there was only a limited amount of time to construct the current dissertation, there is a great deal of further work that may be done in the future to enhance and advance this topic on a scientific level.

With regard to the constraints that were experienced, it is necessary to build more extensive databases and, where it is feasible to do so, to use more stringently supervised techniques, provide a wide range of signals to each subject, and conduct many studies. Techniques that are robust enough to distinguish emotions without any signal, subject, or database reliance won't be viable to create until there is sufficient data and signal diversity. This won't be achievable until more data is collected and more signals are generated. In addition, in addition to the expansion of databases, various methods, such as unsupervised learning, one-shot learning, weakly-supervised learning, and self-supervised learning methodologies, may be considered and used to reduce some of the limitations caused by low data availability. This may be done in conjunction with the expansion of databases. In circumstances when there is a limited amount of data accessible, each of these methods is appropriate and often results in improvements.

In addition, additional potential developments in the future may concentrate on expanding the quantity of data that is already available since. Also, extending the model validation to other ECG datasets for emotion identification, such as amigos. Nevertheless, since the issue is still at the signal level, it is necessary to first eliminate this signal dependence and then concentrate on certain strategies to guarantee high performance with subject-independent settings. This is because the issue is still at the signal level.

Finally, it is possible to say that the objectives of the dissertation were met, and there is a possibility that it will be a scientific work that contains fascinating debates and proposals for future work related to the subject of emotion recognition and affective computing.

Bibliography

- A. Garc'ia, C., Otero, A., Vila, X., M'endez, A., Rodr'iguez-Li~nares, L., & Lado, M. (2016). RHRV Quick Start Tutorial.
- Ab. Aziz , N. A., Hasnul, M. A., Alelyani, S., Mohana , M., & Abd. Aziz , A. (2021). Electrocardiogram-Based Emotion Recognition Systems and Their Applications in Healthcare—A Review. *Sensor*.
- Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE*, 209–22.
- Agrafioti, F. (2011). *ECG in Biometric Recognition :time Dependency and Application Challenges*. University of Toronto.
- Andrew, G., Arora, R., Bilmes, J., & Livescu, K. (2013). Deep canonical correlation analysis. *International conference on machine learning*, 1247-1255.
- Ayata, D., Yaslan, Y., & Kamasak, M. (2017). Emotion Recognition via Galvanic Skin Response: Comparison of Machine Learning Algorithms and Feature Extraction Methods. *IU-Journal of Electrical & Electronics Engineering*, 3147-3156.
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2481 - 2495.
- Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 614–636.
- Bashyal, S., & Venayagamoorthy, G. (2008). Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence*, 1056-1064.
- BETTS, J., DESAIX, P., JOHNSON, E., JOHNSON, J., KOROL, O., KRUSE, D., . . . YOUNG, K. (2013). *Anatomy & Physiology*. OpenStax.
- Bhaskar, P., & Uplane, M. (2016). High frequency electromyogram noise removal from electrocardiogram using FIR low pass filter based on FPGA. *Procedia Technology*, 497-504.
- Butterworth, S. (1930). On the theory of filter amplifiers. *Wireless Engineer*, 536-541.
- Chung, W. J., & Cho, T. H. (2018, August). A Multi-Path Routing Determination Method for Improving the Energy Efficiency in Selective Forwarding Attack Detection Based MWSNs. *International Journal of Wireless & Mobile Networks (IJWMN)*, 10(4), 9-19.
- Cohn, J., & Torre, F. (2015). Automated face analysis for affective computing. *The Oxford handbook of affective computing*, 131–150.

- Cohn, J., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The handbook of emotion elicitation and assessment*, 203-221.
- Dahl, G., Sainath, T., & Hinton, G. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 8609-8613.
- Dar , M., Akram , M., Khawaja, S., & Pujari , A. (2020). CNN and LSTM-Based Emotion Charting Using Physiological Signals. *Sensors*.
- Dauphin, Y., Fan, A., Auli, M., & Grangier, D. (2017). Language Modeling with Gated Convolutional Networks. *International Conference on Machine Learning*, 933-941.
- Delashmit, W., & Manry, M. (2005). Recent Developments in Multilayer Perceptron Neural Networks.
- El Ayadi, M., Kamel, M., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern recognition*, 572-587.
- Etemad, A., & Sarkar, P. (2020). Self-Supervised ECG Representation Learning for Emotion Recognition. *IEEE*, 1541 - 1554.
- Frank, E. (1956). An Accurate, Clinically Practical System For Spatial Vectorcardiography. *Circulation*, 737-749.
- Gent, P., Farah, H., Nes, N., & Arem, B. (2019). HeartPy: A novel heart rate algorithm for the analysis of noisy signals. 368-378.
- Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 315-323.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Greene, S., Thapliyal, H., & Caban-Holt, A. (2016). A Survey of Affective Computing for Stress Detection: Evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine*, 44 - 56.
- Hammad, D., & Monkaresi, H. (2022). Detecting Naturalistic Emotions Using Deep Learning Techniques Applied to ECG Signals. *Webology*, 6146-6180.
- Hammad, D., & Monkaresi, H. (2022). ECG Based Emotion Detection via Parallel - Extraction of Temporal and Spatial Features Using Convolutional Neural Network. *Traitement du Signal*, 43-57.
- Harper, R., & Southern, J. (2020). A Bayesian Deep Learning Framework for End-To-End Prediction of Emotion From Heartbeat. *IEEE*, 985 - 991.
- J. Bota, P., Wang, C., Fred, A. L., & Da Silva, H. P. (2019). A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE*, 140990 - 141020.

- Karlik, B., & Olgac, A. (2011). Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. *International Journal of Artificial Intelligence and Expert Systems*, 111-122.
- Katsigiannis, S., & Ramzan, N. (2018). DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE*, 98–107.
- Kennedy, A., Finlay, D. D., Guldenring, D., Bond, R., Moran, K., & McLaughlin, J. (2016). The cardiac conduction system: generation and conduction of the cardiac impulse. *Critical Care Nursing Clinics*, 269-279.
- Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE*, 2067 - 2083.
- Klabunde, R. E. (2012). *Cardiovascular Physiology Concepts*. Lippincott Williams & Wilkins.
- Ko, B. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors*.
- Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M., & Kraaij, W. (2014). The SWELL Knowledge Work Dataset for Stress and User Modeling Research. *In Proceedings of the 16th international conference on multimodal interaction*, 291–298.
- Krizhevsky, A., Sutskever, I., & E. Hinton, G. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 84-90.
- Kuo, P.-H., & Huang, C.-J. (2018). A Green Energy Application in Energy Management Systems by an Artificial Intelligence-Based Solar Radiation Forecasting Model. *Energies*.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 436–444.
- Levenson, R. W. (2014). The Autonomic Nervous System and Emotion. *Emotion Review*, 100-112.
- Li, D., & Chen, C. (2022). Research on exercise fatigue estimation method of Pilates rehabilitation based on ECG and sEMG feature fusion. *BMC Medical Informatics and Decision Making*, 1-11.
- Li, J., Cheng, J.-h., Shi, J.-y., & Huang, F. (2012). Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement. *Advances in Intelligent and Soft Computing*. Springer.
- Li, S., & Deng, W. (2022). Deep Facial Expression Recognition: A Survey. *IEEE*, 1195-1215.
- Lin, M., Chen, Q., & Yan, S. (2013). Network In Network.
- Liu, W., Qiu, J.-L., Zheng, W.-L., & Lu, B.-L. (2019). Multimodal Emotion Recognition Using Deep Canonical Correlation Analysis.

- Luthra, A. (2019). Nomenclature of ECG Deflections. In *ECG Made Easy* (pp. 1-14). JAYPEE BROTHERS MEDICAL PUBLISHERS.
- Maas, A., Hannun, A., & Ng, A. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models.
- Marieb, E. N., & Hoehn, K. (2013). *Human Anatomy & Physiology*. Pearson.
- McCraty, R. (2022). Following the Rhythm of the Heart: HeartMath Institute's Path to HRV Biofeedback. *Applied Psychophysiology and Biofeedback*, 305–316.
- Mellouk, W., & Handouzi, W. (2020). Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science*, 689-694.
- Miranda-Correa, J., Abadi, M., Sebe, N., & Patras, I. (2018). AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Transactions on Affective Computing*, 479-493.
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- Nair, V., & Hinton, G. (2010). Rectified linear units improve restricted boltzmann machines. *In Proceedings of the 27th international conference on machine learning*, 807-814.
- Neal, R. (1992). Connectionist learning of belief networks. *Artificial intelligence*, 71-113.
- Nicholson, J., Takahashi, K., & Nakatsu, R. (2000). Emotion Recognition in Speech Using Neural Networks. *Neural Computing & Applications*, 290–296.
- Nita, S., Bitam, S., Heidet, M., & Mellouk, A. (2022). A new data augmentation convolutional neural network for human emotion recognition based on ECG signals. *Biomedical Signal Processing and Control*.
- Nwe, T., Foo, S., & De Silva, L. (2003). Speech emotion recognition using hidden Markov models. *Speech communication*, 603-623.
- Pinto, J. (2017). *Continuous biometric identification on the steering wheel*.
- Pranav, E., Kamal, S., Chandran, C., & Supriya, M. (2020). Facial Emotion Recognition Using Deep Convolutional Neural Network. *IEEE*, 317-320.
- Pyakillya, B., Kazachenko, N., & Mikhailovsky, N. (2017). Deep Learning for ECG Classification. *Journal of Physics*. IOP Publishing.
- Ramachandran, P., Zoph, B., & Le, Q. (2017). Searching for Activation Functions.
- S, J., Murugappan, M., Wan, K., & Yaacob, S. (2014). Electrocardiogram-based emotion recognition system using empirical mode decomposition and discrete Fourier transform. *Expert Systems*, 110-120.
- Sampson, M., & McGrath, A. (2015). Understanding the ECG. Part 1: Anatomy and physiology. *British Journal of Cardiac Nursing*, 548-554.
- Scanlon, V. C., & Sanders, T. (2019). *Essentials of Anatomy and Physiology*. FA Davis.

- Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Laerhoven, K. (2018). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 400-408.
- Schuller, B., Rigoll, G., & Lang, M. (2003). Hidden Markov model-based speech emotion recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing*.
- Sepúlveda, A., Castillo, F., Palma, C., & Rodríguez-Fernández, M. (2021). Emotion Recognition from ECG Signals Using Wavelet Scattering and Machine Learning. *Applied Sciences*.
- Sohn, J., Yang, S., Lee, J., Ku, Y., & Kim, H. (2020). Reconstruction of 12-Lead Electrocardiogram from a Three-Lead Patch-Type Device Using a LSTM Network. *Sensors*.
- Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2011). A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE*, 42-55.
- Sultan, H., Salem, N., & Al-Atabany, W. (2019). Multi-Classification of Brain Tumor Images Using Deep Neural Network. *IEEE Access*, 69215 - 69225.
- Sun, B., & Lin, Z. (2022). Emotion Recognition using Machine Learning and ECG signals. *arXiv preprint*.
- Tate, P. (2009). *Seeley's principles of anatomy and physiology*. McGraw-Hill.
- Thompson, M. (2014). Chapter 14: Analog Low-Pass Filters. In M. Thompson, *Intuitive Analog Circuit Design* (pp. 531-583). Newnes: Boston.
- Wang, C., Feng, T., Kim, J., Wang, G., & Zhang, W. (2009). Catching packet droppers and modifiers in wireless sensor networks. *2009 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks* (pp. 1-9). IEEE.
- Wang, H., Zhang, Y., Liang, J., & Liu, L. (2023). DAFA-BiLSTM: Deep Autoregression Feature Augmented Bidirectional LSTM network for time series prediction. *Neural Networks*, 240-256.
- Watson, D., Clark, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 1063-1070.
- Yoo, S., Lee, C., Par, Y., Kim, N., Lee, B., & Jeong, K. (2005). Neural Network Based Emotion Estimation Using Heart Rate Variability and Skin Resistance. *Springer*, 818-824.
- Zeiler, M. D. (2013). On rectified linear units for speech processing. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3517-3521.
- Zhang, H., Jolfaei, A., & Alazab, M. (2019). A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing. *IEEE*, 159081 - 159089.

Zhang, Z., & Wei, D. (2006). A New ECG Identification Method Using Bayes' Teorem.
IEEE, 1-4.

الملخص

لأكثر من قرن، كان الباحثون والعلماء في جميع أنحاء العالم متحمسين للإشارات الكهربائية منذ أن تم إنتاج أول تخطيط كهربائي للقلب (ECG) وكانت هذه الإشارات محور الجهود البحثية والتطويرية الواسعة، وقد أدت إلى تقدمات ملحوظة في مجال القلب. ومن المعروف على نطاق واسع أن حالاتنا العقلية لها تأثير كبير على نشاط جهازنا العصبي الذاتي (ANS)، والذي بدوره يؤثر على إشارات تخطيط القلب الكهربائي. وقد أدى ذلك إلى استكشاف طرق لتحديد الحالات العاطفية باستخدام خوارزميات التعلم العميق للوصول إلى درجة عالية من دقة التعرف على المشاعر. يتم استخدام مجموعة بيانات التعرف على المشاعر Dreamer لتدريب النموذج. بسبب قيود مجموعة البيانات، تم تطوير طريقة جديدة لمعالجة مجموعة البيانات باستخدام أربعة فلاتر منخفضة التردد متميزة لغرض زيادة البيانات لإشارة تخطيط القلب الكهربائي. تم استخدام هيكلية مختلطة من الذاكرة طويلة المدى ثنائية الاتجاه (BiLSTM) وشبكة عصبية تحويلية (CNN) للكشف عن تسعة مشاعر (المرح، الإثارة، السعادة، الهدوء، الغضب، الإشمئزاز، الخوف، والمفاجأة). وقد أدى ذلك إلى معدل دقة مذهل بنسبة 98.6%. ومقارنةً بالدراسات الأخرى، فإن هذا المستوى من الدقة هو أعلى درجة دقة قابلة للتحقيق. يوضح هذا البحث الإمكانيات الكبيرة لإشارات ECG في الكشف عن الحالات العاطفية ويسلط الضوء على أهمية تقنيات المعالجة الشاملة وزيادة البيانات لتعزيز دقة ودقة النماذج. توفر نتائج هذا البحث أساساً قيماً للبحث المستقبلي في هذا المجال، مع تطبيقات محتملة في مجالات مثل الصحة العقلية وعلم النفس والتفاعل الإنسان و الحاسوب.

جامعة البحرين
كلية تقنية المعلومات



التعرف على المشاعر من إشارات تخطيط القلب باستخدام تقنيات التعلم العميق

أطروحة مقدمة كجزء من متطلبات الحصول على درجة الماجستير
في لغة الآلة والحوسبة الذكية

إعداد
منة الله محمد محمد
الرقم الجامعي: ٢٠١٢٠٧٧٣

إشراف
أ. د. فيصل الخطيب
الأستاذ
جامعة البحرين

الإشراف المشترك
د. حصة الجنيد
أستاذ مشارك
جامعة البحرين

مملكة البحرين
أبريل ٢٠٢٣ م