Investigating the use of Multimodal Physiological Signals in developing an Emotion Recognition Model for Social Care use.

A Masters Dissertation
submitted in partial fulfilment
of the requirements for the award
MSc. in Data Science.

by Olumide Odetunde

[redacted]

[redacted]

SUBMITTED ON 9th SEPTEMBER 2022

Abstract

Emotion recognition can be incorporated into social care to improve the health outcome of residents in nursing facilities. The capacity of emotion recognition model to predict dimensions of emotion or discrete emotions can be leveraged to identify the stimuli of these emotion for residents in nursing facilities.While external physical signals like speech, gestures, facial expressions can be used in developing emotion recognition models,they are not reliable because these signals can be controlled by humans at will. Thus the use of physiological signals not under the voluntary control of humans can be used in developing robust emotion recognition system.This thesis begins by thoroughly reviewing the literature of emotion recognition model and systems developed using physiological signals.This thesis then creates a data collection protocol to collect physiological signals, emotions and emotion intensity from research participants. The signals and ground truths obtained are pre-processed with statistical analysis and hypotheses testing conducted. To develop the emotion recognition system, a subject dependent approach was taken using ensemble classifiers. These models were optimised using grid search and evaluated using defined metrics. The best performing model was identified and presented and the framework used in developing the model was also presented.Further works were also suggested within this thesis project.

Acknowledgements

[redacted]

Contents

Contents 4

List of Figures 7

List of Tables 9

1 Introduction 10

1.1 Problem Statement . . . 10
1.2 Aims and Objectives . . . 11
1.3 Solution . . . 11
1.4 Solution Framework . . . 12
1.4.1 Data Gathering . . . 12
1.4.2 Data Annotation Form . . . 13
1.4.3 Data-set Creation, Feature Extraction and Selection . . . 13
1.4.4 Emotion Recognition Model Development . . . 14
1.5 Study Contribution . . . 14
1.6 Thesis Outline . . . 14

2 Literature Review 16

2.1 Introduction . . . 16
2.2 Emotion . . . 16
2.3 Emotion Signals . . . 17
2.3.1 Electroencephalogram (EEG) . . . 18
2.3.2 Electrocardiogram (ECG) . . . 19
2.3.3 Galvanic Skin Response (GSR) . . . 19
2.3.4 Respiratory Signal (RSP) . . . 20
2.3.5 Electromyography (EMG) . . . 20
2.3.6 Skin Temperature (SKT) . . . 20
2.3.7 Heart Rate Variability (HRV) . . . 21
2.4 Measuring Physiological Signals . . . 22
2.5 Emotion Recognition Framework . . . 23
2.5.1 Eliciting Emotions . . . 23
2.5.2 Signal Data Labelling . . . 24
2.5.3 Traditional Machine Learning Methods . . . 25

2.5.4 Model Evaluation and Performance 31
2.6 Summary 32

3 Data Collection and Analysis 33
3.1 Introduction 33
3.2 Data Collection 33
3.2.1 Emotion Induction 34
3.2.2 Smart Device 35
3.2.3 Data Annotation Form 37
3.2.4 Research Participants 38
3.2.5 Data Collection Protocol 38
3.2.6 Data Export 39
3.3 Data Pre-processing 39
3.4 Dataset Creation 44
3.5 Data Analysis 47
3.5.1 Normality Hypothesis Test 48
3.5.2 Significant differences between data observations for emotion groups 50
3.5.3 Significant differences between physiological data observations for sound of audio stimuli groups 51
3.6 Further Findings 52

4 Methodology 54
4.1 Introduction 54
4.2 Emotion Induction 54
4.3 Data Gathering 55
4.4 Data Pre-processing 55
4.5 Traditional Machine Learning Method 57
4.5.1 Feature Extraction 58
4.5.2 Feature Selection 58
4.5.3 Feature Fusion 60
4.5.4 Model Training 60
4.6 Model Evaluation 64
4.7 Summary 65

5 Results and Discussion **66**
5.1 Results 66
5.1.1 Evaluating Feature Selection Subsets 66
5.1.2 Model Evaluation Comparison 68
5.1.3 Model Evaluation Comparison using early level fusion . 72
5.2 Discussion 76
5.2.1 Emotion Recognition Model Framework 76
5.2.2 Explaining Predictions of Emotion Recognition Model . 76
5.2.3 Thesis Findings 79
5.2.4 Application in social care 81

6 Conclusion **82**
6.1 Further Works 83

References **84**

List of Figures

1 Emotion Wheel.(Plutchik, 2001) 17

2 Data collection and analysis pipeline. 34

3 Wrist-worn FitBit Sense Smart Device.(Fit, 2022) 37

4 Data Annotation Form 38

5 Fitbit Dashboard for Data Export 40

6 Fitbit data structure obtained after data export.(Fit, 2022) . . 40

7 heart_rate JSON file before pre-processing visualised in Microsoft Excel. 42

8 heart_rate file after pre-processing visualised in Microsoft Excel. 42

9 skin_temp CSV file before pre-processing visualised in Microsoft Excel. 43

10 skin_temp CSV file after pre-processing visualised in Microsoft Excel. 43

11 skin_temp CSV file before pre-processing visualised in Microsoft Excel. 43

12 df_label visualised in Microsoft Excel. 44

13 Final dataset *(df_dataset)* visualised in Microsoft Excel 46

14 Method Framework for developing emotion recognition model 56

15 Strong sound stimuli 56

16 Illustration of the thumb and index finger during signal collection with Fitbit Sense. (Fit, 2022) 57

17 Subset of Features Selected using Recursive Feature Selection. 60

18 Experiment setup for development of emotion recognition model. 63

19 Model Evaluation on train-validation set using 10 features . . 69

20 Model Evaluation on test set using 10 features 69

21 Confusion Matrix of Model developed with Xgboost classifier using test set 70

22 Confusion Matrix of Model developed with Random Forest classifier using test set 70

23 Confusion Matrix of Model developed with AdaBoost classifier using test set 71

24 Comparison of emotion recognition rate on train-validation set and test set 71

25 Model evaluation after executing early level fusion on training-validation set 73
26 Model evaluation after executing early level fusion on training-validation set 74
27 Confusion matrix of model developed with Xgboost classifier after early level fusion 74
28 Confusion matrix of model developed with random classifier after early level fusion 75
29 Confusion Matrix of Model developed with AdaBoost classifier after early level fusion 75
30 Emotion Recognition Framework 77
31 SHAP summary plot using the 10 selected Features 78

List of Tables

1 Related works within Emotion Recognition Field 30

2 Breakdown of the IADS-E Audio Stimuli.(Yang et al., 2018) . 35

3 Modules used for the purpose of this study 41

4 Explanation of variables in *df_final* 46

5 Descriptive Statistics for all three participants 47

6 Normality Test Results of skin_temp across emotion groups. . 49

7 Normality Test Results of skin_temp across audio stimuli groups. 49

8 Normality Test Results of heart_rate across emotion groups. . 49

9 Normality Test Results of heart_rate across audio stimuli groups. 50

10 Feature list and domain of features extracted with the TSFEL module 59

11 Emotion recognition rate using n features on training-validation set 67

12 Emotion recognition rate using n features on test set 67

13 Emotion recognition rate on training-validation set and test set after executing early level fusion 73

14 Values of hyperparameter used in developing emotion recognition model 78

15 Features used in developing the emotion recognition model . . 79

Chapter One

1 Introduction

The health of residents within nursing facilities can be improved if more attention is paid to the emotions of residents (Holst and Skär, 2017; K et al., 2021). The ability to identify the emotional states of residents helps to understand the stimulus behind experienced emotions. This understanding can be used to improve the overall health of residents by avoiding these stimuli. Further, the safety of residents and the health staff looking after residents can be further assured with an understanding of the emotional state of residents and the corresponding stimuli.

In recent times, numerous works have been done to predict emotions using different approaches. The use of external human physical signals like facial expressions, gestures, speech has been explored (Lin et al., 2013; Zhang et al., 2016; Tarnowski et al., 2017). The major challenge encountered is that external physical signals are under the voluntary control of humans. Thus, humans can voluntarily conceal and control these signals making the reliability of these signals questionable (Shu et al., 2018). Therefore this makes it difficult to develop reliable emotion recognition systems with these signals.Physiological signals which are not under the control of humans are reliable in developing emotion recognition algorithms. Recent studies have utilized this extensively (Shu et al., 2018; Ayata et al., 2020; Cosoli et al., 2021). Measuring these physiological signals however could be invasive, especially if done with medical equipment (Cosoli et al., 2021). This makes the portability and ease of use inadequate to scale within nursing facilities.

1.1 Problem Statement

To improve the overall health of residents within nursing facilities, emotion-aware systems and algorithms must be developed using a different approach. Multi-modal physiological signals which are reliable must be obtained with minimal and non-invasive methods and used in developing emotion recognition models that can be implemented within nursing facilities.

1.2 Aims and Objectives

This thesis project aimed to investigate the use multimodal physiological signals obtained from a wrist-worn smart device in developing an emotion recognition model.This was a pilot study.To achieve the aim of this study, the following steps had to be completed:

- A thorough and critical literature review on machine learning methods and frameworks used in creating emotion recognition model
- Selection and acquisition of a wrist-worn smart device embedded with sensor technology and the development of a data annotation form
- Collection of physiological signals and ground truth under a controlled environment using the wrist-worn device and data annotation form
- Creation of dataset by aligning the physiological signals and ground truth for each participant
- Data pre-processing, feature extraction and feature selection to improve model performance
- Development, validation, and testing of the emotion recognition model
- Presentation of a framework that can be used to develop an emotion recognition model using multimodal physiological signals obtained from a wrist-worn smart device.

1.3 Solution

The solution developed in this study classified emotion experienced by using multimodal physiological signals obtained from a wrist-worn smart device.This solution was developed as it was ideal for problems discussed in section 1.1. The use of a wrist-worn smart device to capture physiological signals used to develop an emotion recognition algorithm is particularly ideal and scalable within nursing facilities. The advantage lies in the ease of use, non-invasiveness, and portability, thus making it a better and more viable alternative to the use of medical equipment within a nursing facility setting.

Another advantage of using wrist-worn smart devices embedded with sensor technology is the ability of these devices to collect multiple or multi-modal physiological signals simultaneously(Cosoli et al., 2021). The use of a multi-modal approach in collecting physiological signals using just smart devices is uniquely advantageous because it contributes to the production of a robust emotion recognition model (Cosoli et al., 2021).

This solution validates the hypothesis surrounding the viability of smart devices in building emotion-aware systems. Thus, this makes it possible to classify periods in which different emotions are experienced by resident and understand the cause of such emotions especially when negative. Further, the emotion recognition model developed makes it possible for professional staff to use findings about the causes of extreme emotions in creating better care management plans to improve the health outcomes of residents.

1.4 Solution Framework

To build the emotion recognition model using physiological signals, this study was positioned and executed as a supervised learning task that required the physiological signals obtained from the smart devices be correctly aligned with the self-reported emotion and emotion intensity experienced. The resulting dataset was pre-processed and used to build, train, validate and test the model.Sections 1.4.1 - 1.4.4 explain all the major steps of this solution in greater detail.

1.4.1 Data Gathering

This study differs from other studies in that it does not use unreliable signals like external physical signals like gestures, and facial expression signals in creating and gathering the required dataset for the study (Cosoli et al., 2021). The study focuses on obtaining multi-modal physiological signals not with medical equipment that are invasive, non-portable and expensive, rather with a wrist-worn smart device. The reason for this approach was scalability within nursing facilities to help improve outcomes for both residents and health staff in nursing facilities. The methodology employed within this study was to gather physiological signals and appropriately align them with self-reported emotions and emotion intensity. The resulting dataset was used to develop the emotion recognition model which can be implemented in nursing facilities.

Participants were invited to wear the wrist-worn smart device and listen to different audio stimuli that were grouped to simulate daily sounds encountered in day-to-day living. Participants were then asked to report emotions experienced and rate the intensity of the emotions using a data annotation form developed for the purpose of this study.

1.4.2 Data Annotation Form

In collecting self-reported emotional states that would be aligned with physiological signals within this study, a data annotation form was developed from which participants could click to listen to the day-to-day audio stimuli and appropriately select the emotions experienced and rate the intensity of the emotions using three button clicks. Thus, using this data annotation form, participants could click to listen to the sound, select the emotions experienced and rate the emotion experienced on a Likert scale of 1 to 9.

The data annotation form collected the identification number of each sound, the time stamps when participants start to listen to each sound, the emotion experienced, emotion intensity ratings, and the time stamp when each report is submitted per sound listened to. All of these were captured by the data annotation form and saved as a comma separated value (CSV) file. This served as the ground truth for this study.

1.4.3 Data-set Creation, Feature Extraction and Selection

With the collection of ground truth using the data annotation form and multi-modal physiological signals using the wrist-worn smart device, both data must be aligned to create the data-set with which to train and develop an emotion recognition system using supervised machine learning methods. Details of the dataset creation, pre-processing methods, and hypotheses tested are discussed in chapter 3.

1.4.4 Emotion Recognition Model Development

A personalised or subject-dependent approach was used to develop an emotion recognition model after data analysis and findings obtained from chapter 3.

1.5 Study Contribution

There are three major contributions presented in this thesis:

- Data annotation form that can be adapted to collect data ground truth required to build emotion recognition models using multi-modal physiological signals.
- Presentation of an emotion recognition framework with which to create emotion recognition models using physiological signals from wrist-worn smart devices.
- Development of an emotion recognition model using an ensemble classifier to predict intensity of emotion stimuli with an accuracy of 94.44% for intensity of emotion on test set.

1.6 Thesis Outline

This section provides an overview of each chapter within this thesis project.In chapter 2, the literature review which served as the basis of this research was discussed. The review first explained emotion and how emotion is measured.The review then described different approaches taken by researchers to build emotion recognition algorithms. This was followed by a detailed explanation of physiological signals, classifiers that have been used to create emotion recognition systems in previous works.Chapter 3 describes the data preparation and data analysis techniques used within the study.Due to the use of sensor technology to obtain physiological signals from participants,data preparation techniques were employed in obtaining sequenced physiological signals.Further,more detail about the data annotation form was presented and how the physiological signals obtained were aligned which the ground truth to create the research data set. Statistical analysis and hypothesis testing were also explored in this chapter.Chapter 4 presents a detailed explanation of the methodology and machine learning techniques employed in developing the emotion recognition model. Chapter 5 presents the results of models developed after training validation and testing, the chapter also discusses the results obtained.Chapter 6 summarises this thesis.The chapter ascertains that the aims and objectives of the study have been met. Limitations and future research areas for the study were also discussed.

In summary, this project aims to investigate the use of mulitmodal physiological signals obtained from a wrist-worn smart device in developing an emotion recognition model scalable within nursing facilities. The next chapter present a thorough review of related works which served as the basis of this thesis project.

Chapter Two

2 Literature Review

2.1 Introduction

This chapter presents the related works done within the emotion recognition field.It explains emotions and how emotion is identified and measured. This chapter also presents signals used in capturing emotions as done in related studies. This chapter concludes by conducting an in-depth survey into the different emotion recognition frameworks used in related research.

2.2 Emotion

Emotion, as explained by Plutchik (2001), is a chain of events. It starts with a stimulus and involves feelings and physiological changes that lead to goal-oriented behaviour which aims to restore the body to equilibrium.Drawing from works within the emotion recognition field, researchers identify emotions discretely or with dimensions. These are the two ways psychologists tend to model emotions (Shu et al., 2018). Using discrete emotion categorization, Ekman (1992) explained nine characteristics of emotions and summarized six basic emotions (happy, sad, anger, hear, surprise, and disgust) using these characteristics. Ekman (1992) also considered other emotions to be a combination of these basic emotions. Similarly, Plutchik (2001) proposed an emotion wheel model presented in figure 1 which included eight basic emotions (joy, trust, fear, surprise, sadness, disgust, anger, and anticipation) arranged as four opposite pairs.

With more research recognition, degrees of correlation between discrete emotions were discovered. Also, discrete emotions could have varying intensities, for example, anger can be described as upset or aggressive. Further, with the rising number of categorized discrete emotions, evaluating, and labelling these emotions can be difficult. (Shu et al., 2018; Dzedzickis et al., 2020). Thus, researchers use emotional dimensions; arousal, valence, and dominance to identify and classify emotions (Feidakis et al., 2011). By way

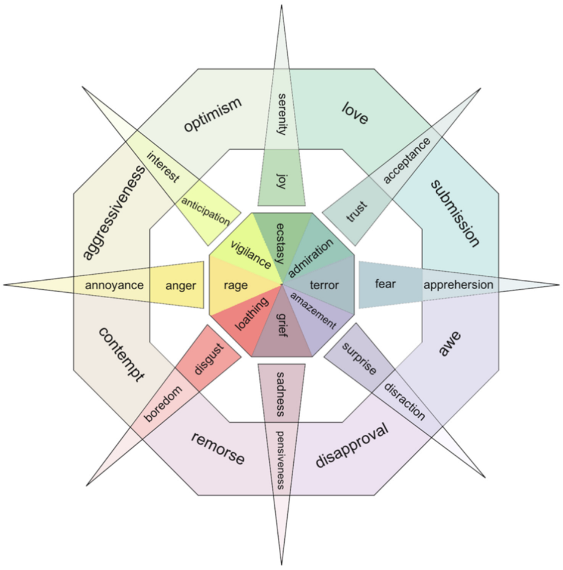


Figure 1: Emotion Wheel.(Plutchik, 2001)

of illustration, Lang (1995) explained that emotion can be categorized in 2D by valence (degree of the pleasantness of emotion) and arousal (a measure of the intensity of the emotion). Notably, Mehrabian (1997) expanded the 2D model to 3D by adding dominance (a measure that reflects the control ability of humans when experiencing discrete emotions).

In summary, discrete emotions and emotional dimensions are the two ways that have been used majorly in describing and identifying emotions.

2.3 Emotion Signals

Several works have used different physiological signals to create emotion recognition models. These signals could be external human physical signals like speech, facial, gesture, voice intonation, postures, or physiological signals (Shu et al., 2018).The use of external human physical signals in developing emotion recognition models has been explored in different studies. For example, Lin et al. (2013) used students' facial expressions to identify emotions and determine if students understood the taught content or not in real-time. Li et al. (2021) detailed the application of speech emotion recognition in identifying emotions between physicians using artificial neural networks.

While these external physical signals and labels are easy to collect, the signals are unreliable for building emotion recognition models (Cosoli et al., 2021; Shu et al., 2018). Humans can intentionally mask behavioural signs by controlling facial muscles or changing speech patterns and gestures to hide their real emotions (Shu et al., 2018). This impacts the reliability of emotion recognition models developed using external human physical signals. Furthermore, some of these signals (facial expressions, gestures) are collected using cameras installed under certain lighting conditions and specific resolutions (K et al., 2021). This may not be applicable in some research fields due to privacy concerns, hence a need for another approach.

Another approach to building emotion recognition models is the use of physiological signals. The human nervous system is divided into two: the central nervous system (CNS) and the peripheral nervous system (PNS). The PNS can be split into autonomic and somatic nervous systems. The autonomic nervous system comprises sensory and motor neurons which are relay between the CNS and some organs of the body like the heart, eyes, lungs, etc. Depending on the stimuli humans receive at every given time, the physiological signals from these visceral organs change and the signals aren't under the voluntary control of humans, making these signals reliable for building robust emotion recognition models (Shu et al., 2018).Thus, the use of physiological signals in developing an emotion recognition model is reliable and has been employed within this study. Sections 2.3.1 - 2.3.7 explains different physiological signals in developing emotion recognition models.

2.3.1 Electroencephalogram (EEG)

An electroencephalogram measures the electric potential of the brain in a time series. This implies that EEG is a direct measure of the change in brain activities thus making it a vital signal in detecting and recognizing emotions (Lan et al., 2016). A study by Lin et al. (2010) proposed an EEG – based emotion algorithm during music listening that achieved an accuracy of 82.29% with 32 channels across 26 subjects. Four emotions were determined using differential symmetry of hemispheric power spectra as features and a support vector machine as the classifier.

2.3.2 Electrocardiogram (ECG)

An electrocardiogram is a method for interpreting the electrical activity of the heart. The heart is directly related to the human central system making ECG useful for analyzing the activity of the heart and in emotion recognition (Goshvarpour et al., 2017). The QRS complex of the ECG is mostly used in emotion recognition because it defines the activation of the heart in relation to emotional state, hence making it a viable indicator of discrete emotions (Dzedzickis et al., 2020). A study by Cheng et al. (2017) presented a novel method for negative emotion recognition using a single channel ECG extracted seven linear-derived features, ten non-linear features, six time-frequency domain features and four-time domain features.

2.3.3 Galvanic Skin Response (GSR)

GSR also electrodermal activity or skin conductance is a continuous measurement of the electrical attributes of the human skin. Changes in emotion induce sweat reaction, sweat reaction causes variations in the concentration of salt in human skin leading to changes in the electrical resistance of the skin (Dzedzickis et al., 2020). The electrical parameters of the skin are not under the voluntary control of humans because changes in electric parameters of the skin reflect changes in the sympathetic nervous system. A study by Wu et al. (2010) recognized six emotions from GSR signals using the Fischer classifier. Statistical features were extracted using functions like range, maximum, and minimum of the GSR signals. A total of 30 features were extracted. Features were reduced using immune hybrid particle and swarm optimization. The model built achieved an average accuracy of 78.72%, 73.37%, 70.48%,62.65%, 62.52% and 44.93% for fear, disgust, grief, happiness, and anger respectively.

2.3.4 Respiratory Signal (RSP)

Respiratory signals give vital information about emotional states. Fast and in-depth breathing depicts excitement which could be associated with discrete emotions like happiness, anger, or afraid. Fast and shallow breathing often depicts tension. Hence depth and velocity of respiration vary with different states of human emotions. (Dzedzickis et al., 2020). In a study by Wong and Tan (2010), breathing rate, amplitude and statistical features like mean and standard deviation were extracted from RSP and used with three other signals to classify four emotions. An accuracy of 86% was achieved for four emotions using the four signals.

2.3.5 Electromyography (EMG)

Electromyography is a method used to evaluate and record the electric potential generated by muscle cells (Zong and Chetouani, 2009). Most of the research in emotion recognition using EMG focuses on the analysis of facial expression because of the hypothesis that facial mimicry contributes to the emotional response to different stimuli (Dzedzickis et al., 2020). EMG signals are viable input signals in emotion recognition. For example, Gouizi et al. (2011), utilized EMG, RSP, skin temperature (SKT), heart rate (HR), skin conductance (SKC) and blood volume pulse (BVP) as input signals to classify six different emotions. The features extracted from the EMG included temporal and frequency parameters. Temporal parameters were mean, standard deviation, and mean of the absolute values of the first and the second difference. The model was developed using a support vector machine classifier, and a recognition rate of 85% for the different emotional states was achieved.

2.3.6 Skin Temperature (SKT)

Skin temperature is a bio-signal that represents a reaction of the autonomic nervous system which is beyond the voluntary control of humans. Skin temperature is directly related to human heart activities and sweat reactions, thus making it a signal used in automatic emotion recognition (Dzedzickis et al., 2020). An advantage of skin temperature signal is the possibility of non-contact measurement which ensures high comfort to participants. However, SKT is useful when a stimulus last for a long duration and SKT is often used in combination with other signals because SKT alone often cannot detect an exact emotion (Dzedzickis et al., 2020). For example, research by Kim et al. (2004) used ECG, SKT and GSR all obtained with minimal discomfort and reflective of the autonomic nervous system to classify discrete emotions. Correct classification ratios of 78.4% and 61.8% were obtained for three (sadness, anger, stress) and four (sadness, anger, stress, surprise) discrete emotion categories respectively using a support vector machine classifier.

2.3.7 Heart Rate Variability (HRV)

Heart rate variability is a method based on the measurement of beat-to-beat variation on time with a certain period of sinus rhythm (Dzedzickis et al., 2020). In simpler terms, HRV examines the time variance and regularity of each heartbeat cycle (Hsieh and Chin, 2011). Variability in heart rate depicts the effect of the two branches (sympathetic and parasympathetic) of the autonomic nervous system. While the parasympathetic branch slows the heart rate, the sympathetic increases the heart rate. The changes in the effect of the parasympathetic and sympathetic branches with respect to HRV can be influenced by emotions (Benezeth et al., 2018).HRV can be measured and determined classically with ECG. The drawback of HRV based on ECG is the complexity of sensors and complex requirements for the measurement procedure (Dzedzickis et al., 2020).

An alternative to HRV based on ECG is PPG. Photoplethysmography is a method to detect the change in micro-vascular blood volumes in tissues. In recent times, there have been increased works in remote PPG which makes it possible to recover cardiovascular pulse waves by measuring variations of back-scattered light remotely using only ambient light (Benezeth et al., 2018). PPG signals have been used in emotion recognition. A recent paper by Lee et al. (2019) investigated the use of PPG to classify emotions by adopting a two-dimensional emotion model (valence and arousal). PPG signal features were extracted using a one-dimensional convolutional neural network. An emotion recognition accuracy of 75.3% and 76.2% for valence and arousal respectively was achieved using the DEAP dataset.

2.4 Measuring Physiological Signals

Most of the physiological signals discussed in section 2.3 can be measured using medical equipment. Medical equipment can be expensive and invasive. Thus, collecting physiological signals with medical equipment is not easily accessible, requires training in some cases and there is no ease of collection using this method (Cosoli et al., 2021). With progressive development in technology, it is possible to measure and collect these physiological signals using smart wearable devices (Cosoli et al., 2020). These wearable devices use embedded sensor technology to measure and capture needed physiological signals (Shany et al., 2012). Depending on physiological signals required for research and tele-health applications, wearable devices with the required accuracy in measuring these signals are selected and used (Cosoli et al., 2020).Advantageously, these smart wearable devices allow continuous data collection over time under different environments useful for monitoring (Cosoli et al., 2021). Furthermore, these devices can measure different physiological signals simultaneously, thus giving the unique advantage of obtaining multi-modal physiological signals with the same device. There is however no generalized standard protocol for the validation of wearable devices. Massive data generated from these wearable devices require immense computational resources for preparation and analysis. Furthermore, different wearable devices have varying accuracy in measuring various physiological signals. (Cosoli et al., 2020).

These smart devices could be worn on different parts of the body like the wrist or chest. The device used in this study was worn on the wrist for comfortability, and the wrist-worn device captured physiological signals used in building the emotion recognition model. With an understanding of emotion, the reason for using physiological signals and the choice of smart device, the next section discusses the end-to-end framework used by researchers in developing emotion recognition models.

2.5 Emotion Recognition Framework

Building emotion recognition models require a sequenced process that begins with inducing emotions and ends with the validation and testing of the developed emotion recognition model. This section discusses the entire framework using related works that have developed emotion recognition systems from multimodal physiological signals.

2.5.1 Eliciting Emotions

Emotions occur due to stimulus; this emotion can be captured naturally or induced and captured under a controlled environment. Thus, in developing emotion recognition models, researchers have used various stimuli (visual, audio etc.) to elicit emotion in participants within experimental setups. By way of illustration, the national institute of mental health (1997) proposed the International Affective Picture System (IAPS) contains a series of standardized emotion-eliciting pictures (Bradley and Lang, 2017).A combination of visual and auditory stimuli (movies, music videos) has been used progressively in eliciting emotions. For example, Nasoz et al. (2004) initially built a multi-modal affective user interface to gather emotions and related data. After a pilot study to determine top-quality movies, Nasoz et al. (2004) selected twenty movie clips to elicit anger, sadness, amusement, surprise, fear, and disgust. Similarly, a study by Koelstra et al. (2012) had 32 subjects watch 40 one-minute-long music videos that were selected from an initial 120 music videos using affective tags and affective highlighting algorithms. This was done to elicit emotion that was self-reported by participants. Recently, Zhang et al. (2017) proposed the affective virtual reality system (AVRS) which was made of eight emotive virtual reality scenes. A three-dimensional emotion index of AVRS was evaluated by one hundred subjects using the self-assessment manikin. This study concluded that emotion was better elicited by the AVRS.

From the literature, emotions have been elicited by researchers using different visual, audio stimuli or sometimes a combination of different stimuli like film, clips, music, sounds etc. to induce emotions which are then captured and used to develop emotion recognition models. Obtaining ground-truth under the experimental setup is discussed in section 2.5.2.

2.5.2 Signal Data Labelling

Labels or ground truth in the context of this study means the emotion experienced either discretely and intensity of the emotion. Labelling physiological signals with emotions experienced under a controlled environment at different moments, especially by filling a self-report requires a lot of effort to obtain reliable labels (Kołakowska et al., 2020). Under experimental setups, a self-report questionnaire could be presented to participants at specified points of the experiment session. It could be at the beginning, end or after each step of the session if the session contains many steps (Mottelson and Hornbæk, 2016; Ciman and Wac, 2018; Exposito et al., 2018). The self-reports are filled accordingly. In some other cases, labels are assigned according to emotions elicited with an assumption that the emotion elicitation was successful (Tikadar et al., 2017; Tikadar and Bhattacharya, 2019).

There are several ways self-report forms can be structured for ease of use. The simplest way is to provide option buttons for each emotional state elicited, this supports ease of usage and enables the binary representation of selected emotion. Within the self-report forms, emotion can be reported using a Likert scale and may be applied to discrete emotional states. Self-assessment manikin: a pictorial version of the Likert scale can be applied to emotional dimensions (Valence, Arousal, and Dominance). There are other scales designed specifically for emotional states using a self-report form. Positive and negative affective schedule (PANAS) is one of those scales. PANAS is a form that contains two 10-item mood scales designed to measure positive and negative affect.Drawing from literature, this study employed a self-report that captured emotion experienced and the self-report also used a Likert scale to rate the intensity of emotion. The combination of the physiological signal and ground truth is required to build an emotion recognition model. Section 2.5.3 focuses on emotion recognition development using machine learning (ML) techniques.

2.5.3 Traditional Machine Learning Methods

This section focuses on using cleaned and pre-processed physiological signals obtained from sensor devices aligned with ground truth as input to developing emotion recognition models using traditional machine learning methods.Traditional machine learning methods use features from physiological signals manually extracted in developing an emotion recognition model.Deriving features from physiological signals is the first step. This is referred to as feature extraction. This is followed by selecting the optimum number of extracted features in developing the model. This is vital to achieving improved model performance, this process is referred to as feature selection. Afterwards, classifiers capable of using extracted, optimized and sometimes fused features in making classification and predictions are utilized in developing the emotion recognition model.Feature extraction, feature selection, feature fusion, and traditional machine learning classifiers used in building emotion recognition models with physiological signals are discussed in details in this section using relevant literature.

Feature Extraction

Feature extraction derives information from raw physiological signals and creates a new feature subspace. The aim is to compress the data into a new feature set that contains most of the required information needed for emotion recognition. Relevant feature extraction techniques used by researchers in related works are explained below.

Fast Fourier Transform (FFT) AND Short Time Fourier Transform (STFT)

Physiological signals are complex and non-stationary. Statistical features of these signals are required to build emotion recognition models. Fast-Fourier transform (FFT) can be utilised to achieve this. For example, EEG signals which are non-stationary and with features like power spectral density and spectral entropy which are also non-stationary were needed for a study that aimed to find the relationship between EEG signals and human emotions. FFT was adopted to calculate the spectrogram of the EEG channels as was done in a study by (Nie et al., 2011).Due to the inability of FFT to effectively handle unstable features required for emotion recognition, a short-time Fourier transform can be utilized to initially decompose signals into several same-length pieces, thus making each piece almost stationary after which FFT can then be utilized to extract the required features. By way of illustration, a study by Lin et al. (2009) that focused on the approach for recognizing music-induced emotional responses from brain activity utilized a short-time Fourier transform (STFT) with hanning window of time interval and with a 50% overlap to extract the power spectral density value over time from EEG signals. Overall, FFT and STFT make it possible to extract statistical features needed for emotions from physiological signals that are complex and unstable.

Wavelet Transform (WT)

With the understanding that physiological signals are non-stationary, in extracting features from these signals, small windows fit high frequency and large windows fit low frequency. The window of STFT is however fixed (Shu et al., 2018). Wavelet transform allows a free window to analyse the time-space frequency. This makes the WT a better fit to decompose physiological signals into different time and frequency scales (Shu et al., 2018).The wavelet function contains the scale parameter, the translational parameter, and the mother wavelet. The performance of a WT is greatly affected by the mother wavelet (Shu et al., 2018).There are several types of mother wavelets like Haar, Coif, Bior wavelets etc. The coefficient obtained after a wavelet transform can be used to reproduce the original signal. For example, a discrete wavelet transforms with a Db-5 wavelet was used to analyse the high-frequency coefficient at every of the five-level for the EEG frequency band (Verma and Tiwary, 2014).In summary, wavelet transform allows a flexible window to analyse and extract features within the time and frequency domain from non-stationary physiological signals.

Autoencoder

Autoencoder is an unsupervised algorithm that works based on a back propagation mechanism. Autoencoder comprises an input layer, one or more hidden layers and an output layer where the dimension of the input layer and output layer are equal. The input layer is also referred to as the encoder network and the output layer is referred to as the decoder network. The autoencoder functions by initiating the weights of the encoder network and decoder network after which the autoencoder is trained to minimize the error between the raw data and reconstructed data (Shu et al., 2018)

Feature Selection

After feature extraction, some features might be irrelevant or there could be a possible correlation between features. In cases like this, not all features are important in building an emotion recognition system. Using all features may cause over fitting to the training data and reduce the ability of the model to generalize on new data (Shu et al., 2018).This necessitates the need for feature optimization prior to emotion recognition model development. Several feature optimization techniques have been used in different related works. They include ReliefF algorithm, maximum relevance and minimum redundancy, sequential backward selection (SBS), sequential forward selection (SFS), principal component analysis (PCA) etc.

ReliefF algorithm is used due to its effectiveness and simplicity of computation (Zhang et al., 2016).The key concept behind the relief algorithm is to evaluate the quality of features according to the capacity to discriminate among samples close to each other. This capacity is qualified as the weight of each feature.A study by Zhang et al. (2016) that used a probabilistic neural network to recognise the emotions of participants while watching videos utilized the refleifF algorithm to rank 128 features. Features of channels with larger weights were selected and used in probabilistic neural networks that achieved an average accuracy of 81%.Maximum relevance and minimum redundancy use mutual information to characterize the importance of features in developing an emotion recognition system. mrMR's two criteria which are maximum relevance and minimum redundancy.Maximum relevance aims to maximize the average mutual information between each feature and the specific label while minimum redundancy aims to minimize the average mutual information between two features (Liu et al., 2016).A study by Liu et al. (2016) that aimed to model the relationship between EEG signals and emotions adopted the mrMR in feature selection. The study achieved the best result after adopting the mrMR algorithm achieving the highest performance of 71.23% and 69.97% for arousal and valence respectively.

Sequential backward selection and sequential forward selection are techniques used in selecting features relevant to the emotion recognition model. SBS starts with a full set of features and iteratively eliminates redundant features while SFS starts with an empty set of features and iteratively selects the important feature that would improve model performance. A study by Wong and Tan (2010) which proposed a particle swarm optimization of synergetic neural classifier for multimodal emotion recognition adopted the SBS together with the proposed method and this led to an increase in emotion recognition rate from 71% to 86%.Principal component analysis is a technique that projects high-dimensional data to a lower-dimensional space with a minimal loss of information. This implies that new features are created by a linear transformation of original feature values instead of selecting from a feature pool (Valenza et al., 2012). In the study by Valenza et al. (2012) PCA was adopted to arrive at features needed to develop the ER model using a Quadratic Discriminant Classifier (QDA). A recognition rate of 90.36% and 92.29% for valence and arousal was achieved.

Generally, feature selection aim to eliminate redundant features by either selecting very important features from a pool of features using algorithms like ReleifF, SBS, SFS, RFE, and mrMR or using algorithms like PCA, and independent component analysis (ICA) that transform original features into new features with minimal information loss.

Feature Fusion

Feature fusion techniques combine extracted and selected features at different levels.There are three categories of feature fusion which are early fusion (feature-level fusion), intermediate fusion and late fusion (decision-level fusion). Early fusion and Late fusion are widely used in integrating physiological signals within the emotion recognition field (Shu et al., 2018).Early and Late fusion are discussed further.

Early fusion involves integrating extracted, and selected features into a single feature set before the classification or learning phase. Hence it is also referred to as feature-level fusion. The advantage of early fusion is the use of one learning phase and the benefit of mutual information from the data (Ayata et al., 2020). A study by (Verma and Tiwary, 2014) that investigated emotion representation models and recognized emotion from multimodal physiological signals (EEG and eight peripheral physiological signals) using a multi resolution approach adopted the early fusion technique. Feature-level fusion was used to combine the energy-based features extracted from 32-channel EEG signals. The combined feature vector was trained and achieved a recognition rate of 81.45% for 13 emotional states using a support vector machine classifier.Late fusion entails processing the classification results of prior learning stages by combining the results from multiple classifiers to give a fused final decision, hence it is also referred to as decision-level fusion or fusions of classifiers (Ayata et al., 2020). The goal of late fusion is to utilize the redundancy of a set of independent classifiers to achieve greater robustness by combining their results (Ayata et al., 2020).The study by Ayata et al. (2020) proposed a novel emotion recognition algorithm from multimodal physiological signals for emotion-aware healthcare systems by adopting the decision-level fusion. Random forest classifiers were used to train features extracted from RB, PPG, and FTT signals individually. The result from each classifier was then combined to yield a fused result.The study achieved a significant increase in accuracy from 69.86% to 73.08% for arousal and from 69.53% to 72.18% for valence. The study indicated that multimodal physiological signals and decision level fusion improved the accuracy of emotion recognition systems.

Emotion Recognition Model Development

Emotion recognition models typically use feature set from physiological data source to assign discrete emotional states or emotional dimensions. Several traditional machine learning classifiers have been used in developing these ER systems, they include support vector machine (SVM), random

Aim of the study	ML Methods	Physiological signals	Accuracy (%)
To identify the elicitation of acoustic stimulation using a wearable device on a test population.	SVM,RF,DT, NB,KNN, BAGGING, BOOSTING,LR	IBI, EDA, SKT, BVP	75 (SVM), 72.62 (LR)
To propose a novel personalized probabilistic framework able to characterize the emotional state of a subject using heartbeat dynamics	SVM	ECG	79.15 (Valence), 83.35(Arousal)
To presents a thorough analysis of psychological properties of ECG signals	LDA	ECG	76.19 (arousal), 52 to 89 (valence)
To report a new methodology for the automatic assessment of emotional response	QDA	ECG, EDR, RSP	90
To present an innovative and intelligent solution for the monitoring of patients with anxiety disorders during therapeutic sessions.	ANN, SVM	BVP, ECG, EDR, RSP	84

Table 1: Related works within Emotion Recognition Field

forest (RF), linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA),linear regression (LR), k-nearest neighbour (KNN), decision tree (DT), artificial neural network (ANN) to mention a few. Table 1 presents a related works that have developed emotion recognition model using physiological signals and identifies the traditional machine learning classifiers used within the studies.The emotion recognition rate have been.

2.5.4 Model Evaluation and Performance

Model Evaluation

After the development of emotion recognition models, these models need to be evaluated for reliability and ability to generalise on new data. The concept of testing this ability can be referred to as model evaluation. The ability of the models to generalise well can be tested by inputting new physiological signals for the model to predict the target or ground truth. The testing error on the new data set is the generalization error (Shu et al., 2018). Two major methods are used in model evaluation: The hold-out method and the cross-validation method.The hold-out method divides the dataset into a training set and a test set while maintaining consistency of data distribution within both sets. About 60% - 80% of the dataset is usually used to train the model with the remainder used as the test set. This method runs several iterations using random division and presents an evaluation result based on the test set.The cross-validation method could either be K-fold cross-validation or Leave-one-out cross-validation. The leave one out cross validation uses just one subset as the testing set and the remainder as the training set. The test set is not introduced to the model at all but is just used to evaluate the model once. For K-Fold cross-validation, the data set is divided into K- subsets. One subset is used as the test set and the K-1 minus subset left is used as the training set. The cross-validation is repeated K times. Each subset is validated once, and the average of the k times is the final result.

Model Evaluation

There are defined metrics used to evaluate models developed to infer the abilities of the emotion recognition model on new data. These metrics are explained below.

Precision, Recall, and F1

Precision is a measure that determines how precise a model is by how many "true positives" were out of the entire predicted outputs. Recall is a measure that determines how many true positives are there out of the actual positives. F1 is defined as the harmonic mean of precision and recall.

Receiver Operating Curve

A receiver operating curve (ROC) is created by plotting the true positive rate against the false positive rate. Conventionally, the false positive rate is plotted on the horizontal axis. When the ROC is plotted, the closer the plot is to the top left, the higher the accuracy and the better the classifier. The area under the ROC is termed Area under the curve (AUC). The bigger the AUC, the better the classifier.

2.6 Summary

In summary, this chapter discussed emotion and it means of measurement. This chapter expanded on emotional signals used in developing emotion recognition models. This was followed by an in-depth explanation about different means of capturing and recording emotional signals. Finally, the emotion recognition model development framework using traditional machine learning methods was discussed making references to related works.The next chapter presents the data collection methods used to obtain physiological signals using the wrist-worn smart device. Statistical analysis of the physiological signals obtained was done and discussed within chapter 3 .

Chapter Three

3 Data Collection and Analysis

3.1 Introduction

This chapter presents the steps and methods utilised in collecting physiological signal from research participants. Physiological signals obtained were aligned with ground truth and analysed in this chapter. Figure 2 presents the overview of the chapter. This chapter is divided into two major parts: data collection and data analysis. Data collection explains the techniques and methods used in collecting, pre-processing, and creating the dataset. Data analysis discusses the statistical analysis and hypotheses testing done.The results of the hypotheses tests are presented in this chapter. In greater detail, this chapter describes:

- The choice of emotion induction used during data collection
- The selection of smart device used during data collection
- The development of data annotation form for collection of ground truth
- The collection, data pre-processing creation of the dataset
- The statistical tests conducted using the created dataset.

3.2 Data Collection

Data was collected in a controlled environment using a chosen smart device embedded with sensor technology to achieve the aim of the study. Ground truth was collected using data annotation form developed for the purpose of this study. Sections 3.2.1 - 3.2.6 explains emotion induction method used, the wrist-worn smart device selected, development of data annotation form, research participants, data collection protocol and data export.

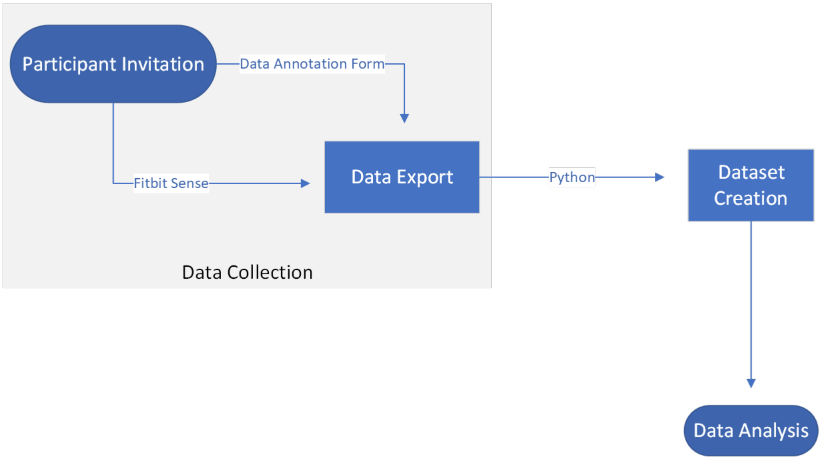


Figure 2: Data collection and analysis pipeline.

3.2.1 Emotion Induction

In day-to-day living, humans experience different emotions due to different stimuli received Plutchik (2001). This occurs naturally. As such, in building the emotion recognition model, a dataset of physiological signals correctly labelled with emotion experienced at different time intervals must be obtained.In previous works as discussed in section 2.5.1, emotions have been induced using audio, visual, and audio-visual stimuli like music, video clips, movies, sounds in different related studies (Koelstra et al., 2012; Shu et al., 2018; Ayata et al., 2020). This study explored the use of audio stimuli to elicit emotions.

The IADS-E audio stimuli database which is an expanded version of the international affective digitized sounds (IADS-2) was within this study (Yang et al., 2018). The IADS-E database contains standardised, emotionally evocative stimuli with sounds from a greater range of categories (Yang et al., 2018). For the purpose of this research, the IADS-E stimuli database excluded the audio stimuli from the original international affective digitized sound database (IADS-2). In total, 768 audio stimuli of the IADS-E divided into 10 different categories were available to be used within this study. Table 2 shows the breakdown of audio stimuli groups available within the IADS-E database. 4 of the 10 groups of were selected to elicit emotion during data

Categories	Number of Sound Files (N)
Animal	36
People	22
Nature	63
Daily Routine	159
Transport	44
Electric	59
Breaking Sound	52
Music	158
Scenarios	04
Sound Effects	171

Table 2: Breakdown of the IADS-E Audio Stimuli.(Yang et al., 2018)

collection: Animal, Nature, Transport and People. These four groups were selected because:

- The sound file within these groups encompassed and mimicked sound stimuli humans receive and are exposed to in day-to-day living
- The sound file within these groups are natural sounds readily available in the real world and thus applicable in social care setting.

3.2.2 Smart Device

The choice of smart device used for this study was informed by usability within a nursing facility. Hence the device had to fit the following:

- Easy to use: The device had to be easy to wear by residents in nursing facilities. Generally, these devices could be worn on the wrist, chest, or ankle but a wrist-worn device was preferred because it can be worn on an easily accessible part of the body

- Multi-modal physiological signal collection: The device had to be able to collect multimodal physiological signals using sensor technology simultaneously.

Considering these criteria, several wrist-worn smart devices fit these criteria and could possibly have been explored for the study. However, this study explored the Fitbit Sense wrist-worn smart-watch because of its low-cost.The smart device was used within this study after considering the affordability of the smart device for residents within nursing facilities.The Fitbit Sense has the following specification:

- Screen: 1.58 inches organic light-emitting diodes
- Case size: 40.5 millimetres
- Case thickness: 12.35 millimetres
- Weight: 45.9 grams
- Water resistance: IP68, 50 metres (5 atmospheres)

The Fitbit Sense is made up of the following sensors which was relevant to this study:

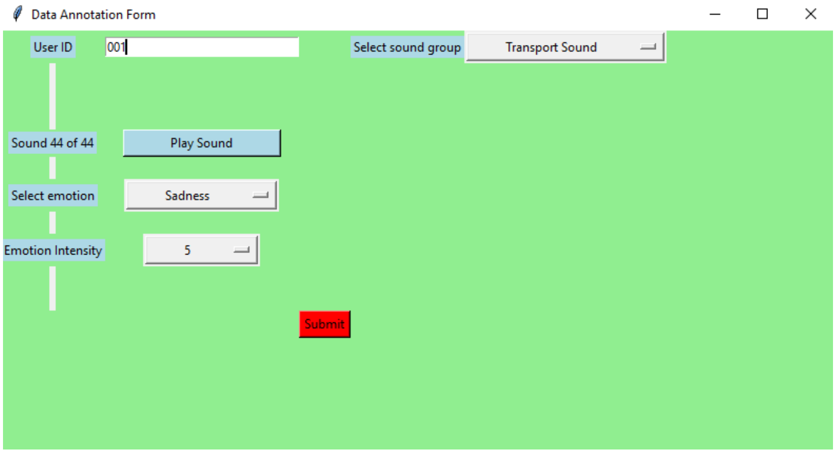
- Multi-path optical heart rate tracker
- Multipurpose electrical sensors
- On-wrist skin temperature sensor.

The multi-path optical heart rate tracker reads the heart rate from multiple positions and combines the reading to give an accurate read of the heart rate ((Fit, 2022). The multipurpose electrical sensors captures ECG signals while the on-wrist temperature sensor records the skin temperature (Fit, 2022). This study utilized all three sensors to obtain multimodal physiological signals which would be analysed and used to develop the emotion recognition model.

Figure 3: Wrist-worn FitBit Sense Smart Device.(Fit, 2022)

3.2.3 Data Annotation Form

Data annotation form was developed to induce and collect self report of emotions and intensity of emotions experienced by participants.Figure 4 shows the outlook of the data annotation form. The form was developed using python programming language and the graphical user interface module Tkinter (Lundh, 1999).The form collected research participants' unique identification numbers called *userid*. The form worked by reading sound files from selected audio stimuli groups. Participants selected one audio stimuli group containing its sound files and listened to the sound files within this group one at a time.This induced emotion. Emotions experienced (fear, sadness and happiness) is selected and intensity rated using a Likert scale ranged from 1 to 9. All selections made by the participants using the platform were saved and stored as a CSV file.



Data Annotation Form

User ID 001

Select sound group Transport Sound

Sound 44 of 44

Select emotion Sadness

Emotion Intensity 5

Figure 4: Data Annotation Form

3.2.4 Research Participants

The research was conducted with three participants who were all male, apparently healthy and between the ages of 18 to 40. All participants were briefed about the experiment and written informed consent was obtained from the participants. Data obtained from research participants were anonymised before processing.

3.2.5 Data Collection Protocol

Participants were invited to a soundproof room and asked to create a Fitbit account. Participants were guided through the data collection procedure using practice sounds. Twelve sound files which were not part of the selected audio stimuli groups were randomly selected and labelled “practice sound”. These practice sounds were used to guide the participants through the data collection. The participants sat about 1 metre from the display screen in the soundproof room. Participants wore the Fitbit Sense smart device on the dominant wrist. The experiment was conducted in a soundproof room to eliminate distraction and aid participants’ concentration.

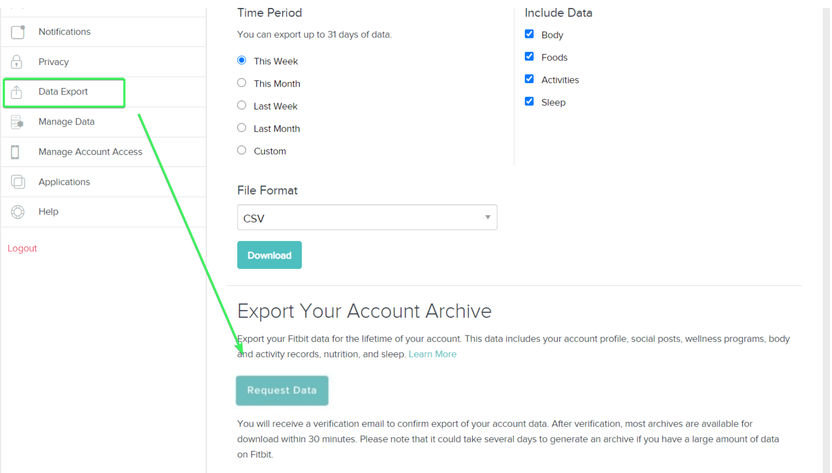
All sound within the selected audio stimuli groups were six seconds long. Data collection was executed in sessions. Each session entailed selecting one of the four audio stimuli groups, listening to all sound files within the group, selecting the emotion experienced and rating the intensity of the emotion experienced for each sound stimuli. All of these was done using the data annotation form described in section 3.2.3. A 2 to 5 minute break was taken between each session to ensure each participants were not fatigued. Physiological signals were recorded continuously during each session with the Fitbit sense. At the end of the data collection, each participant had their physiological signals captured and recorded with the Fitbit Sense device and the ground truth captured by the data annotation form and stored as a CSV file.

3.2.6 Data Export

The ground truth was readily available as a CSV file after data collection, but the physiological signals and data were stored in different formats by the Fitbit Sense device and only available after 24 hours to be exported.Each Participant used the created Fitbit account to access the Fitbit dashboard from which the physiological signal data was exported. Figure 5 illustrates how the the physiological data was exported from the Fitbit dashboard.Following a successful data export, a zipped folder containing all data captured by the device is obtained. The different data captured by the device are stored in sub-folder data as JSON files or CSV files. Physiological signals could be extracted from the appropriate sub-folder for processing.The folder structure obtained after successfully exporting data is visualised in figure 6.

3.3 Data Pre-processing

This section explains the data pre-processing steps carried out after data collection.The pre-processing steps executed within this study aimed to align the physiological signals obtained from the Fitbit Sense smart device with the ground truth obtained form the data annotation form. This implies that at the end of data pre-processing, a dataset that contains the physiological signals aligned with the ground truth would have been created.



Time Period

You can export up to 31 days of data.

● This Week
○ This Month
○ Last Week
○ Last Month
○ Custom

Include Data

☑ Body
☑ Foods
☑ Activities
☑ Sleep

File Format

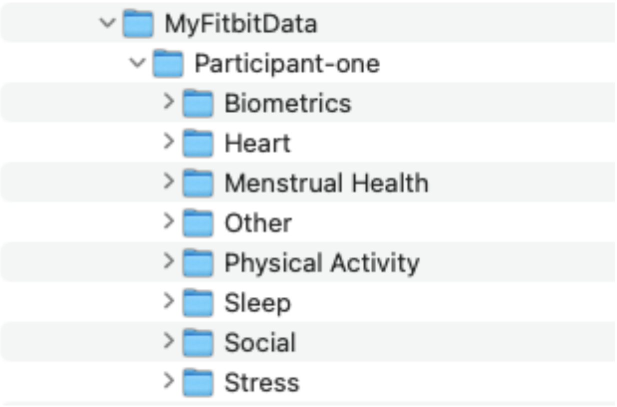
CSV

Export Your Account Archive

Export your Fitbit data for the lifetime of your account. This data includes your account profile, social posts, wellness programs, body and activity records, nutrition, and sleep. Learn More

You will receive a verification email to confirm export of your account data. After verification, most archives are available for download within 30 minutes. Please note that it could take several days to generate an archive if you have a large amount of data on Fitbit.

Figure 5: Fitbit Dashboard for Data Export



```
MyFitbitData
  Participant-one
    Biometrics
    Heart
    Menstrual Health
    Other
    Physical Activity
    Sleep
    Social
    Stress
```

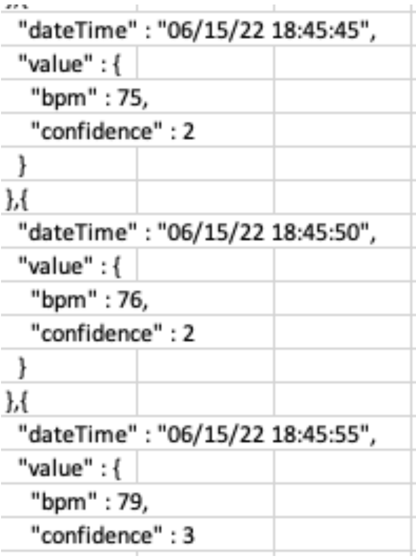
Figure 6: Fitbit data structure obtained after data export.(Fit, 2022)

Modules	Version
python	3.9.13
numpy	1.22.4
pandas	1.4.2
matplotlib	3.5.2
seaborn	0.11.2
scikit-learn	1.1.1
scipy	1.8.1

Table 3: Modules used for the purpose of this study

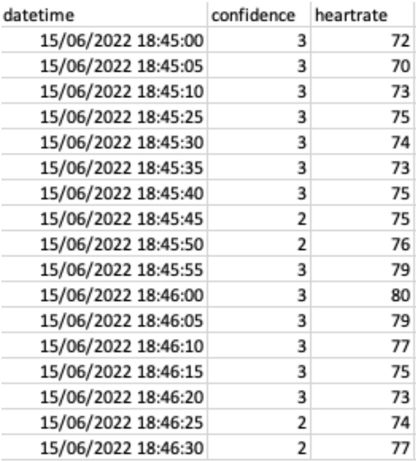
Data pre-processing was carried out using python programming language in Jupyter Notebook run on a local computer. Versions of modules used within this study are listed in Table 3. The heart rate and skin temperature are the physiological data captured automatically by Fitbit Sense during data collection.These signals were pre-processed. These signals were used because the smart device could capture and record these physiological signals automatically without further steps, indicating a suitable application in a social care setting.

The heart rate was captured by the Fitbit Sense Smart Device and stored as a JSON object named *heart_rate*.This file was obtained from the "physical activity" sub-folder.Each object in the JSON file contained date-time, beat per minute (bpm), and confidence. The datetime indicated the time the heart rate was captured. The bpm is the heart rate recorded. The bpm was aggregated every 5 seconds with a single value recorded per object in the JSON file. The confidence indicated the accuracy of the Fitbit device with 1 signifying low confidence and 3 indicating the highest confidence. The first step to processing was reading the data into a data frame and saving the file in a tabular format which is easy to work with. Figure 7 visualises the JSON



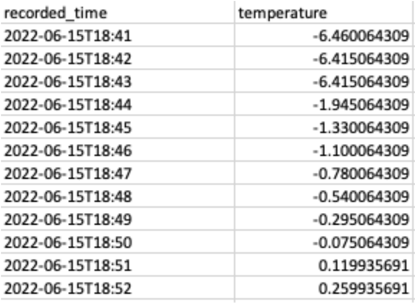
```
 "dateTime" : "06/15/22 18:45:45",
 "value" : {
   "bpm" : 75,
   "confidence" : 2
 }
},{
 "dateTime" : "06/15/22 18:45:50",
 "value" : {
   "bpm" : 76,
   "confidence" : 2
 }
},{
 "dateTime" : "06/15/22 18:45:55",
 "value" : {
   "bpm" : 79,
   "confidence" : 3
```

Figure 7: heart_rate JSON file before pre-processing visualised in Microsoft Excel.



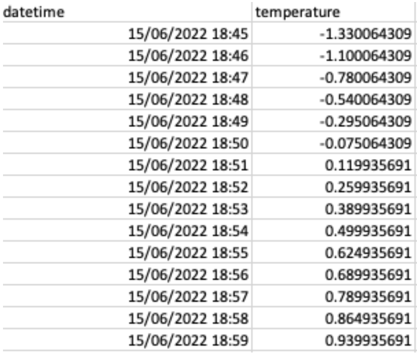
datetime	confidence	heartrate
15/06/2022 18:45:00	3	72
15/06/2022 18:45:05	3	70
15/06/2022 18:45:10	3	73
15/06/2022 18:45:25	3	75
15/06/2022 18:45:30	3	74
15/06/2022 18:45:35	3	73
15/06/2022 18:45:40	3	75
15/06/2022 18:45:45	2	75
15/06/2022 18:45:50	2	76
15/06/2022 18:45:55	3	79
15/06/2022 18:46:00	3	80
15/06/2022 18:46:05	3	79
15/06/2022 18:46:10	3	77
15/06/2022 18:46:15	3	75
15/06/2022 18:46:20	3	73
15/06/2022 18:46:25	2	74
15/06/2022 18:46:30	2	77

Figure 8: heart_rate file after pre-processing visualised in Microsoft Excel.



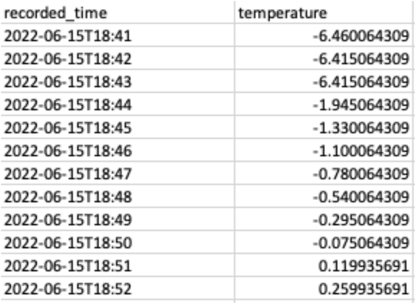
recorded_time	temperature
2022-06-15T18:41	-6.460064309
2022-06-15T18:42	-6.415064309
2022-06-15T18:43	-6.415064309
2022-06-15T18:44	-1.945064309
2022-06-15T18:45	-1.330064309
2022-06-15T18:46	-1.100064309
2022-06-15T18:47	-0.780064309
2022-06-15T18:48	-0.540064309
2022-06-15T18:49	-0.295064309
2022-06-15T18:50	-0.075064309
2022-06-15T18:51	0.119935691
2022-06-15T18:52	0.259935691

Figure 9: skin_temp CSV file before pre-processing visualised in Microsoft Excel.



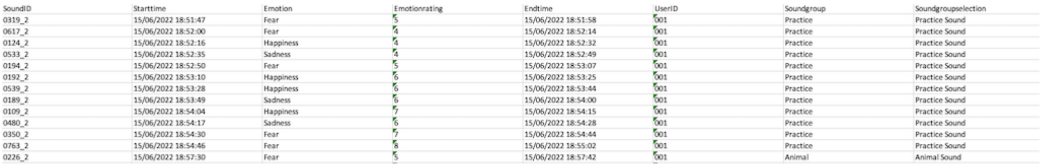
datetime	temperature
15/06/2022 18:45	-1.330064309
15/06/2022 18:46	-1.100064309
15/06/2022 18:47	-0.780064309
15/06/2022 18:48	-0.540064309
15/06/2022 18:49	-0.295064309
15/06/2022 18:50	-0.075064309
15/06/2022 18:51	0.119935691
15/06/2022 18:52	0.259935691
15/06/2022 18:53	0.389935691
15/06/2022 18:54	0.499935691
15/06/2022 18:55	0.624935691
15/06/2022 18:56	0.689935691
15/06/2022 18:57	0.789935691
15/06/2022 18:58	0.864935691
15/06/2022 18:59	0.939935691

Figure 10: skin_temp CSV file after pre-processing visualised in Microsoft Excel.



recorded_time	temperature
2022-06-15T18:41	-6.460064309
2022-06-15T18:42	-6.415064309
2022-06-15T18:43	-6.415064309
2022-06-15T18:44	-1.945064309
2022-06-15T18:45	-1.330064309
2022-06-15T18:46	-1.100064309
2022-06-15T18:47	-0.780064309
2022-06-15T18:48	-0.540064309
2022-06-15T18:49	-0.295064309
2022-06-15T18:50	-0.075064309
2022-06-15T18:51	0.119935691
2022-06-15T18:52	0.259935691

Figure 11: skin_temp CSV file before pre-processing visualised in Microsoft Excel.



SoundID	Starttime	Emotion	Emotionrating	Endtime	UserID	Soundgroup	Soundgroupselection
0319_2	15/06/2022 18:51:47	Fear	5	15/06/2022 18:51:58	001	Practice	Practice Sound
0617_2	15/06/2022 18:52:00	Fear	4	15/06/2022 18:52:14	001	Practice	Practice Sound
0124_2	15/06/2022 18:52:16	Happiness	4	15/06/2022 18:52:32	001	Practice	Practice Sound
0533_2	15/06/2022 18:52:35	Sadness	4	15/06/2022 18:52:49	001	Practice	Practice Sound
0194_2	15/06/2022 18:52:50	Fear	5	15/06/2022 18:53:07	001	Practice	Practice Sound
0192_2	15/06/2022 18:53:10	Happiness	6	15/06/2022 18:53:25	001	Practice	Practice Sound
0539_2	15/06/2022 18:53:28	Happiness	6	15/06/2022 18:53:44	001	Practice	Practice Sound
0189_2	15/06/2022 18:53:49	Sadness	6	15/06/2022 18:54:00	001	Practice	Practice Sound
0109_2	15/06/2022 18:54:04	Happiness	7	15/06/2022 18:54:15	001	Practice	Practice Sound
0480_2	15/06/2022 18:54:17	Sadness	6	15/06/2022 18:54:28	001	Practice	Practice Sound
0350_2	15/06/2022 18:54:30	Fear	7	15/06/2022 18:54:44	001	Practice	Practice Sound
0763_2	15/06/2022 18:54:46	Fear	8	15/06/2022 18:55:02	001	Practice	Practice Sound
0226_2	15/06/2022 18:57:30	Fear	5	15/06/2022 18:57:42	001	Animal	Animal Sound

Figure 12: df_label visualised in Microsoft Excel.

file before processing while Figure 8 shows the JSON file after processing.Skin temperature was obtained from the “sleep” folder as a CSV file. The skin temperature captured by the device was aggregated every 60 seconds with one value recorded every minute. The file contained the date-time, and the skin temperature recorded every minute. The date-time was harmonized, and the column header was renamed during processing. Figure 11 shows the skin temperature dataset before while figure 10 after data processing.The ground truth that was captured by the data annotation form as a CSV file was read as df_label. df_label is displayed in figure 12 using microsoft excel.

3.4 Dataset Creation

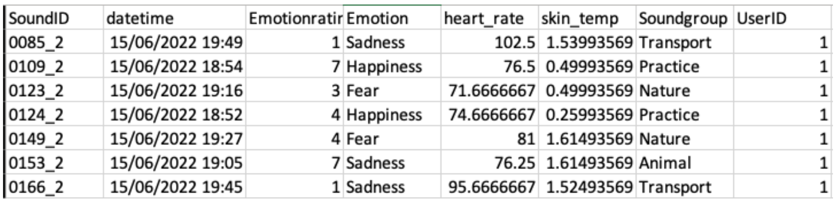
This subsection discusses how the pre-processed physiological data and the ground truth are aligned to create the dataset for statistical analysis.The *heart_rate*, *skin_temp* and ground truth obtained after data pre-processing were merged using date-time as the primary key. For the ground truth, the *starttime* column was retained, and a new column *listentime* was created from it by adding six seconds to the start time. *listentime* column indicates the time each participant finished listening to each sound as each sound was exactly six seconds long. The *endtime* column was retained.*endtime* signified the time the participant submitted each entry.Using these three date-times (*starttime*, *listentime*, *endtime*) from *df_label*, three new dataframes were created using the datetimes. The *starttime* was the date time in the first data frame *(df_start)*, the *listentime* was the date time in the second data frame *(df_listen)* and the *endtime* was the datetime in the third data frame *(df_end)*. For each of these new three data frames created the emotion experienced, intensity of emotion and sound identification number (soundID) were retained.The heart_rate was read into a data frame *(df_heartrate)* while the skin_temperature was read into *df_skintemp.*The skin temperature which contained a value per 60 seconds was duplicated each value into 12 to match the length of *df_heartrate.*This implied an assumption was made that the skin temperature remained constant for every 60 seconds. The two dataframes *(df skin_temp)* were merged using the nearest date times to create a new data frame *df_signals.*

The three new data frames (*starttime*, *listentime*, *endtime*) created using *df_label* was then merged with the dataframe containing the physiological signals *(df_signals).* The merge was done on date time using the nearest date-time from each file to merge each other. An allowance of three seconds was given to ensure that each row from the physiological signal dataframe *(df_signals)* was merged to the respective data label row on date time. This created three data frames containing merged signals and respective data labels. The three data frames were concatenated to create one data frame *(df_combined)* such that each sound had three instances.The groupby method was applied to the combined dataframe *(df_combined)* to group the dataframe by soundID. The end time (datetime), emotion rating, and the emotion experienced. The mean of the heart rate and skin temperature at *starttime*, *listentime*, and *endtime* were grouped to create the dataset *(df_dataset).* The *endtime* was retained in *df_dataset.* It was saved as a CSV file to be used in statistical data analysis. A sample of the resulting data frame following the process is shown in figure 13.

The data creation steps was repeated for all the participants to create a *df_dataset* for each participant with which statistical analysis can be conducted. At the end of data creation. Each dataset contained 8 variables and 167 instances for each participant. Each instance represented an instance for each sound listened to by the participant during data collection. Table 4 explains in greater detail the 8 variables.

S/N	Variable name	Variable details	Capture Device
1	Heart rate	This is the heart rate of the participants aggregated every 5 seconds using the optical heart rate sensor	Fitbit
2	Skin_temperature	This is the temperature variation of the wrist captured every 60 seconds by the on-wrist temperature sensor	Fitbit
3	SoundID	The sound file name used to identify each of the sounds played. Unique to every sound	Data annotation form
4	Emotion	The emotion experienced after listening to each sound	Data annotation form
5	Emotionrating	The Likert scale rating for emotion experienced. Ranged from 1 to 9	Data annotation form
6	Soundgroup	The group of sound stimuli which could be people, animals, transport, or Nature	Data annotation form
7	Datetime	The exact datetime the submit button was clicked to log and submit each entry for each sound	Data annotation form
8	UserID	Assigned identification number unique to research participants	Data annotation form

Table 4: Explanation of variables in *df_final*



SoundID	datetime	Emotionratir	Emotion	heart_rate	skin_temp	Soundgroup	UserID
0085_2	15/06/2022 19:49	1	Sadness	102.5	1.53993569	Transport	1
0109_2	15/06/2022 18:54	7	Happiness	76.5	0.49993569	Practice	1
0123_2	15/06/2022 19:16	3	Fear	71.6666667	0.49993569	Nature	1
0124_2	15/06/2022 18:52	4	Happiness	74.6666667	0.25993569	Practice	1
0149_2	15/06/2022 19:27	4	Fear	81	1.61493569	Nature	1
0153_2	15/06/2022 19:05	7	Sadness	76.25	1.61493569	Animal	1
0166_2	15/06/2022 19:45	1	Sadness	95.6666667	1.52493569	Transport	1

Figure 13: Final dataset *(df_dataset)* visualised in Microsoft Excel

3.5 Data Analysis

The descriptive statistics for heart rate and skin temperature is presented in table 5. Using table 5 for participants one, the mean of the heart rate of was 85.24, and the median was 82. The skin temperature had a mean value of 1.38. Observation from participant two returned mean values of 74.97 and -0.90 and median values of 70 and -0.9 for heart rate and skin temperature respectively. Observations from participant three returned mean values of 64.96 and 0.11 and median values of 64.33 and 0.21 for heart rate and skin temperature.

This gives an indication that the heart rate and skin temperature are unique to each participant. The range of value from minimum to maximum for heart rate and skin temperature were unique to each participant .This is expected because physiological signals of humans differ based on a number of factors which include age, sex, lifestyle (Lin et al., 2021). This finding gave an indication that a personalised approach to developing an emotion recognition model might be a good methodology to execute and obtain a robust emotion recognition model per participant.

Statistics	Participant 1		Participant 2		Participant 3	
Descriptive	heart_rate	skin_temp	heart_rate	skin_temp	heart_rate	skin_temp
count	167.00	167.00	167.00	167.00	167.00	167.00
mean	85.24	1.38	74.97	-0.90	64.96	0.12
std	12.56	0.38	15.39	0.176	4.20	0.44
min	65.00	0.17	54.00	-1.32	56.50	-0.60
25%	74.00	1.46	60.00	-1.05	62.00	-0.37
50%	82.00	1.54	70.00	-0.91	64.33	0.22
75%	97.00	1.59	91.00	-0.80	67.33	0.54
max	111.50	1.66	102.33	-0.58	79.67	0.66

Table 5: Descriptive Statistics for all three participants

Considering the observations collected from the participants, statistical assumptions had to be validated.Thus certain hypotheses were tested to validate and test statistical assumptions. The hypotheses are listed below:

- Are the data samples normally distributed across emotions and audio stimuli groups?
- Are there significant differences between physiological signal data when different emotions are experienced by participants?
- Are there significant differences between physiological signal data when groups of audio stimuli were listened to by participants?

3.5.1 Normality Hypothesis Test

Due to the relatively small size of observations (N=167) collected from the three participants, determining the normality of distribution of the heart rate and skin temperature was important for selecting the appropriate statistical method required to test the other hypotheses.Thus, the normality hypothesis test checks if the population the sample represents is normally distributed. The null hypothesis states that the sample is normally distributed. A conventional p-value of 0.05 was set. The results of Shapiro-Wilk test conducted to test the normality hypothesis of the distribution of heart rate and skin temperature observations per participant are presented.As presented in table 6, skin temperature observations from the three participants revealed that for all emotion groups (fear, sadness and happiness) distributions did not appear normal or gaussian except participant 3 in the sadness group (W=0.97, p-value=0.09). As presented in table 7, skin_temperature observations from the three participants for all audio stimuli groups (animal, nature, people, transport) did not appear normal or gaussian except participant 1 in the people group (W=0.91, p-value=0.05).

Group	Participant 1		Participant 2		Participant 3	
	statistic (W)	p-value	statistic (W)	p-value	statistic (W)	p-value
Fear	0.71	0.00	0.88	0.00	0.93	0.00
Sadness	0.59	0.00	0.77	0.00	0.97	0.09
Happiness	0.56	0.00	0.87	0.00	0.93	0.03

Table 6: Normality Test Results of skin_temp across emotion groups.

Group	Participant 1		Participant 2		Participant 3	
	statistic (W)	p-value	statistic (W)	p-value	statistic (W)	p-value
Animal	0.84	0.00	0.79	0.00	0.88	0.00
Nature	0.44	0.00	0.92	0.00	0.89	0.00
People	0.91	0.05	0.85	0.01	0.89	0.02
Transport	0.28	0.00	0.92	0.01	0.82	0.00

Table 7: Normality Test Results of skin_temp across audio stimuli groups.

Group	Participant 1		Participant 2		Participant 3	
	statistic (W)	p-value	statistic (W)	p-value	statistic (W)	p-value
Fear	0.86	0.00	0.93	0.00	0.86	0.00
Sadness	0.91	0.00	0.89	0.03	0.84	0.00
Happiness	0.90	0.00	0.96	0.01	0.89	0.00

Table 8: Normality Test Results of heart_rate across emotion groups.

Table 8 and 9 presents the Shapiro-wilk test results of heart rate observations both sound groups and emotion groups. As shown in table 8, all heart rate observation for all three participants for emotions experienced did not

Group	Participant 1		Participant 2		Participant 3	
	statistic (W)	p-value	statistic (W)	p-value	statistic (W)	p-value
Animal	0.80	0.00	0.96	0.01	0.97	0.43
Nature	0.84	0.00	0.91	0.00	0.83	0.00
People	0.95	0.35	0.93	0.11	0.89	0.02
Transport	0.74	0.00	0.98	0.78	0.61	0.00

Table 9: Normality Test Results of heart_rate across audio stimuli groups.

appear gaussian or normal. The result in table 9 showed that the heart rate observations for the audio groups did not appear normal except for participant 3 in animal group (W=0.97,p-value=0.43) and participant in people sound group (W=0.95,p-value=0.35).

Overall, it was observed that most of the observations for skin temperature and heart rate across emotion and sound groups did not appear normally distributed.Hence a non-parametric statistical method was chosen to evaluate the other hypotheses.

3.5.2 Significant differences between data observations for emotion groups

As established in section 3.5.1, for participants, most observation of the audio stimuli (sound) and emotion groups were not normally distributed. Hence, a Kruskal-wallis one-way ANOVA which is a non-parametric statistical test was conducted to determine if there are significant differences between skin temperature and heart rate observation when different emotions (happiness, sadness and fear) are experienced.A conventional p-value of 0.05 was set.With the Kruskal-wallis ANOVA, small p values lesser than the set p-value led us to reject the null hypothesis and assume that there are significant differences within our observation.

The null hypothesis states that heart rate observations do not differ significantly among the three different emotions (happiness, sadness and fear).

For participant one (H=0.30, p=0.86), participant 2 (H=0.39, p=0.82), and participant 3 (H=0.43, p=0.81) there was no significant difference. Thus when different emotions were experienced by all three participants there was no significant difference in observations of the heart rate capture.Hence, we fail to reject the null hypothesis.For the skin temperature observations for the three participants, the null hypothesis states that the observations do not differ significantly when the three emotions were experienced.Except participant 1 (H=9.34, p=0.01) that had significant difference in skin temperature observations when experiencing different emotions, there was no significant difference in skin temperature observations for participant 2 (H=5.10, p=0.08) and participant 3 (H=5.78, p=0.06).Thus we fail to reject the null hypothesis for participant 2 and 3.

3.5.3 Significant differences between physiological data observations for sound of audio stimuli groups

Similar to section 3.5.2 ,Kruskal-wallis one-way ANOVA was done to determine if there are significant differences between skin temperature and heart rate observation when sounds from the different audio stimuli group were listened to by the participants.

With the heart rate observation, the hypothesis states that there are no significant differences between heart rate observations when different sounds of each audio stimuli group were listened to. For participant 1 (H=73.61, p=0.00), participant 2 (H=20.45, p=0.00), and participant 3 (H=104.70, p=0.00) there were significant differences in observations of the heart rate when different sound were listened to from the audio stimuli groups. Thus we reject null hypothesis and accept the alternative hypothesis that there were significant differences in heart rate observation when participants listened to sounds from the different audio stimuli groups.For skin temperature observations, the hypothesis states that there are no significant difference in skin temperature observations when different sounds of the audio stimuli groups were listened to. For participant 1 (H=43.82, p=0.00), participant 2 (H=127.80, p=0.00), and participant 3 (H=137.93, p=0.00), there were significant differences in skin temperature observations for all participants when they listened to sounds of the different audio stimuli groups. Thus we reject the null hypothesis and accept the alternative hypothesis that there was significant differences in skin temperature observations when participants listened to sounds from the different audio stimuli groups.

From the descriptive and inferential statistical test conducted, the following were established:

- Physiological signal data (heart rate and skin temperature) for each participant differs, with different range for heart rate and skin temperature observed per participant. This indicates the need for a personalised or subject-dependent approach in developing an emotion recognition model that will be applicable within nursing facilities.
- There were significant differences between observations of the skin temperature and heart rate when sound from the different audio stimuli group were listened to. This indicates that physiological signals of participants differed upon receiving external stimuli.(Cosoli et al., 2021)

These findings informed the development of emotion recognition model using a subject dependent approach.

3.6 Further Findings

After establishing that physiological signals obtained from each participant differs after receiving stimuli, an attempt to develop a subject-dependent emotion recognition model for each participants was made.Using the dataset created in section 3.4 which contained skin temperature and heart rate as the two features to develop an emotion recognition model. The results of the models developed using the two features achieved an accuracy of 44% for classification of three discrete emotion:sadness, fear and happiness.

The results achieved was due to the fact that the skin temperature and heart rate captured automatically by Fitbit Sense were not raw physiological signals. The signals have been pre-processed and aggregated by Fitbit. Thus, feature extraction could not be used to derive features needed to train and develop an emotion recognition model using traditional machine learning methods. Hence, these signals (heart rate and skin temperature) were insufficient to develop a robust emotion recognition model.With these findings, a new approach was taken to create and develop a robust emotion recognition model. Chapter 4 presents the approach used to develop the emotion recognition model.

Chapter Four

4 Methodology

4.1 Introduction

This section presents the methods and experimental set-up used to develop a subject-dependent emotion recognition model for this thesis project. The method framework is visualised in figure 14. With the findings explained in section 3.6, a personalised approach was taken to develop the emotion recognition model.Thus a second experiment to collect physiological signals using just one participant (Participant 1) was conducted.Therefore,the emotion recognition model developed was done with physiological signals obtained from this research participant.

4.2 Emotion Induction

The audio stimuli groups of the the IADS-E database selected in section 3.2.1 was retained to induce emotions.Sounds from the audio stimuli initially rated by the research participant during data collection in section 3.2.5 was divided into two broad groups: strong and weak emotions. For a better understanding,sounds rated 1 to 4 by the participant were the sounds that elicited weak emotions while sounds rated from 5 to 9 were the sounds that elicited strong emotions.For both strong and weak emotions reported by the participant, the sound files were selected into sub-groups of discrete emotions (happiness, sadness and fear). This means weak emotions had sounds selected into fear, sadness and happiness sub-groups, same for strong emotions.Figure 15 shows a breakdown of the strong emotions folder structure.Weak emotions had a total of forty-five sound stimuli.Using the ground truth obtained during data collection discussed in chapter 3.2.5, 15 sounds each were selected into sadness, fear and happiness subgroups. Strong emotions had a total of forty-two. 15 sounds were selected in to fear and happiness sub groups respectively and 12 into the sadness subgroup. This meant a total of 87 sounds were selected to induce emotions.

4.3 Data Gathering

Data annotation form developed and discussed in section 3.2.3 was refactored to collect ground truth for this second experiment. The research participants listened to the selected sounds and followed the data collection protocol discussed in section 3.2.5. However, for the purpose of this second experiment, ECG signals (waveform samples) and aggregated heart rate reading were collected.The research participant listened to the sound files for 6 seconds and physiological signals was collected for 30 seconds accounting for sustained physiological changes experienced after listening to the audio stimuli.The physiological signal was collected by holding the alternate diagonal corners of the FitBit Sense with the thumb and index finger. Figure 16 illustrates the positioning of the thumb and index finger during data collection.Data annotation form collected the same variables as discussed in section 3.2.3. These variables were saved as a CSV file. ECG signals was exported from the Fitbit dashboard as discussed in section 3.2.6.

4.4 Data Pre-processing

The data exported from the Fitbit sense contained of relevance the following; timestamp of data collection, heart rate and ECG signals. The ECG signals obtained was raw physiological signal, thus features would be extracted from the signals. However, the heart rate captured was already aggregated by Fitbit Sense. The ECG signals (waveform samples) was collected at a sampling rate of 256Hz. Thus a total of 7500 waveform samples was collected every 30 seconds per audio stimuli listened to.From the ground truth obtained the endtime, emotion group and the intensity of the emotion was aligned with the extracted ECG signal, heart rate and timestamp exported from the Fitbit Sense. This was done using the python programming language.At the end of data preprocessing, the ECG signals and heart rate was aligned with discrete emotion (fear, happiness and sadness) and emotion in-

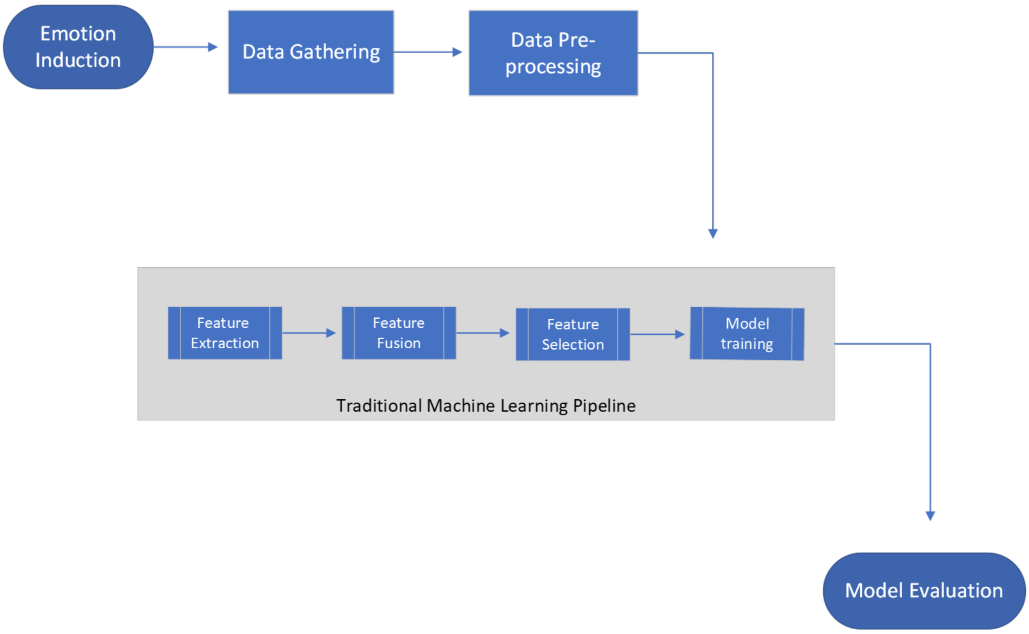
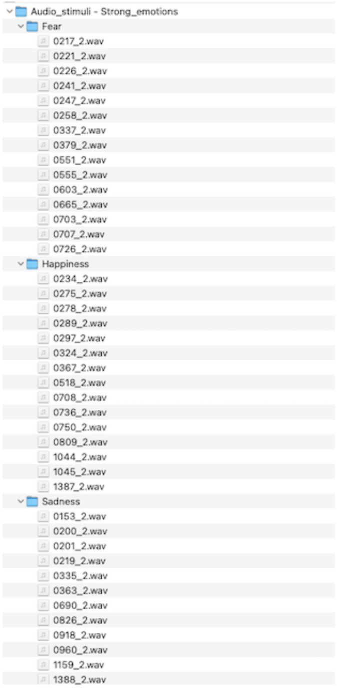


Figure 14: Method Framework for developing emotion recognition model



- Audio_stimuli - Strong_emotions
  - Fear
    - 0217_2.wav
    - 0221_2.wav
    - 0226_2.wav
    - 0241_2.wav
    - 0247_2.wav
    - 0258_2.wav
    - 0337_2.wav
    - 0379_2.wav
    - 0551_2.wav
    - 0555_2.wav
    - 0603_2.wav
    - 0665_2.wav
    - 0703_2.wav
    - 0707_2.wav
    - 0726_2.wav
  - Happiness
    - 0234_2.wav
    - 0275_2.wav
    - 0278_2.wav
    - 0289_2.wav
    - 0297_2.wav
    - 0324_2.wav
    - 0367_2.wav
    - 0518_2.wav
    - 0708_2.wav
    - 0736_2.wav
    - 0750_2.wav
    - 0809_2.wav
    - 1044_2.wav
    - 1045_2.wav
    - 1387_2.wav
  - Sadness
    - 0153_2.wav
    - 0200_2.wav
    - 0201_2.wav
    - 0219_2.wav
    - 0335_2.wav
    - 0363_2.wav
    - 0690_2.wav
    - 0826_2.wav
    - 0918_2.wav
    - 0960_2.wav
    - 1159_2.wav
    - 1388_2.wav

Figure 15: Strong sound stimuli

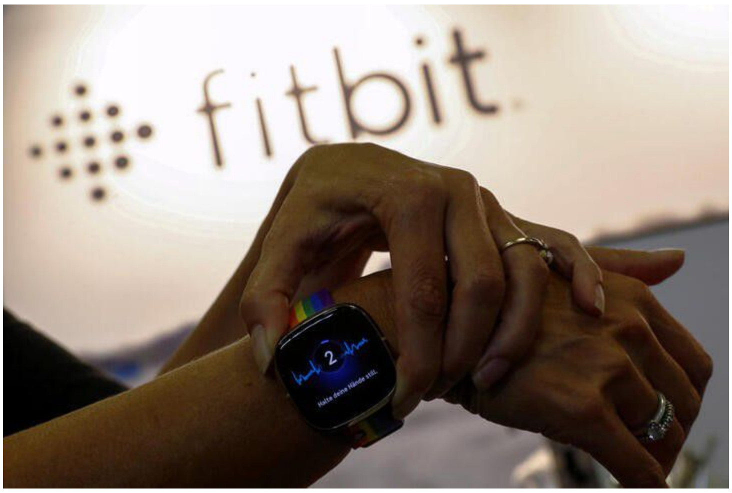


Figure 16: Illustration of the thumb and index finger during signal collection with Fitbit Sense. (Fit, 2022)

tensity (strong or weak) using the timestamp and datetime obtained from fitbit and data annotation form as primary keys.

4.5 Traditional Machine Learning Method

This section explains sequentially the steps taken in developing the emotion recognition model with the pre-processed dataset using traditional machine learning methods. The steps include:

- Feature Extraction
- Feature Selection
- Feature Fusion
- Model Development
- Model Evaluation.

4.5.1 Feature Extraction

In capturing the ECG signals which is non-stationary, noise could have been introduced into the signal (Shu et al., 2018).The noise could be due to motion artefact and baseline wandering (Prasad and Parthasarathy, 2018).Baseline wandering could be due to factors like respiration and body movement during signal capture while motion artefact could be caused by body movement during data collection (Prasad and Parthasarathy, 2018).With this understanding, the noise had to filtered. The Fast Fourier Transform (FFT) was used to denoise the captured ECG waveform samples.FFT is represented by equation 1:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j(\omega)t}dt \tag{1}$$

The Fourier transform is the sum over all time of the signal f(t) multiplied by the complex exponential. The result of this transform are Fourier coefficients .Constituent sinusoidal component of the original signals are generated when the Fourier transform are multiplied by a sinusoid of frequency.After denoising and filtering of the ECG signals using FFT, features were extracted using time series feature extraction Library (TSFEL).TSFEL extracts features from the waveform samples on the statistical, temporal and spectral domains (Barandas et al., 2020).Table 10 presents the list of features TSFEL module could extract from the filtered and denoised waveform samples. Upon execution of TSFEL to extract features from the filtered ECG signals, 390 features were extracted in total on all three domains.

4.5.2 Feature Selection

After extracting features using filtered waveform sample, 390 features were extracted on statistical, temporal and spectral domains. The 390 extracted features could not be used to train with selected classifiers because it would result to:

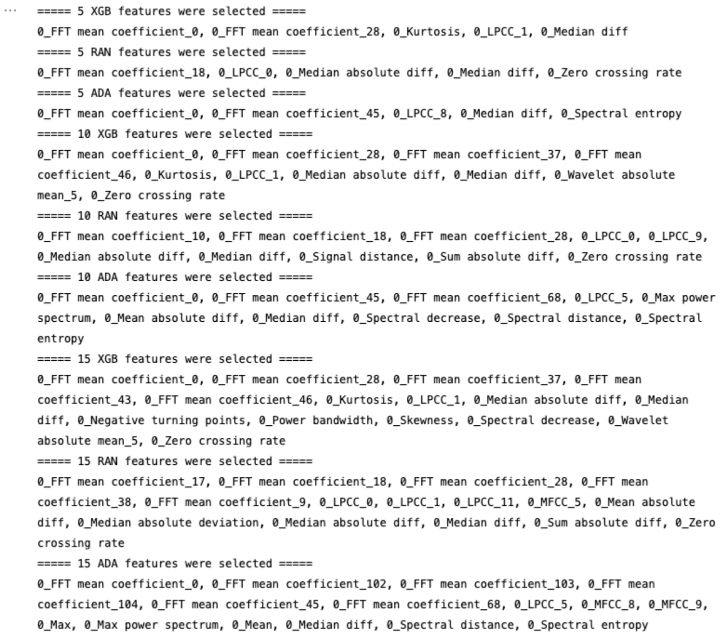
- Developing an emotion recognition with redundant and useless features
- The development of an emotion recognition model that over fits to training data and thus would not generalise well on new or test data

Domain	Feature Names
Statistical	ECDF, ECDF Percentile,ECDF Percentile Count, Histogram,Interquartile range, Kurtosis, Max, Mean,Mean absolute deviation,Median,Median absolute deviation, Min, Root mean square, Skewness,Standard deviation, Variance
Temporal	Absolute energy, Area under the curve, Autocorrelation, Centroid, Entropy, Mean absolute diff, Mean diff,Median absolute diff,Median diff, Negative turning points, Peak to peak distance, Positive turning points, Signal distance, Slope, Sum absolute diff, Total energy, Zero crossing rate, Neighbourhood peaks
Spectral	FFT mean coefficient, Fundamental frequency, Human range energy LPCC, MFCC, Max power spectrum, Maximum frequency, Median frequency, Power bandwidth, Spectral centroid, Spectral decrease, Spectral distance, Spectral entropy, Spectral kurtosis, Spectral positive turning points, Spectral roll-off, Spectral roll-on, Spectral skewness, Spectral slope, Spectral spread, Spectral variation, Wavelet absolute mean, Wavelet energy, Wavelet standard deviation, Wavelet entropy, Wavelet variance

Table 10: Feature list and domain of features extracted with the TSFEL module

- Curse of dimensionality in that we have a small research dataset: 87 observations and 390 features.

As such, this presents the need to select optimum features needed to train the model and guarantee improved model performance on new and test data.Recursive feature elimination (RFE) implemented using the scikit learn framework was utilised to select n numbers of features (Pedregosa et al., 2011). RFE works by using a defined classifier to search from the 390 features n subset of features optimum to train the model. RFE starts with all 390 features and continues to eliminate till n subset of features optimum for model development is selected.For this study n was set to 5, 10 and 15. Thus, 5, 10 and 15 subset of features from 390 features were selected respectively to develop the emotion recognition model.Figure 17 presents the subset of features selected by the ensemble classifiers, where ADA denotes AdaBoost



```
...  ===== 5 XGB features were selected =====
     0_FFT mean coefficient_0, 0_FFT mean coefficient_28, 0_Kurtosis, 0_LPCC_1, 0_Median diff
     ===== 5 RAN features were selected =====
     0_FFT mean coefficient_18, 0_LPCC_0, 0_Median absolute diff, 0_Median diff, 0_Zero crossing rate
     ===== 5 ADA features were selected =====
     0_FFT mean coefficient_0, 0_FFT mean coefficient_45, 0_LPCC_8, 0_Median diff, 0_Spectral entropy
     ===== 10 XGB features were selected =====
     0_FFT mean coefficient_0, 0_FFT mean coefficient_28, 0_FFT mean coefficient_37, 0_FFT mean
     coefficient_46, 0_Kurtosis, 0_LPCC_1, 0_Median absolute diff, 0_Median diff, 0_Wavelet absolute
     mean_5, 0_Zero crossing rate
     ===== 10 RAN features were selected =====
     0_FFT mean coefficient_10, 0_FFT mean coefficient_18, 0_FFT mean coefficient_28, 0_LPCC_0, 0_LPCC_9,
     0_Median absolute diff, 0_Median diff, 0_Signal distance, 0_Sum absolute diff, 0_Zero crossing rate
     ===== 10 ADA features were selected =====
     0_FFT mean coefficient_0, 0_FFT mean coefficient_45, 0_FFT mean coefficient_68, 0_LPCC_5, 0_Max power
     spectrum, 0_Mean absolute diff, 0_Median diff, 0_Spectral decrease, 0_Spectral distance, 0_Spectral
     entropy
     ===== 15 XGB features were selected =====
     0_FFT mean coefficient_0, 0_FFT mean coefficient_28, 0_FFT mean coefficient_37, 0_FFT mean
     coefficient_43, 0_FFT mean coefficient_46, 0_Kurtosis, 0_LPCC_1, 0_Median absolute diff, 0_Median
     diff, 0_Negative turning points, 0_Power bandwidth, 0_Skewness, 0_Spectral decrease, 0_Wavelet
     absolute mean_5, 0_Zero crossing rate
     ===== 15 RAN features were selected =====
     0_FFT mean coefficient_17, 0_FFT mean coefficient_18, 0_FFT mean coefficient_28, 0_FFT mean
     coefficient_38, 0_FFT mean coefficient_9, 0_LPCC_0, 0_LPCC_1, 0_LPCC_11, 0_MFCC_5, 0_Mean absolute
     diff, 0_Median absolute deviation, 0_Median absolute diff, 0_Median diff, 0_Sum absolute diff, 0_Zero
     crossing rate
     ===== 15 ADA features were selected =====
     0_FFT mean coefficient_0, 0_FFT mean coefficient_102, 0_FFT mean coefficient_103, 0_FFT mean
     coefficient_104, 0_FFT mean coefficient_45, 0_FFT mean coefficient_68, 0_LPCC_5, 0_MFCC_8, 0_MFCC_9,
     0_Max, 0_Max power spectrum, 0_Mean, 0_Median diff, 0_Spectral distance, 0_Spectral entropy
```

Figure 17: Subset of Features Selected using Recursive Feature Selection.

classifier, XGB denotes XGBoost classifier, and RAN denotes Random Forest Classifier respectively.

4.5.3 Feature Fusion

The aggregated heart rate obtained alongside the ECG signals during data collection using Fitbit sense was combined with n selected features. This can be referred to as early fusion or feature-level fusion.The result of models developed using early fusion are presented and discussed in section 5.1.3.

4.5.4 Model Training

This section explains how the emotion recognition model was developed. Selected classifiers utilised extracted features to develop the emotion recognition model by ”learning” with these features during model training. This section explains rationale behind developing the emotion recognition model making reference to unsuccessful attempts encountered while developing the model.The choice of classifiers and experimental setup for model training is discussed within this section. Model evaluation techniques used within this thesis are presented in this section.

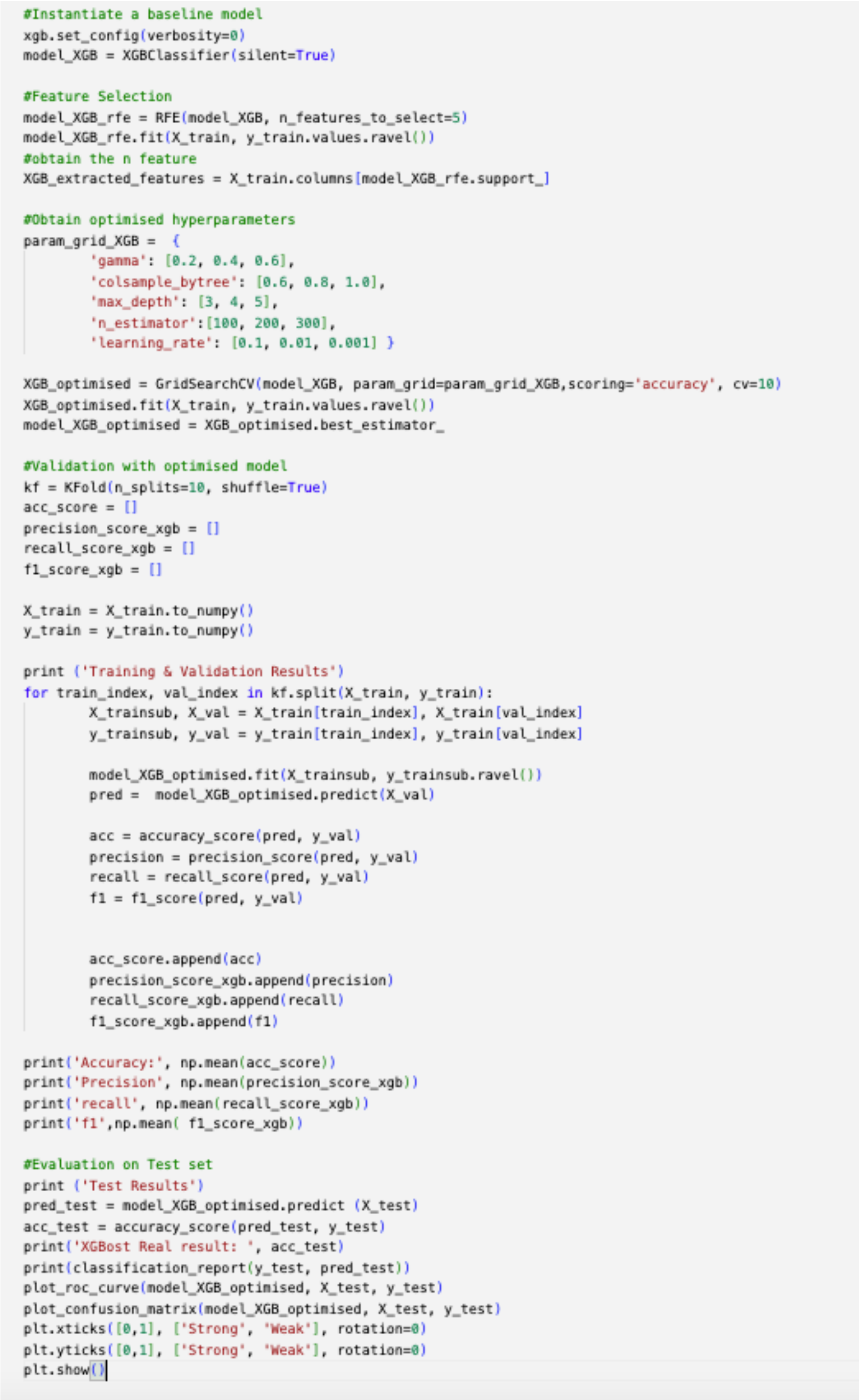
An initial attempt to classify the three classes of discrete emotions (happiness, sadness and fear) using extracted features from ECG signals of emotions reported as strong (5 to 9 on Likert scale) by the participant alone returned an accuracy of 51.4%.This implied the model couldn't differentiate between the three discrete emotions reported as strong using extracted features from the waveform samples. This maybe due to the fact that the ECG signals captured when the participant experienced the strong emotions (happiness, sadness and fear) were indistinguishable for all discrete emotions.Thus, the emotion recognition model developed sort to classify emotions using the intensity of the emotions. Emotions rated from 1 to 4 using were identified as weak emotions. Emotions rated from 5 to 9 were identified as strong emotions.This implies that the model aimed to classify an emotion experienced as strong or weak.This positioned the model training phase as a binary classification task.As explained in section 4.2, a total of 87 observation for both strong and weak emotions were available to be used. Extracted features for all observations were split into training/validation set and test set using an 80%:20% ratio respectively. The training/validation set was used for model training. The test set was used to further evaluate the abilities of the models developed to generalise on new data.

Ensemble classifiers were used to develop the models.Ensemble classifiers use different methods to obtain a better predictive performance than can be obtained from any of the constituent algorithms alone. Ensemble learning involves using techniques like bagging, boosting or stacking to obtain better predictive performance than the use of a lone classifier.Bagging and boosting methods were explored within this thesis project. Bagging involves the use of multiple models on different subsets of the training dataset and combining the prediction from all classifiers to obtain robust predictions. Boosting use weak classifier learned sequentially and adaptively to improve model predictions of a ensemble algorithm.The use of ensemble classifiers within this study was to ensure the development of a robust emotion recognition model. Thus, the selected ensemble classifiers used for model training are:

- Random Forest Classifier: Bagging method
- AdaBoost Classifier: Boosting method
- XGBoost Classifier: Boosting method.

In creating the experimental setup, the dataset was divided into training-validation set and test set using a 80% : 20% ratio respectively.A 10-fold cross-validation was used to evaluate the model with the training-validation set. The test or hold-out set was used to further evaluate the model performance on unseen data. Each of the classifier parameter was optimised using the grid search CV algorithm. This experiment was setup using scikit learn framework (Pedregosa et al., 2011).The setup as executed for XGBoost classifier is presented in figure 18. The same setup was initiated and executed for other ensemble classifiers.From figure, the steps are explained below:

- 1. A baseline instance of the classifier was defined. Recursive feature selection was used to select *n* number for features using the train-validation set. *n* could be 5, 10 or 15.Selected features were stored
- 2. Gridsearch CV algorithm was used to tune the hyperparameters using the train-validation set
- 3. An optimised classifier obtained after hpyerparameter tuning was defined inheriting the best parameter values obtained using the Gridsearch CV algorithm.
- 4. Training-validation set was used to train the optimised instance of the classifier using the selected (n) number of features. A 10-fold cross validation was used to evaluate the accuracy or emotion recognition rate of the training-validation set
- 5. Optimised model obtained after training and validation was further evaluated using defined metrics on the test or holdout set (20%).



```
#Instantiate a baseline model
xgb.set_config(verbosity=0)
model_XGB = XGBClassifier(silent=True)

#Feature Selection
model_XGB_rfe = RFE(model_XGB, n_features_to_select=5)
model_XGB_rfe.fit(X_train, y_train.values.ravel())
#obtain the n feature
XGB_extracted_features = X_train.columns[model_XGB_rfe.support_]

#Obtain optimised hyperparameters
param_grid_XGB =  {
        'gamma': [0.2, 0.4, 0.6],
        'colsample_bytree': [0.6, 0.8, 1.0],
        'max_depth': [3, 4, 5],
        'n_estimator':[100, 200, 300],
        'learning_rate': [0.1, 0.01, 0.001] }

XGB_optimised = GridSearchCV(model_XGB, param_grid=param_grid_XGB,scoring='accuracy', cv=10)
XGB_optimised.fit(X_train, y_train.values.ravel())
model_XGB_optimised = XGB_optimised.best_estimator_

#Validation with optimised model
kf = KFold(n_splits=10, shuffle=True)
acc_score = []
precision_score_xgb = []
recall_score_xgb = []
f1_score_xgb = []

X_train = X_train.to_numpy()
y_train = y_train.to_numpy()

print ('Training & Validation Results')
for train_index, val_index in kf.split(X_train, y_train):
        X_trainsub, X_val = X_train[train_index], X_train[val_index]
        y_trainsub, y_val = y_train[train_index], y_train[val_index]

        model_XGB_optimised.fit(X_trainsub, y_trainsub.ravel())
        pred =  model_XGB_optimised.predict(X_val)

        acc = accuracy_score(pred, y_val)
        precision = precision_score(pred, y_val)
        recall = recall_score(pred, y_val)
        f1 = f1_score(pred, y_val)


        acc_score.append(acc)
        precision_score_xgb.append(precision)
        recall_score_xgb.append(recall)
        f1_score_xgb.append(f1)

print('Accuracy:', np.mean(acc_score))
print('Precision', np.mean(precision_score_xgb))
print('recall', np.mean(recall_score_xgb))
print('f1',np.mean( f1_score_xgb))

#Evaluation on Test set
print ('Test Results')
pred_test = model_XGB_optimised.predict (X_test)
acc_test = accuracy_score(pred_test, y_test)
print('XGBost Real result: ', acc_test)
print(classification_report(y_test, pred_test))
plot_roc_curve(model_XGB_optimised, X_test, y_test)
plot_confusion_matrix(model_XGB_optimised, X_test, y_test)
plt.xticks([0,1], ['Strong', 'Weak'], rotation=0)
plt.yticks([0,1], ['Strong', 'Weak'], rotation=0)
plt.show()
```

Figure 18: Experiment setup for development of emotion recognition model.

4.6 Model Evaluation

As stated in section 4.5.4, the development of the emotion recognition model was positioned as a binary classification task that utilised extracted features from waveform samples to classify emotions as weak or strong.The model performances was evaluated using accuracy, recall, precision and F1 score.

$$Accuracy = (1 - \frac{|Ncco - Nto|}{Nti}) * 100 \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{5}$$

Accuracy as shown in equation 2 calculates the percentage of the total number of correctly identified observation. Using equation 2 *Ncco* represents the number correctly identified observation and *Nto* the number of total of observation utilised by the classifier.Precision calculates the number of observations correctly predicted and is defined as equation 4. Recall on the other hand calculates the number of observations that were actual correct observations that the model correctly identified.Recall is defined as equation 3. F1 is the harmonic mean of precision and recall defined as equation 5. TP, FP, TN, FN denotes true positive, false positive, true negative, false negative respectively.Additionally, confusion matrices were used to visualise evaluating the performance of the model on test set.

4.7 Summary

In this chapter, the emotion induction stimuli and data collection protocol used in gathering physiological signals for developing the emotion recognition model have been discussed. Data preprocessing method have been also been discussed.This chapter also presented an in-depth explanation into feature extraction, feature selection and feature fusion type used within the study.Classifier choice used for model training have been discussed. Chapter 5 presents the model performance using evaluation metrics defined in section 4.6.

Chapter Five

5 Results and Discussion

5.1 Results

This chapter presents model performances using selected features, and fused features obtained after early fusion.This chapter also presents the emotion recognition model and framework. An explanation of model predictions is discussed and thesis findings are discussed. Finally the application of the emotion recognition model within social care setting is discussed.This section presents the results of emotion recognition models developed using the ensemble classifiers from section 4.5.4 and presents a comparison of models using the defined evaluations metrics on training-validation set, test set and with fused features.

5.1.1 Evaluating Feature Selection Subsets

As discussed in section 4.5.2, 390 features were derived from the filtered waveform samples on the statistical, frequency and spectral domains. The first task was identifying the optimum number of features and feature types needed to build an emotion recognition model.Table 11 presents the result of selecting 5, 10, and 15 features using recursive feature elimination on both training-validation set and test set. As observed, for models developed with random forest classifier and xgboost, the accuracy or emotion recognition rate declined with an increase in the number of features. Table 12 displays the emotion recognition rates of models developed with classifiers using 5, 10, 15 selected features. As expected, the emotion recognition rate of model developed with AdaBoost classifier remained constant regardless of the number of features used.For model developed with random forest classifier, the emotion recognition rate declined with an increase in the number of features while the model developed with XGBoost declined when 15 features were used.

Feature Subset	Random Forest(%)	AdaBoost (%)	XGboost(%)
5 features	64.76	65.71	86.90
10 features	66.66	65.71	85.23
15 features	65.71	65.71	85.23

Table 11: Emotion recognition rate using n features on training-validation set

Feature Subset	Random Forest(%)	AdaBoost (%)	XGboost(%)
5 features	72.22	88.88	89.00
10 features	66.67	88.88	94.44
15 features	61.11	88.88	88.88

Table 12: Emotion recognition rate using n features on test set

Emotion recognition rates observed could be associated with predictive abilities of the features. RFE works by selecting the best features that outputs the best emotion recognition rate. The more the number of features being added, the lesser the predictive abilities of the additional features. In this context, the first 5 features are very important in achieving a high emotion recognition rate, an increase to 10 features means 5 additional features with lesser predictive abilities have been added, thus reducing the performance of the model especially on the training-validation set.In selecting the optimum number of features needed to develop the emotion recognition model,the abilities of the models developed must be scrutinized on the test set.Except for the model developed with random forest classifier, the accuracy remained constant or increased when 10 features were used. This indicated that developing the model with 10 features is optimum for this participant.

5.1.2 Model Evaluation Comparison

In section 5.1.1, it was established that the optimum features to be used in developing the emotion recognition model was 10. This section presents the overall model performance of models developed with all classifiers using metrics defined in 4.6.Figure 20 visualises the performance of models developed on training and validation set. As observed, the best performing model is the model developed with XGBoost classifier using metrics precision, recall and f1 score. Figure 20 presents the performances of the three models on test set.As anticipated, the best performing model was the model developed with XGBoost classifier.The excellent performance of XGBoost classifier also extreme gradient boosting could be associated to the fact that with XGBoost, trees are built in parallel and the algorithm follows a level-wise strategy, scanning across gradient values and using these partial sums to evaluate the quality of splits during training.

The predictions of the models was visualised to provide more context into how well the models performed in predicting emotions using the test data with confusion matrices. Figure 29, 22, 23 displays the confusion matrices of models developed using XGBoost classifier, random forest classifier, adaboost classifier respectively. As seen, confusion matrix gave the best prediction misclassifying one observation as a strong emotion instead of a weak emotion.Model developed with adaboost was second best misclassifying two emotions while the model developed using random forest classifier was the least accurate misclassifying six emotions.

The abilities of the models to generalise on test data as was compared to understand if the developed emotion recognition model overfit or underfit on training data.Thus, the accuracy obtained with training-validation set and test set was compared for the three emotion recognition models developed. Figure 24 displays the comparison using a bar chart.As observed, the three models developed using 10 features did not over fit to training-validation set.It is noticed that the model performed better on the test set for models developed using XGBoost classifier and AdaBoost classifier. This could have been due to the small size of the test set (N=18). Nevertheless, all models showed ability to generalise well on unseen data.

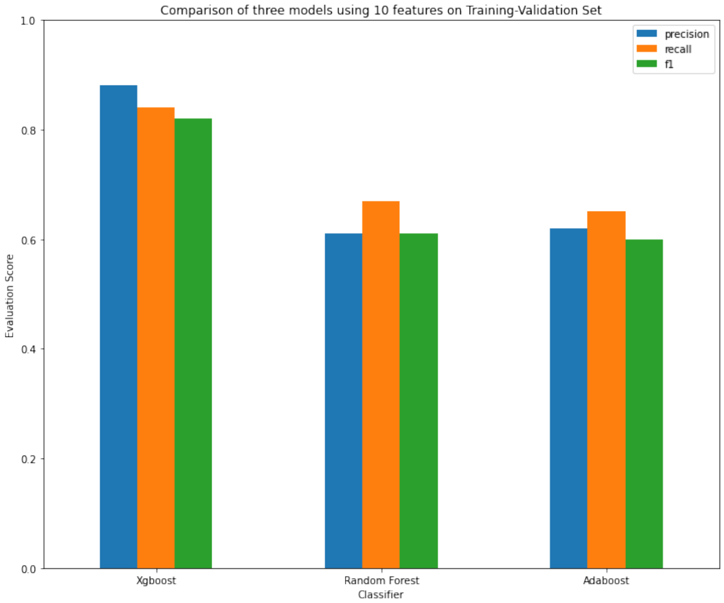


Figure 19: Model Evaluation on train-validation set using 10 features

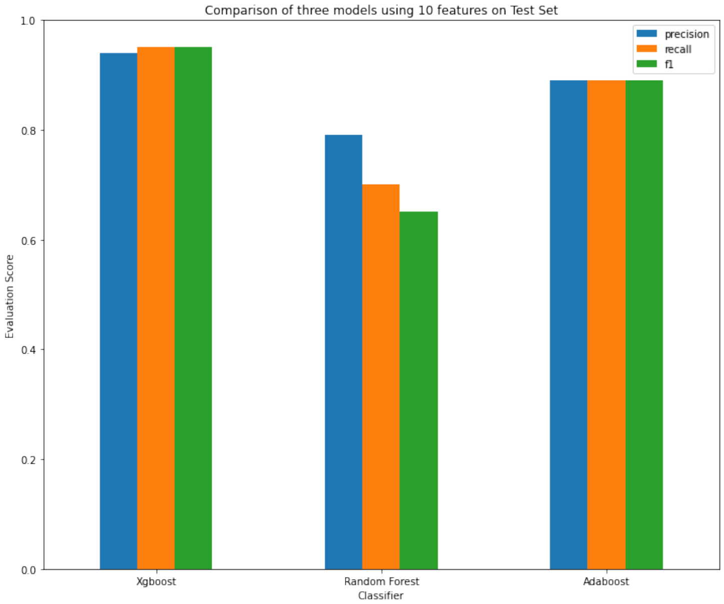


Figure 20: Model Evaluation on test set using 10 features

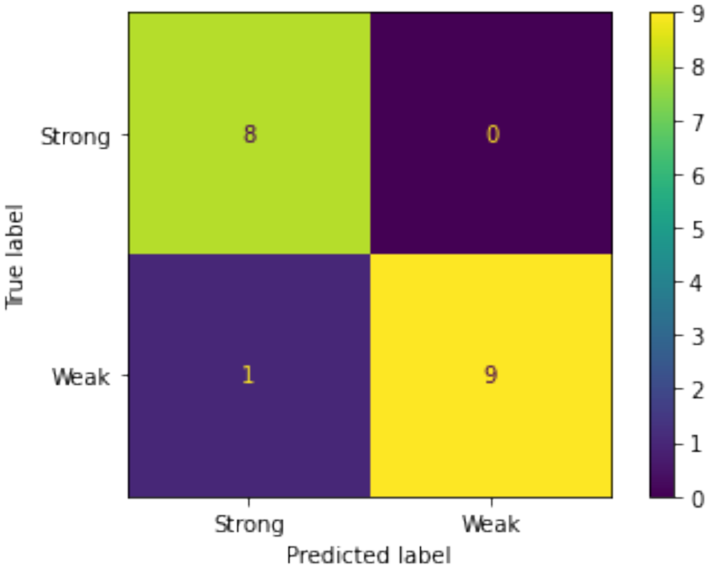


Figure 21: Confusion Matrix of Model developed with Xgboost classifier using test set

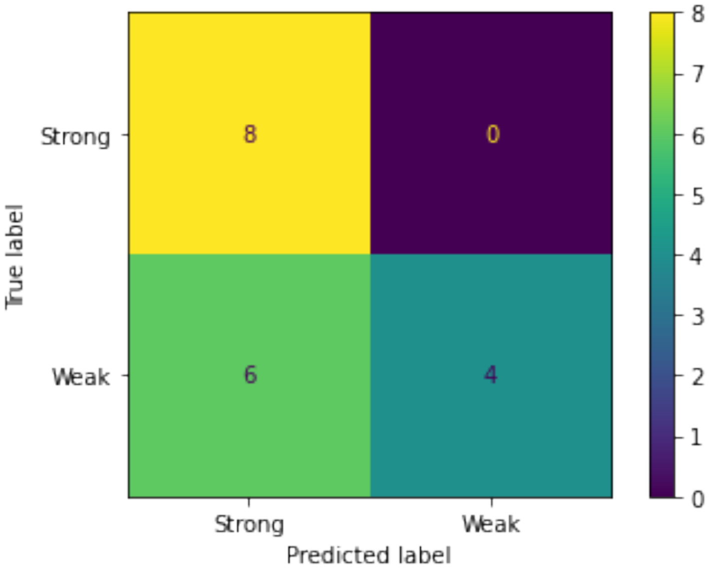


Figure 22: Confusion Matrix of Model developed with Random Forest classifier using test set

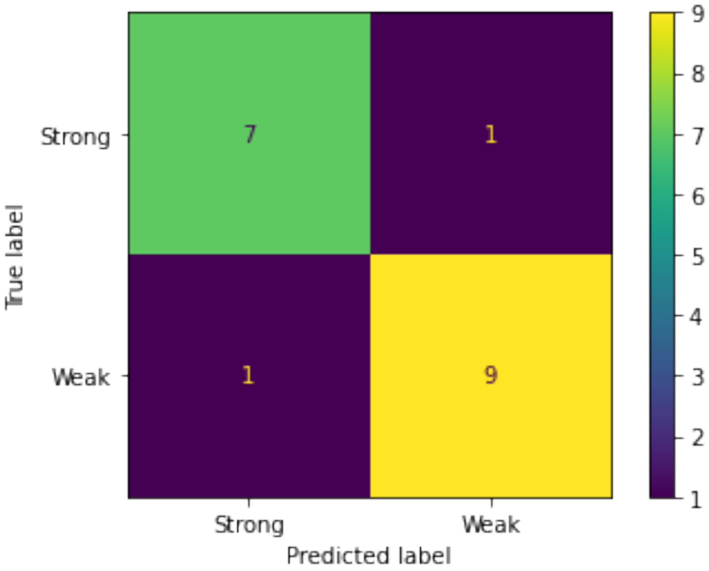


Figure 23: Confusion Matrix of Model developed with AdaBoost classifier using test set

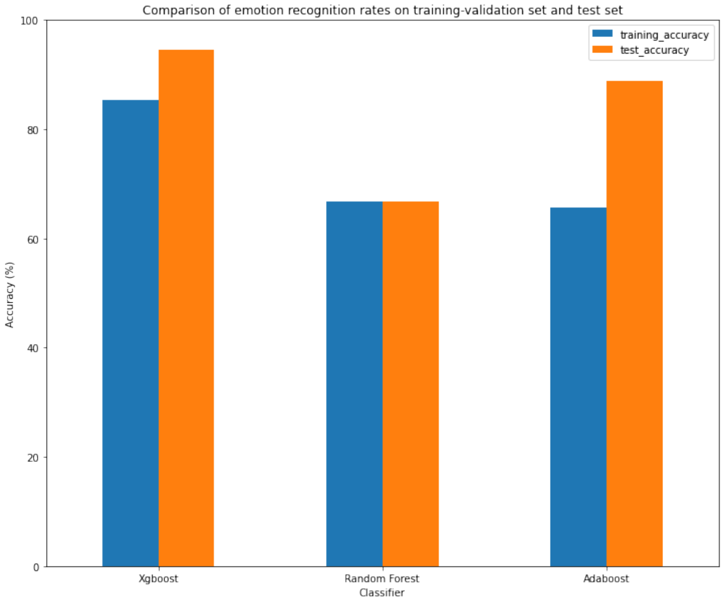


Figure 24: Comparison of emotion recognition rate on train-validation set and test set

5.1.3 Model Evaluation Comparison using early level fusion

As discussed in section 4.5.3, early level fusion was investigated to develop a robust emotion recognition model. In executing early level fusion, the aggregated heart rate obtained directly from Fitbit sense was fused with 10 features selected by each classifier. The fused feature was used to develop emotion recognition models. Emotion recognition model was evaluated using training/validation set and test set. The models developed after early level fusion were evaluated using defined metrics in 4.6. Figure 25 shows the performance of the emotion recognition models using precision, recall and f1 score. When compared to performance discussed in section 5.1.2 a marked improvement in the ability of the models to classify emotions was noticed after executing early level fusion using training-validation set.Similarly, the performance of the models on test set shown in figure 26 reflected an increase in the performance of model developed with random forest classifier when compared to performance explained in section 5.1.2. There was a noticeable drop in performance of model developed with AdaBoost classifier especially in recall and f1 score.However, overall using both training and test set remarkable increase in performance of emotion recognition models was noticed after early level fusion was executed.

The improvement in performance of the emotion recognition models developed with random forest becomes visible when the confusion matrix is considered.Compared to matrix discussed in section 5.1.2, the emotion recognition model developed using random forest classifier had 1 emotion misclassifed after executing early level fusion, the model developed with XGBoost classifier had no changes on the confusion matrix while the model developed with AdaBoost had 4 emotion misclassified.Considering the accuracy of the emotion recognition models developed using fused features obtained after early fusion using results displayed in table 13. The models improved in accuracy when compared to performance obtained using only features extracted from the ECG signals.This is a clear indication that the fusion of physiological signals (features from waveform samples and aggregated heart rate from fitbit sense) resulted in the development of a robust emotion recognition model.

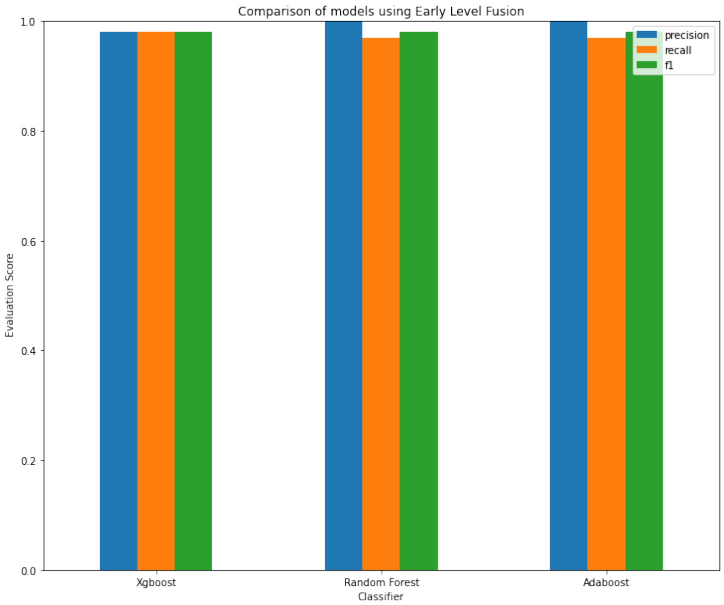


Figure 25: Model evaluation after executing early level fusion on training-validation set

Data	Random Forest(%)	AdaBoost (%)	XGboost(%)
Training set	98.57	98.57	97.14
Test set	94.44	77.78	94.44

Table 13: Emotion recognition rate on training-validation set and test set after executing early level fusion

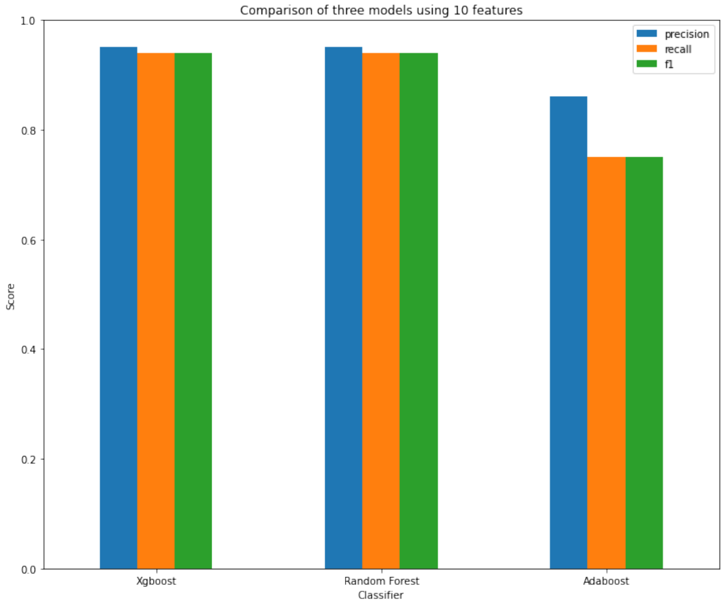


Figure 26: Model evaluation after executing early level fusion on training-validation set

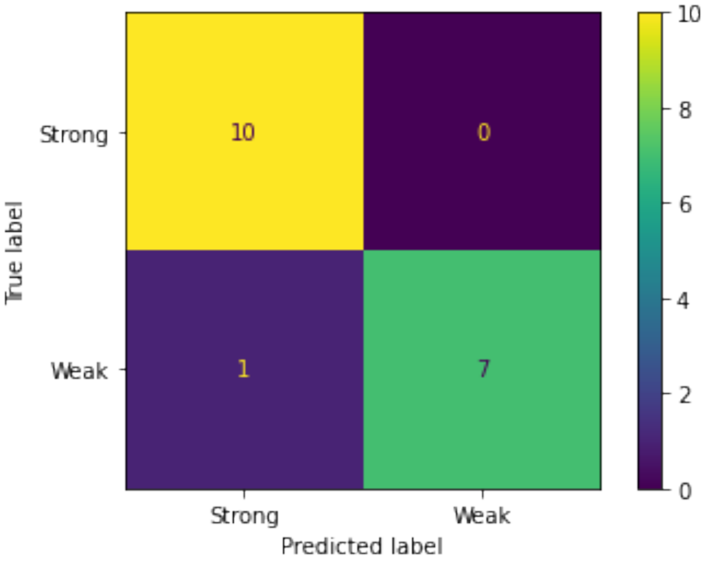


Figure 27: Confusion matrix of model developed with Xgboost classifier after early level fusion

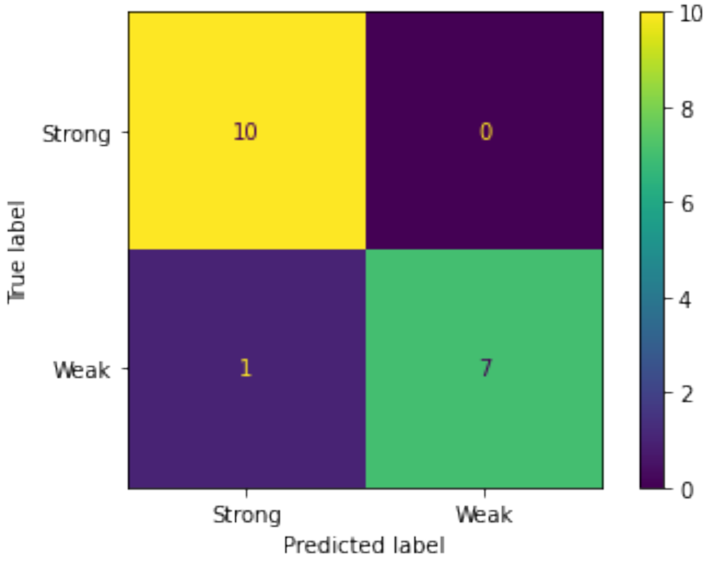


Figure 28: Confusion matrix of model developed with random classifier after early level fusion

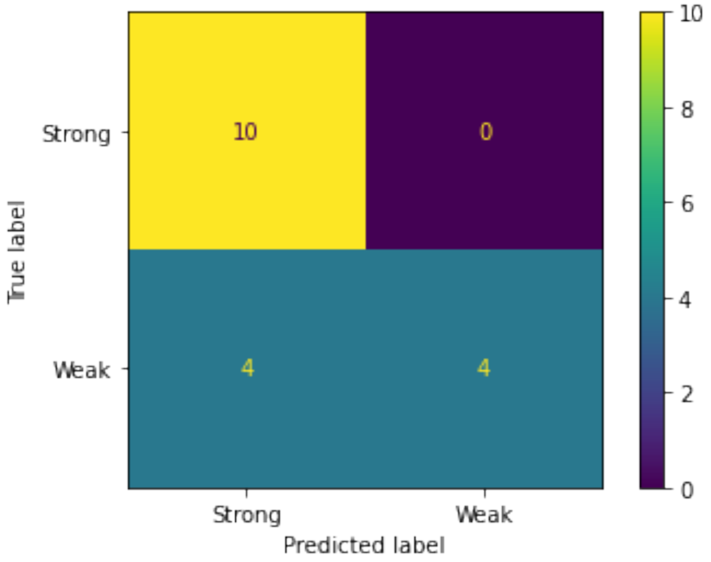


Figure 29: Confusion Matrix of Model developed with AdaBoost classifier after early level fusion

5.2 Discussion

This section presents the emotion recognition framework used in developing the emotion recognition model and explains possible applications of the framework. This section discusses and explains how the features drive the predictions made by the model on test data.Further a holistic discussion of findings made in this thesis is presented.Finally the application of emotion recognition model in social care setting is discussed.

5.2.1 Emotion Recognition Model Framework

Figure 30 presents the emotion recognition framework. The pipeline can be replicated and scaled to develop a personalised or subject dependent emotion recognition model for residents within nursing facilities. At the end of this pipeline, an emotion recognition model for residents would be developed using ECG waveform samples and aggregated heart rate that can classify and predict occurrence of an emotion as the intensity of emotion experienced by the resident.Within this thesis, the same pipeline was used in developing the emotion recognition model.The model developed with XGboost classifier using this pipeline achieved an accuracy of 85.23% on training-validation set and 94.44% on test set.

5.2.2 Explaining Predictions of Emotion Recognition Model

In developing the best performing emotion recognition model, the hyperparameters were tuned to obtain the best emotion recognition rate. These hyperparameters are presented in table 14. Further,the predictions obtained from the emotion recognition model developed after executing the emotion recognition framework can be explained by understanding how the features drive the decision and prediction made by the model. Thus, the features and description of these features for the best performing model are presented in table 15.SHAP module implemented in python programming language was utilised to give insight into how the features drive the predictions of the emotion recognition model. SHAP (Shapley Additive Explanations) module

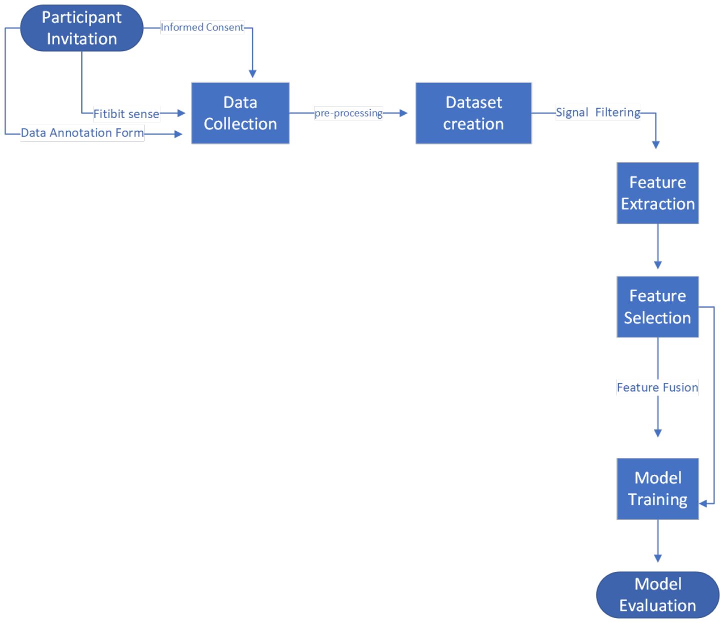


Figure 30: Emotion Recognition Framework

interprets a model using shapely values.Shapely values are a measure of contribution each feature has a in machine learning model.Figure 31 presents a summary plot obtained after implementing SHAP module on the created. Figure 31 presents the features in a descending order.For the emotion recognition model developed using physiological obtained from the research participant, the median differences of the signals contributed the most to deciding if an emotion is strong or weak while the zero crossing rate feature contributed next in order of importance to deciding if an emotion is weak or strong. Overall for the emotion recognition model, the feature of highest importance in predicting if an emotional stimuli experienced is strong or weak is the Median differences while the least feature is 0_LPPCC_1 which is the linear prediction of cepstral coefficients.

Hyperparameter	Values
colsample_bytree	0.6
gamma	0.2
learning_rate	0.01
max_depth	3
n_estimator	100

Table 14: Values of hyperparameter used in developing emotion recognition model

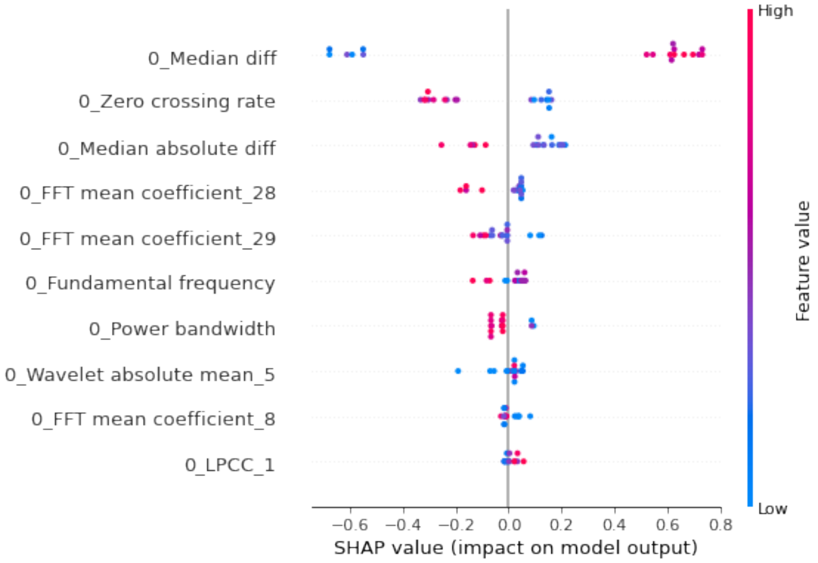


Figure 31: SHAP summary plot using the 10 selected Features

Feature Name	Description
0_FFT mean coefficient_28	The mean value of each spectrogram frequency.
0_FFT mean coefficient_29	The mean value of each spectrogram frequency.
0_FFT mean coefficient_8	The mean value of each spectrogram frequency.
0_Fundamental frequency	The fundamental frequency of the signal.
0_LPCC_1	The linear prediction cepstral coefficients.
0_Median absolute diff	The median absolute differences of the signal.
0_Median diff	The median of differences of the signal.
0_Power bandwidth	The power spectrum density bandwidth of the signal.
0_Wavelet absolute mean_5	The absolute mean value of each wavelet scale.
0_Zero crossing rate	The Zerocrossing rate of the signal.

Table 15: Features used in developing the emotion recognition model

5.2.3 Thesis Findings

After testing hypotheses listed in section 3.5 using non parametric statistical methods. For all participants, there were significant differences in physiological signals captured when distinct audio stimuli were received by participants.This indicates upon humans experience, physiological signals changes after receiving stimuli by humans,and thus validates the use of physiological signals in developing an emotion recognition model which has been executed by researchers within the field Ayata et al. (2020); Dzedzickis et al. (2020); Koelstra et al. (2012); Cosoli et al. (2021); Shu et al. (2018).

In developing the emotion recognition model, three ensemble algorithms were utilised. A 10-fold cross-validation was executed during training and validation.Models were evaluated using training/validation set and test set.Using accuracy as performance metric, the emotion recognition rate ranged from 65.71% to 85.23% on training/validation set. Emotion recognition rate on test set ranged from 66.67% to 94.44%. When compared to a similar study that aimed to identify the elicitation of acoustic stimulation using a wrist worn empatica device that captured electrodermal activity and skin temperature and achieved an accuracy of 75.00% and 72.62% for support vector machine and linear regression respectively, this study achieved a higher emotion recognition rate on test data of 94.44% using ECG signal only and a subject dependent approach.The higher results in accuracy on test data obtained within this study could be associated with evaluation on relatively small dataset (N=18). Further this study utilised bagging and boosting methods.The best performing model that achieved an emotion recognition rate (accuracy) of 94.44% was developed using the boosting method.Boosting technique uses weak classifier learned sequentially and adaptively to improve model prediction.Thus the boosting technique used could be associated to the improved performance obtained when compared to the results presented by Cosoli et al. (2021) obtained using support vector machine and linear regression classifiers.

As discussed and presented in section 5.1.3, the use 10 extracted features from ECG waveform sample and selected using recursive feature elimination combined with the aggregated heart rate provided by Fitbit Sense device improved the emotion recognition rate for model developed using random forest classifier. This suggests that the use of multimodal physiological signals obtained from Fitbit sense device can lead to the development of a robust emotion recognition model. This is line with findings presented by Cosoli et al. (2021). The use of mulitmodal physiological signals (IBI, BVP, EDA, SKT) within this study in developing an emotion recognition model using support vector machine (72.62%) was better compared to the model was developed using HRV signals only (66.67%).

Overall, this study established that extracted features from ECG signals only or fused with aggregated heart rate obtained from Fitbit Sense (wrist-worn device) can be used in developing an emotion recognition model. As such, this study confirms that the use of a subject dependent approach in developing an emotion recognition model using physiological signals obtained from smart device embedded with sensor technology can identify the degree of emotional intensity (weak or strong) experienced by residents of a nursing facilities. The application of the emotion recognition model in social care is discussed in the next section.

5.2.4 Application in social care

In social care setting, the emotion recognition model developed using multimodal physiological signal obtained from a Fitbit sense device can classify intensity of emotion stimuli as weak or strong. This model makes it possible to identify moments or periods when weak or strong stimuli are experienced by residents of a nursing facilities.This is useful because when a resident receives a strong stimuli, a discrete emotion would be experienced by the resident(Plutchik, 2001). A correct classification and prediction of the intensity of the emotion using physiological signals obtained from a wrist-worn smart device can be achieved by the emotion recognition model. Correctly classifying periods when resident experienced strong emotions enables health professionals to investigate the cause of the emotion and use findings to improve the health plan of the resident. For example, if a resident experienced a strong emotion like anger, the emotion recognition model can classify and identify that a strong emotion was experienced. This can be investigated by health professionals and the stimuli of the emotion can be eliminated to improve the overall health of the patient

Thus, this emotion recognition model is an important contribution to social care. The model can aid classification of strong and weak emotions using physiological signals obtained from a wrist-worn smart device.This can be leveraged to ensure residents receive improved care.

Chapter Six

6 Conclusion

This thesis project aimed to investigate the use of multimodal physiological signal obtained from wrist-worn smart device in developing an emotion recognition model.This thesis project presented the emotion recognition model developed and an emotion recognition framework. In greater details, the following steps were taken to achieve the aim of this thesis:

- Thorough review of related works within the emotion recognition research field
- Creation of data annotation form to collect ground truths
- Collection of physiological signals from research participants using discussed data collection protocol
- Creation of data set by aligning ground truth and ECG signals using python programming language.
- Filtering ECG waveform samples and extraction of features from the ECG waveform samples using TSFEL package with python programming language.
- Utilising three ensemble classifier for model training and development using features selected by recursive feature elimination.
- Development and validation of emotion recognition models.
- Presentation of the emotion recognition model and framework.

The execution of the listed steps implied that all the objective set out in section 1.2 was achieved.

6.1 Further Works

The challenges encountered within this project was the choice of smart device. The Fitbit sense smart device is a fitness (health) smart device. It presents most of the capture data in aggregated and processed format and doesn't present physiological signals in the raw form except ECG signals. Thus filtering, denoising, and feature extraction could not be executed with these aggregated data exported from the Fitbit Sense device.Further,the sample size worked with in developing the emotion recognition model was small (N=87).Therefore some further works can be done to develop and improve the emotion recognition model.

- First is the use of a different wrist-worn smart device that collects raw physiological signals simultaneously automatically. This would enable the use of multiple multimodal raw physiological signals in developing a robust emotion recognition model
- Second is an improvement of data gathering to collect enough observations of physiological signals and utilise in developing a precise emotion recognition model
- Third, this thesis project focused on the use of traditional machine learning classifier in creating the emotion recognition model, however further works could look into using deep learning architectures in creating emotion recognition model using obtained physiological signals
- Fourth, a personalised or subject-dependent approach was taken to develop the emotion recognition model. Further works could look to work with a larger research population and develop a generalised emotion recognition model that is scalable and applicable within the health sector.

References

(2022). Fitbit sense user manual. fitbit user manual. https://help.fitbit.com/manuals/manual$_sense_en_US.pdf$.

Ayata, D., Yaslan, Y., Mustafa, , and Kamasak, E. (2020). Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. *Journal of Medical and Biological Engineering*, 40:149–157.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). Tsfel: Time series feature extraction library. *SoftwareX*, 11:100456.

Benezeth, Y., Li, P., Macwan, R., Nakamura, K., Gomez, R., and Yang, F. (2018). Remote heart rate variability for emotional state monitoring. pages 153–156. IEEE.

Bradley, M. M. and Lang, P. J. (2017). *International Affective Picture System*. Springer International Publishing.

Cheng, Z., Shu, L., Xie, J., and Chen, C. L. P. (2017). A novel ecg-based real-time detection method of negative emotions in wearable applications. pages 296–301.

Ciman, M. and Wac, K. (2018). Individuals' stress assessment using human-smartphone interaction analysis. *IEEE Transactions on Affective Computing*, 9:51–65.

Cosoli, G., Poli, A., Scalise, L., and Spinsante, S. (2021). Measurement of multimodal physiological signals for stimulation detection by wearable devices. *Measurement*, 184:109966.

Cosoli, G., Spinsante, S., and Scalise, L. (2020). Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics. *Measurement*, 159:107789.

Dzedzickis, A., Kaklauskas, A., and Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. *Sensors*, 20:592.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6:169–200.

Exposito, M., Hernandez, J., and Picard, R. W. (2018). Affective keys. pages 139–145. ACM.

Feidakis, M., Daradoumis, T., and Caballe, S. (2011). Endowing e-learning systems with emotion awareness. pages 68–75. IEEE.

Goshvarpour, A., Abbasi, A., and Goshvarpour, A. (2017). An emotion recognition approach based on wavelet transform and second-order difference plot of ecg. *Journal of AI and Data Mining*, 5:211–221.

Gouizi, K., Reguig, F. B., and Maaoui, C. (2011). *Analysis Physiological signals for Emotion Recognition*.

Holst, A. and Skär, L. (2017). Formal caregivers' experiences of aggressive behaviour in older people living with dementia in nursing homes: A systematic review. *International Journal of Older People Nursing*, 12:e12158.

Hsieh, P.-Y. and Chin, C.-L. (2011). The emotion recognition system with heart rate variability and facial image features. pages 1933–1940. IEEE.

K, Z., Zheng1, E, J., Staubitz2, Weitlauf2, A. S., Staubitz, J., Pollack4, M., Shibley, L., Hopton2, M., Martin, W., Swanson, A., Juárez, P., Warren, Z. E., and Sarkar, N. (2021). A predictive multimodal framework to alert caregivers of problem behaviors for children with asd (premac) a predictive multimodal framework to alert caregivers of problem behaviors for children with asd. 21:370.

Kim, K. H., Bang, S. W., and Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical Biological Engineering Computing*, 42:419–427.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). Deap: A database for emotion analysis ;using physiological signals. *IEEE Transactions on Affective Computing*, 3:18–31.

Kołakowska, Szwoch, and Szwoch (2020). A review of emotion recognition methods based on data acquired via smartphone sensors. *Sensors*, 20:6367.

Lan, Z., Sourina, O., Wang, L., and Liu, Y. (2016). Real-time eeg-based emotion monitoring using stable features. *The Visual Computer*, 32:347–358.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50:372–385.

Lee, M. S., Lee, Y. K., Pae, D. S., Lim, M. T., Kim, D. W., and Kang, T. K. (2019). Fast emotion recognition based on single pulse ppg signal with convolutional neural network.

Li, H. C., Pan, T., Lee, M. H., and Chiu, H. W. (2021). Make patient consultation warmer: A clinical application for speech emotion recognition. *Applied Sciences (Switzerland)*, 11.

Lin, K. C., Huang, T. C., Hung, J. C., Yen, N. Y., and ju Chen, S. (2013). Facial emotion recognition towards affective computingbased learning. *Library Hi Tech*, 31:294–307.

Lin, S.-Y., Jao, C.-W., Wang, P.-S., Liou, M., Wu, J.-L., Chun, H., Tseng, C.-T., and Wu, Y.-T. (2021). Differences in physiological signals due to age and exercise habits of subjects during cycling exercise.

Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., and Chen, J.-H. (2010). Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57:1798–1806.

Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K., and Chen, J.-H. (2009). *EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine; EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine.*

Liu, J., Meng, H., Nandi, A., and Li, M. (2016). Emotion detection from eeg recordings. pages 1722–1727. IEEE.

Lundh, F. (1999). An introduction to tkinter. *URL: www. pythonware. com/library/tkinter/introduction/index. htm.*

Mehrabian, A. (1997). Comparison of the pad and panas as models for describing emotions and for differentiating anxiety from depression. *Journal of Psychopathology and Behavioral Assessment*, 19:331–357.

Mottelson, A. and Hornbæk, K. (2016). An affect detection technique using mobile commodity sensors in the wild. pages 781–792. ACM.

Nasoz, F., Alvarez, K., Lisetti, C. L., and Finkelstein, N. (2004). Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology Work*, 6:4–14.

Nie, D., Wang, X.-W., Shi, L.-C., and Lu, B.-L. (2011). Eeg-based emotion recognition during watching movies. pages 667–670. IEEE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89:344–350.

Prasad, B. V. P. and Parthasarathy, V. (2018). Detection and classification of cardiovascular abnormalities using fft based multi-objective genetic algorithm. *Biotechnology Biotechnological Equipment*, 32:183–193.

Shany, T., Redmond, S. J., Narayanan, M. R., and Lovell, N. H. (2012). Sensors-based wearable systems for monitoring of human movement and falls. *IEEE SENSORS JOURNAL*, 12.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., and Yang, X. (2018). A review of emotion recognition using physiological signals. *Sensors*, 18:2074.

Tarnowski, P., Kołodziej, M., Majkowski, A., and Rak, R. J. (2017). Emotion recognition using facial expressions. *Procedia Computer Science*, 108:1175–1184.

Tikadar, S. and Bhattacharya, S. (2019). *A Novel Method to Build and Validate an Affective State Prediction Model from Touch-Typing.*

Tikadar, S., Kazipeta, S., Ganji, C., and Bhattacharya, S. (2017). *A Minimalist Approach for Identifying Affective States for Mobile Interaction Design.*

Valenza, G., Lanata, A., and Scilingo, E. P. (2012). The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE Transactions on Affective Computing*, 3:237–249.

Verma, G. K. and Tiwary, U. S. (2014). Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage*, 102:162–172.

Wong, W. M. and Tan, A. W. (2010). Pso optimization of synergetic neural classifier for multichannel emotion recognition. *2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC)*, pages 15–17.

Wu, G., Liu, G., and Hao, M. (2010). The analysis of emotion recognition from gsr based on pso. pages 360–363. IEEE.

Yang, W., Makita, K., Nakao, T., Kanayama, N., Machizawa, M. G., Sasaoka, T., Sugata, A., Kobayashi, R., Hiramoto, R., Yamawaki, S., Iwanaga, M., and Miyatani, M. (2018). Affective auditory stimulus database: An expanded version of the international affective digitized sounds (iads-e). *Behavior Research Methods*, 50:1415–1429.

Zhang, W., Xiangminxu, L. S., and Liao, D. (2017). Affective virtual reality system (avrs): Design and ratings of affective vr scenes; affective virtual reality system (avrs): Design and ratings of affective vr scenes.

Zhang, Y.-D., Yang, Z.-J., Lu, H.-M., Zhou, X.-X., Phillips, P., Liu, Q.-M., and Wang, S.-H. (2016). Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4:8375–8385.

Zong, C. and Chetouani, M. (2009). *Hilbert Huang transform based physiological signals analysis for emotion recognition; Hilbert-Huang transform based physiological signals analysis for emotion recognition.*