AN INVESTIGATION OF FRONTAL ALPHA ASYMMETRY AND DELTA–BETA CORRELATION DURING COGNITIVE IMPAIRMENT IN INDIVIDUALS WITH PUBLIC SPEAKING ANXIETY

BEH SOON TATT

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

AN INVESTIGATION OF FRONTAL ALPHA ASYMMETRY AND DELTA–BETA CORRELATION DURING COGNITIVE IMPAIRMENT IN INDIVIDUALS WITH PUBLIC SPEAKING ANXIETY

BEH SOON TATT

This report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Computer Engineering with Honours

Faculty of Electronic and Computer Technology and Engineering Universiti Teknikal Malaysia Melaka

2023/2024

UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : AN INVESTIGATION OF FRONTAL ALPHA ASYMMETRY AND DELTA–BETA CORRELATION DURING COGNITIVE IMPAIRMENT IN INDIVIDUALS WITH PUBLIC SPEAKING ANXIETY

Sesi Pengajian : 2023/2024

Saya B̲ mengaku membenarkan laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (✓):

☐ **SULIT*** (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

☐ **TERHAD*** (Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan.

☑ **TIDAK TERHAD**

Disahkan oleh:

Soon

DR. FARAH SHAHNAZ BINTI FEROZ
Pensyarah Kanan
Fakulti Kejuruteaan Elektronik Dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)

_____   _____
(TANDATANGAN PENULIS)   (COP DAN TANDATANGAN PENYELIA)

Alamat Tetap: 15,Jalan Utama 53,Taman Jaya Utama,42500 Telok Panglima

Garang,Kuala Langat,Selangor

Tarikh :   12 Januari 2024   Tarikh : 12 Januari 2024

*CATATAN: Jika laporan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali tempoh laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I declare that this report entitled "AN INVESTIGATION OF FRONTAL ALPHA ASYMMETRY AND DELTA–BETA CORRELATION DURING COGNITIVE IMPAIRMENT IN INDIVIDUALS WITH PUBLIC SPEAKING ANXIETY" is the result of my own work except for quotes as cited in the references.
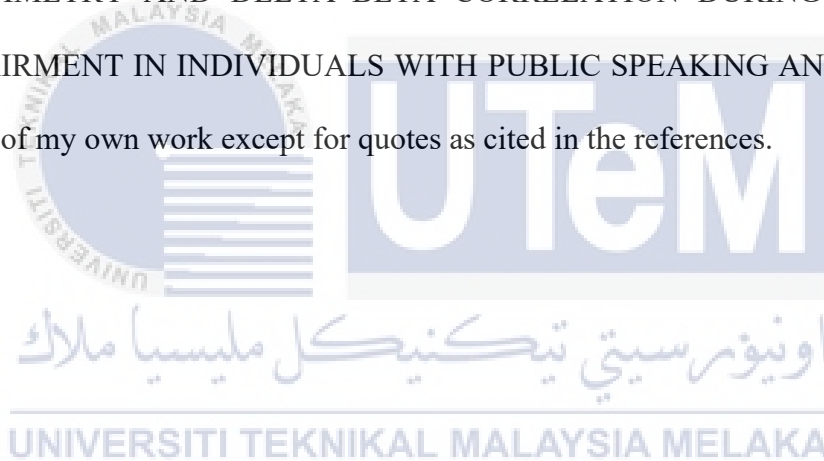
Signature   :   *Soon*
....

Author   :   Beh Soon Tatt
....

Date   :   12-01-2024
....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Computer Engineering with Honours.

Signature   :   …………………………….

Supervisor Name   :   DR. FARAH SHAHNAZ BINTI FEROZ

                                   12-01-2024

Date   :   ………………………….

DEDICATION

To the exciting findings of new EEG Biomarkers in frequency domain.

ABSTRACT

Public Speaking Anxiety (PSA) is the fear of speaking in front of an audience. This study aimed to investigate the brain-behavior mechanisms underlying PSA through the analysis of Electroencephalogram (EEG) data and to accurately classify PSA subjects using performance and EEG biomarkers. 24 subjects, categorized into high (HPSA) and low (LPSA) PSA groups, engaged in the Stroop Task. Their raw EEG data are recorded preprocessed using Independent Component Analysis (ICA) for artifact removal, and Fast Fourier Transformation (FFT) to obtain power density values. Frontal Alpha Asymmetry (FAA) refers to the asymmetrical distribution of alpha brainwave activity in the frontal region of the brain, and Delta-Beta Correlation (DBC) refers to the relationship between delta and beta brainwave frequencies in the context of neuroscience. Both FAA and DBC were investigated in the power density values between the groups of subjects. Findings align with expectations: HPSA subjects exhibit higher right FAA compared to subjects with LPSA, and statistically significant DBC is identified in electrodes. The mean accuracy of categorizing PSA subjects using Logistic Regression is 78.12%, and with Random Forest, it is 69%. FAA and DBC are useful in biomedical engineering as they successfully unveil the biomarkers of cognitive impairment during PSA.

ABSTRAK

Ketakutan Berbicara di Depan Umum (PSA) merujuk kepada rasa takut berbicara di hadapan orang ramai. Kajian ini bertujuan menyelidik mekanisme tindak balas otak terhadap PSA melalui analisis EEG, serta mengklasifikasikan subjek dengan penanda prestasi dan biomarker EEG. 24 subjek, terbahagi kepada kumpulan tinggi (HPSA) dan rendah (LPSA) PSA, melibatkan diri dalam tugas Stroop. Data EEG mentah diproses dengan Analisis Komponen Bebas (ICA) untuk penyingkiran artefak, dan Transformasi Fourier Cepat (FFT) untuk mendapatkan nilai ketumpatan kuasa. Asimetri Alpha Frontal (FAA) merujuk kepada pengagihan tidak simetri aktiviti gelombang otak alpha di kawasan frontal, dan Korelasi Delta-Beta (DBC) merujuk kepada hubungan antara frekuensi gelombang otak delta dan beta dalam konteks neurosains. Kedua-dua FAA dan DBC disiasat dalam nilai ketumpatan kuasa di antara kumpulan subjek. Hasil sejajar dengan jangkaan: subjek HPSA menunjukkan nilai FAA kanan yang lebih tinggi berbanding LPSA dan DBC yang signifikan dikenalpasti di beberapa bahagian otak subjek. Ketepatan purata mengklasifikasikan subjek PSA menggunakan Regresi Logistik ialah 78.12%, dan dengan Rawak Hutan, 69%. FAA dan DBC berguna dalam kejuruteraan bioperubatan dengan mengungkap biomarker kecacatan kognitif semasa PSA.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

PSA   :   Public Speaking Anxiety

EEG   :   Electroencephalogram

ICA   :   Independent Component Analysis

FAA   :   Frontal Alpha Asymmetry

DBC   :   Delta-Beta Correlation

LPSA   :   Low Public Speaking Anxiety

HPSA   :   High Public Speaking Anxiety

RT   :   Reaction Time

ERP   :   Event-Related Potential

FFT   :   Fast Fourier Transformation

SDG   :   Sustainable Development Goals

DFT   :   Discrete Fourier Transformation

SAD   :   Social Anxiety Disorder

ML   :   Machine Learning

LR   :   Logistic Regression

RF   :   Random Forest

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

The introduction, project overview, problem statement, objectives, scope of work, project significance and thesis outlines of this study will be cover in this chapter.

1.1   Introduction

Public Speaking Anxiety (PSA) is a prevalent anxiety disorder affecting approximately 73% of the global population[1]. According to [2]. PSA is a typical type of social anxiety disorder, characterized by an individual's exaggerated nervousness and trepidation concerning speeches or presentations. It is a normal condition for PSA individuals to feel worries and nervous for extended period of time, sometimes stretching weeks or even months before the actual speaking engagement. Furthermore, the PSA individuals will find themselves displaying some physical symptoms like trembling and upset stomach when it comes to a real-life public

speaking situation. This widespread condition has prompted numerous investigations into PSA, aiming to comprehend its nuances and devise more effective therapeutic interventions. Notably, research, as cited in [3], has sought to provide insights into PSA and enhance our understanding of the underlying mechanisms and develop more effective therapeutic approaches. Despite its dominance, there are currently no EEG biomarkers in the frequency domain to identify individuals with PSA. Efforts to establish such biomarkers could revolutionize our approach in diagnosing and treating PSA.

Frontal Alpha Asymmetry (FAA) and Delta-Beta Correlation (DBC) are suggested electroencephalographic (EEG) measurements that can serve as potential EEG biomarkers for anxiety disorders, with a specific focus on PSA. A pivotal research [4] focused on evaluating FAA in children with anxiety during a task centered on discerning threatening stimuli revealed a pronounced void in the existing literature. To bridge this gap, the research incorporated 77 children aged 8 to 12, with 36 identified as individuals displaying heightened anxiety levels. The outcomes illuminated distinct patterns of alpha power fluctuations among participants when confronted with threatening stimuli. This research advances our understanding of the neurobiological aspects of anxiety disorders and it underscores the significance of FAA as a pivotal biomarker in the context of anxiety disorders.

A different study [5] inquiry further explored the examination of the DBC, concentrating on its implications and outcomes. The study sought to clarify the intricacies of this correlation within the framework of behavioral inhibition (BI) and anxiety The research involved 118 children, and those who scored high on behavioral inhibition (BI) exhibited an increased correlation between delta and beta brain waves, particularly in frontal and central regions, with a slight elevation in parietal areas. This

heightened correlation suggests a potential neural marker linked to BI, indicating delta-beta correlation as a biomarker for BI in individuals with anxiety.

In the era of Society 5.0, an accurate classification of PSA patients using advanced analysis would help identify individuals with PSA, enabling earlier and better treatment. This study aims to develop an accurate classification of PSA using Logistic Regression and Random Forest. The integration of Logistic Regression and Random Forest promises a sophisticated and nuanced approach to PSA classification, with the ultimate goal of facilitating more effective and targeted interventions for those grappling with the challenges of public speaking anxiety in both personal and professional spheres.

In summary, Public Speaking Anxiety (PSA) is a widespread and worrisome phenomenon in contemporary society, impacting a substantial segment of the global population. Considerable research has been dedicated to unravelling its origins and devising more efficient treatments. Although there are currently no EEG biomarkers in the frequency domain specifically designed to identify individuals with PSA, Frontal Alpha Asymmetry (FAA) and Delta-Beta Correlation (DBC) have emerged as EEG biomarkers for anxiety disorders, offering avenues for exploration within the domain of PSA. A precise classification is required for PSA using EEG and performance biomarkers for the early detection of PSA subjects.

1.2   Project Overview

Public Speaking Anxiety (PSA), a pervasive fear of speaking in front of an audience. This research endeavors to illuminate the intricate brain-behavior mechanisms underlying PSA, employing advanced Electroencephalogram (EEG) data analysis to unravel the nuances of this anxiety disorder. The primary objective is to not only deepen our understanding of the neural correlates associated with PSA but

also to develop a robust classification model for accurately identifying individuals with varying levels of public speaking anxiety.

Twenty-four subjects, categorized into high (HPSA) and low (LPSA) PSA groups, actively participated in a well-established cognitive task—the Stroop Task. This task served as a controlled environment to elicit cognitive responses, allowing us to investigate neural activities during a cognitive challenge. The subjects' raw EEG data underwent meticulous preprocessing, involving Independent Component Analysis (ICA) for artifact removal and Fast Fourier Transformation (FFT) to derive power density values. This methodological rigour ensures the extraction of meaningful neural biomarkers related to PSA.

This study centers on two crucial EEG biomarkers: Frontal Alpha Asymmetry (FAA), which reflects the asymmetrical distribution of alpha brainwave activity in the frontal region, and Delta-Beta Correlation (DBC), which explores the relationship between delta and beta brainwave frequencies. These biomarkers offer unique insights into the neural dynamics associated with Public Speaking Anxiety (PSA). The study also employs Logistic Regression (LR) and Random Forest (RF) machine algorithms for the classification of individuals with PSA.

1.3   Problem Statement

Based on past research papers (see Appendix A), EEG Biomarker analyses in the frequency domain has never been investigated for PSA individuals. This is despite earlier studies underscoring the significance of EEG biomarkers in the frequency domain among patients with anxiety disorders (refer to Section 1.1 Introduction). The limited exploration of the relationship between FAA and DBC in individuals with PSA highlights the need for more comprehensive and interdisciplinary approaches to

understanding the neural mechanisms underlying this fear and its impact on cognitive functioning. Although PSA is a common anxiety disorder that affects a large proportion of the population, the underlying mechanisms of PSA remain incompletely understood. This lack of understanding has led to restricted treatment options and impeded the development of effective interventions for managing this anxiety disorder.

The literature gap of the lack of biomarkers in the frequency domain (i.e. FAA and DBC) in PSA studies is depicted in Appendix A. As demonstrated in the Appendix A, none of the studies has conducted a comparative analysis of Frontal Alpha Asymmetry (FAA) and Delta-Beta Correlation (DBC) between individuals with High Public Speaking Anxiety (HPSA) and Low Public Speaking Anxiety (LPSA) using the Stroop paradigm, despite both analyses yielding EEG biomarkers in patients with anxiety disorders. A gap exists in the literature regarding a direct comparison in this context in subjects with PSA.

ERP/time-domain biomarkers, specifically the N200 and P200, were identified in the study conducted by [8] and [9], suggesting their relevance in understanding the cognitive processes involved in PSA. Further research is imperative to investigate the relationships between these biomarkers and other cognitive processes, as well as to explore their potential applications in the development of treatments and therapies for individuals with PSA disorders. Additionally, there is a pressing need for an accurate classification method for individuals with PSA because the existing method for classifying individuals with PSA relies on subjective self-report measures, which may be biased. Utilizing machine learning approaches for accurate classification could

provide greater insights into factors and biomarkers contributing to the categorization of individuals with PSA.

1.4 Proposed Solution to the Problem Identified

The proposed solution involves leveraging advanced EEG data analysis techniques, including Independent Component Analysis (ICA) and Fast Fourier Transformation (FFT), FAA and DBC to uncover EEG biomarkers in the frequency domain related to PSA. By understanding the neural patterns associated with high PSA, interventions can be designed to target and alleviate cognitive impairment during public speaking situations. Besides that, Logistic Regression and Random Forest machine learning algorithms are applied to classify individuals with PSA and without PSA accurately.

1.5 Objectives

The primary aim of this study is to investigate the brain-behaviour mechanisms underlying Public Speaking Anxiety (PSA) through the analysis of Electroencephalogram (EEG) data. Therefore, the following objectives have been established:

- To identify EEG artifacts and reject them using Independent Component Analysis (ICA).
- To compare the Frontal Alpha Asymmetry (FAA) of cognitive impairment in HPSA and LPSA individuals.
- To identify the brain regions exhibiting statistically significant Delta-Beta Correlation (DBC) of cognitive impairment in HPSA and LPSA individuals
- To classify individuals with HPSA and LPSA by utilizing machine learning logistic regression (LR) and random forest algorithm (RF) based on performance and EEG biomarkers.

1.6 Scope of Work

、 This research project endeavors to delve into the neural underpinnings of Public Speaking Anxiety (PSA) by analyzing EEG biomarkers, specifically Frontal Alpha Asymmetry (FAA) and Delta-Beta Correlation (DBC). Recognizing the inherent constraints and assumptions, this study aims to provide valuable insights into the neural mechanisms of PSA and their impact on cognitive functioning, with the ultimate goal of informing effective interventions for anxiety disorders. Acknowledging the inherent challenges, this study confronts limitations such as participant availability, time constraints affecting data collection and analysis, and resource limitations. These constraints set the boundaries for the achievable outcomes within the project.

The project is built upon key assumptions, including the suitability of the Stroop paradigm to induce PSA and the belief that selected EEG biomarkers (FAA and DBC) accurately reflect the neural mechanisms associated with PSA. The hypothesis posits increased right frontal alpha activation in High PSA (HPSA) subjects. Additionally, the study hypothesizes a significant correlation between changes in delta and beta brainwave patterns (DBC) and PSA levels. These assumptions and constraints are strategic choices aimed at conducting a focused and feasible research study. The justification lies in the imperative need to unravel the neural intricacies of public speaking anxiety, paving the way for targeted interventions and treatments.

The primary focus is the analysis of EEG biomarkers (FAA and DBC) in two distinct groups: 12 individuals with HPSA and 12 with Low PSA (LPSA) utilizing MATLAB. The raw EEG data will be preprocessed using ICA for artifacts removal and FFT to obtain the power spectrum. The power spectrum was analysed by using

FAA and DBC. Both performance and EEG biomarkers were harnessed for machine learning algorithms (Logistic Regression and Random Forest) to accurately classify subjects with HPSA and LPSA, utilizing the Python programming language in Visual Studio Code IDE. Certain aspects will not be covered, including the exploration of alternative experimental paradigms, the investigation of non-EEG biomarkers or physiological measures, and the development of specific interventions or therapies for PSA.

This research project aspires to contribute valuable insights into the neural mechanisms of PSA, navigating through inherent limitations with a focus on feasibility and relevance. The anticipated outcomes aim to not only enhance our understanding of PSA but also pave the way for the development of targeted interventions for individuals grappling with anxiety disorders related to public speaking.

1.7   Project Impact

This project holds significant implications for both the field of anxiety research in biomedical engineering and the development of effective interventions for individuals with Public Speaking Anxiety (PSA).

The impact of this project is that, up until now, there has been no research on EEG biomarkers in the frequency domain. By investigating the brain-behaviour mechanisms underlying PSA through the analysis of EEG data, this study aims to fill the current gap in understanding the EEG biomarkers associated with PSA in the frequency domain. The understanding of the brain-behaviour mechanisms can create a pathway for the evolvement of EEG biomarkers that can aid in the diagnosis and assessment of cognitive impairment in individuals with HPSA.

Moreover, the findings of this study have the potential to shed light on the neural correlates of PSA, advancing our understanding of the condition and its impact on cognitive functioning. This knowledge can inform the development of more targeted and effective interventions for individuals with PSA, leading to improved treatment outcomes and quality of life.

From the point of view of Sustainable Development Goals (SDG), there are three SDGs related to this study. The first SDG is SDG3: Good Health and Well-being, as it aims to study the brain-behaviour mechanism underlying PSA by analysing the EEG data collected, which has the potential to contribute to the new treatments of PSA patients. Besides that, SDG 9: Industry, Innovation, and Infrastructure is related to this study as it uses biomedical engineering techniques to analyse EEG data may have broader applications in the development of medical technologies and devices to support the diagnosis and treatment of a range of health conditions. Moreover, this project resonates with SDG 17: Partnerships for the Goals. The findings have the potential to foster the development of effective interventions for individuals with public speaking anxiety, which may necessitate collaboration and partnerships across different sectors and stakeholders for successful implementation.

Figure 1.1: SDG 3, 9 and 17

Considering the environmental impact, the Emotiv EPOC+ EEG device used in this project demonstrates eco-friendly characteristics. It prioritizes energy efficiency by utilizing rechargeable batteries instead of disposable ones, thus reducing electronic waste. Emotiv also adheres to industry standards for environmentally friendly materials during the manufacturing process of the EPOC+. Additionally, the device is designed for durability, minimizing the need for frequent replacements and further reducing electronic waste.

Overall, the significance of this project lies in its potential to enhance our understanding of the brain-behaviour mechanisms underlying PSA, contribute to the field of anxiety research, and establish a precedent for the evolvement of new biomarkers and interventions to support individuals with PSA.

1.8   Thesis Outlines

In short, Chapter 1 has introduced the research, providing a clear understanding of the project's context, the identified problem, and its significance. The objectives and scope of work have been outlined, setting the stage for the subsequent chapters. The thesis outlines offer a preview of the organization and content of the entire document. This chapter serves as a foundational guide for readers, ensuring they have a comprehensive overview before delving into the details presented in the subsequent chapters. This thesis starts with a revision of the related research and theories regarding crucial components that have been used in chapter two. A description of the specific methods for conducting the entire study is written in chapter three. Chapter Four comprises the discussion about all the results obtained from the research findings while the conclusion and future works are deliberated about in Chapter Five.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction of Literature Review

The title of this study is An Investigation of Frontal Alpha Asymmetry and Delta–Beta Correlation During Cognitive Impairment in Individuals with Public Speaking Anxiety. This chapter reviews the theoretical insights and past research papers related to this study.

2.2 Public Speaking Anxiety (PSA)

PSA is a widespread psychological phenomenon that manifests as intense fear and discomfort when speaking or presenting in front of an audience[14]. It affects individuals from various backgrounds and can significantly hinder their performance and well-being in social and professional settings. One notable aspect of PSA is the cognitive impairment experienced by those affected, which further exacerbates the

challenges associated with public speaking. Therefore, gaining a deeper understanding of the neural correlates underlying cognitive impairment in individuals with PSA is essential for the development of effective interventions and treatments. One study found P200 biomarkers for attentional bias in HPSA but not in LPSA subjects. The reaction time and the P200 biomarkers found are significantly related to the attentional bias for HPSA subjects [6]. Another study found that N200 and P200 biomarkers in HPSA subjects[7]. From this study, Reaction Time (RT) and Event-Related Potential (ERP) have proven to be valuable tools in modern medicine as they have effectively revealed biomarkers that indicate abnormalities during the interplay of emotion and cognition. These measures provide valuable insights into the underlying cognitive processes and can help identify deviations from normal functioning. By analysing RT and ERP, researchers can gain a deeper understanding of how emotions and cognitive processes interact, which can ultimately aid in diagnosing and treating various medical conditions related to emotion and cognition. To the best of our knowledge, no research paper to date has investigated the EEG biomarkers associated with PSA in the frequency domain. Below is the summary of PSA-related past research:

Table 2.1: Summary of PSA-Related Study

Year/Paper	Subjects	Method	Result	Similarities	Differences
Feroz [6]/2021	12 LPSA & 12 HPSA	ERP, RT, ICA	P200 Biomarker	PSA study	ERP (time domain).
Feroz [7]/2021	12 LPSA & 12 HPSA	ERP, RT, ICA	N200, P200 Biomarkers	PSA study	ERP (time domain).

2.3   Stroop Task

Stroop task is a psychology test that tests the reaction time between automatic and controlled processing of information where the meaning of the printed word interferes with the ability to name the colour of the word. Stroop task was first developed by an

American psychologist, John Ridley Stroop in 1935 and has been modified to investigate the additional brain mechanisms[15]. The Stroop task is divided into 2 congruencies of the font name and colour, which are Congruent (the word meaning is the same as its colour) and Incongruent (the word meaning is different from its colour). The task participants will be asked to identify and speak out the colour of the word and the time taken for participants to response will be considered as the Reaction Time (RT).

Figure 2.1: The Stroop Task

The example of the Stroop task is shown in Figure 2.1. The congruent stimuli would be similar to the first row. In this row, the word "Green" is displayed, and its meaning matches the font colour, which is also green. This congruence between the word and the font colour makes it easier for participants to identify the ink colour. On the other hand, the incongruent stimuli, such as the second row in Figure 2.1, involve a conflict between the word and the font colour. For instance, the word "Red" is written in green ink, creating an incongruent stimulus. This incongruence can create interference and make it more challenging for participants to correctly identify the ink colour, leading to longer response times.

2.4 Independent Components Analysis (ICA)

ICA techniques always be used to relate to a problem called "cocktail party problem"[16]. In this problem, there will be two people talking to each other at a cocktail party with two microphones near to both people. The microphones recorded

their voice in different volumes because they were standing far away from each other. Not only that, but the microphone also recorded the noise in their surroundings as they were at a cocktail party. The question arises as to whether it is possible to separate two voices from noisy recordings and how this separation can be achieved, and this can be done by using ICA. The cocktail party problem can be solved by using ICA is shown in Figure 2.2.
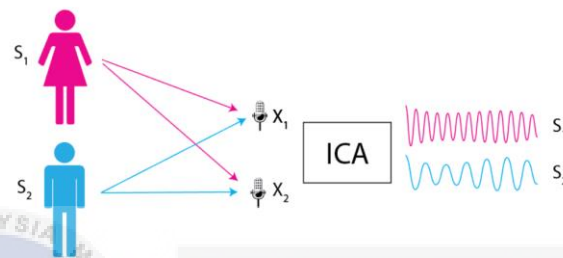
Figure 2.2: Cocktail Party Problem Solved by using ICA

Independent component analysis (ICA) is a statistical technique used to separate a multivariate signal into individual subcomponents that are maximally independent of each other[17]. This method transforms a collection of vectors into a set of components that exhibit high independence. Based on Figure 2.2, different signals from the microphone can be denoted by $X_1(t)$ and $X_2(t)$ where the speaking voice of those 2 people denoted by S_1 and S_2. The output equation of the signal can be expressed as linear equations in (2.1) and (2.2):

$$x_1(t) = a_{11}s_1 + a_{12}s_2 \qquad (2.1)$$

$$x_2(t) = a_{21}s_1 + a_{22}s_2 \qquad (2.2)$$

The variables of a_{11}, a_{12}, s_1 and s_2 are some parameters that depend on the distances of the microphones from the speakers. ICA can be used to extract s_1 and s_2

from $x_1(t)$ and $x_2(t)$ based on the equations above. The mixture signals of brain and muscle movements is the raw EEG data recorded and is denoted by x. The data mixing process is denoted by A whereas s is the representation of the component activations from the mixture signals. The complete equation for the raw EEG data recorded is shown in (2.3) followed by (2.4)

$$X = A \times s \qquad (2.3)$$

$$\text{Inverse Equation: } s = A^{-1} \times X \qquad (2.4)$$

The goal of applying Independent Component Analysis (ICA) in this study is to remove components such as eye blinks and heartbeat signals that affect the accuracy of the analysis. By utilizing the inverse equation, the original components of brain activity (denoted as s) can be obtained. The inverse of the mixing matrix (A^{-1}) can be considered as the topography map of the EEG signals, which provides insights into the spatial distribution of the brain activity components.

Based on [6], [7], [11] the raw EEG data is preprocessed using ICA to identify and remove the artifacts to obtain cleaned EEG data. In conjunction with this research, the brain electrodes act like microphones to pick up the mixture signals of brain and muscle movements and ICA will identify and isolate the pure brain signal from the mixture signals. In this study, a vector-matrix notation is utilized as it offers a more convenient and concise representation compared to the previous summation equations.

2.5 Frontal Alpha Asymmetry (FAA)

According to [12], FAA refers to the asymmetrical cortical activity in the alpha frequency range (8-13 Hz) within the frontal hemispheres of the brain. It is associated with various measures related to motivational states, stress biomarkers, and mood and

stress-related disorders. Left FAA is characterized by greater activation in the left frontal region compared to the right, while right FAA indicates greater activation in the right frontal region compared to the left. The left FAA is associated with better emotional regulation and approach-related-affect while the right FAA indexes behavioral inhibition and withdrawal-related-affect [12]. Therefore, individuals experiencing PSA may exhibit greater relative right frontal activation (Right FAA), reflecting a neural pattern associated with heightened vigilance and withdrawal-related effects. FAA usually involves the fronto-medial electrodes which are F3 and F4[18]. Extensive research has been conducted over the last three decades on the asymmetry of EEG alpha power in the frontal lobe, exploring its potential role as an indicator of emotional and motivational states. It is a reliable marker to study and understand these psychological and physiological processes. There are various formulae used to compute FAA in past research and they are all shown in Table 2.2.

Table 2.2: Formulae of FAA in past study

Paper/Year	FAA Formula
Flasbeck[8]/2023	FAA= ln [Right]-ln [Left]
Schürmann[9]/2022	Asymmetry Index = ln [Right] – ln [Left]
Barros [11]/2022	FAA = ln(F4) −ln(F3)
Glier[12]/2022	Laterality coefficient (LC) $= \frac{R-L}{R+L} \times 100$
Wise[4]/2023	FAA at the F3/F4 site = right alpha power – left alpha power
David[13]/2021	FAA =log (Alpha Right Electrode) − log (Alpha Left Electrode)
F. Heine [18]/2022	FAA = ln [Right] – ln [Left]
S. Song *et al* [19]/2023	FAA Index = ln [Right] – ln [Left]

Some studies have employed the natural logarithm of the power in the right and left frontal regions (ln [right] – ln [left]) as the formula for FAA calculation. This formula has been frequently used and has shown reliability in previous studies. It is important

to note that different studies may utilize different formulas based on their specific research objectives and electrode placements. The formula employing the natural logarithm of the power in the right and left frontal regions is more commonly used in previous research, indicating its relevance and consistency across studies.

In this literature review, we aim to explore the current state of knowledge regarding the FAA and DBC and their implications. The summary of FAA-related past research is depicted in Appendix B.

Flasbeck [8] conducted a study with 36 participants who reported having a fear of flying (FF) and 41 unaffected participants (NFF). The results indicated that individuals with a fear of flying exhibited higher right frontal alpha activity at F8-F7 electrodes compared to the unaffected group. This finding suggests a potential neurophysiological marker related to fear of flying, as indicated by increased activation in the right frontal alpha region.

Schumann's study [9] involved 47 healthy adults, and they utilized techniques such as ICA, FFT and FAA. The study aimed to explore the relationship between alpha power lateralization and depressive symptoms. The findings revealed an association between relative rightward lateralization of alpha power at one electrode pair and depressive symptoms. This suggests that greater activation in the right frontal alpha region may be related to depressive symptoms in healthy adults.

Barros [11] conducted a study comparing 39 older adults ($\geqslant$ 60 years old) and 57 younger adults (between 18 and 35 years old). The researchers employed ICA and FAA to examine age-related differences in alpha asymmetry. The results showed that older adults had higher FAA values than younger adults. This suggests that there may

be age-related variations in frontal alpha asymmetry, with older adults displaying a different pattern of brain activation compared to younger adults.

Glier's study[12] focused on 145 adolescents and investigated the relationship between FAA, trait anxiety, state anxiety, and cortisol reactivities. The study utilized techniques such as ICA, FFT and FAA. The findings revealed that adolescents with rightward FAA activation and high trait anxiety exhibited blunted cortisol reactivities, while those with leftward FAA activation and high state anxiety showed prolonged cortisol recoveries. This suggests that different patterns of frontal alpha asymmetry may be associated with distinct physiological responses to anxiety in adolescents.

Wise[4] conducted a study with 77 children aged between 8 and 12 years old, including 36 high anxious children. The study employed the Gratton method, FFT and FAA. The results showed that during the face and images task, higher alpha power was observed in the left hemisphere in response to threat compared to neutral stimuli, with no significant difference in the right hemisphere. However, there were no significant changes in alpha power values in response to word stimuli. These findings suggest that threat-related stimuli may elicit differential patterns of frontal alpha activation in children, depending on the nature of the task.

David[13] conducted a study with 165 healthy children and adolescents aged between 10 and 16 years. The study employed ICA, FFT and FAA. The researchers investigated the relationship between FAA, state anxiety, and the effectiveness of the REThink intervention. The results indicated that higher scores on FAA were associated with more right-sided alpha activity, suggesting greater inhibition in the right hemisphere associated with negative affect. After the REThink intervention, there was a significant negative correlation between FAA and state anxiety.

Specifically, the group that received the REThink intervention showed a significant increase in right hemisphere inhibition, supporting the efficacy of the intervention in reducing negative modulation. This study used similar methods such as FAA, FFT, and ICA, but had different subjects and an experimental paradigm focused on the effectiveness of the REThink intervention.

Song's study[19] included 62 healthy university students and employed techniques such as Event-Related Potentials (ERP), ICA and FAA. The study aimed to examine the differences in occipital alpha oscillation power and FAA between individuals with social anxiety and healthy controls under different emotional contexts. The findings revealed that in congruently emotional contexts and expressions, the social anxiety group exhibited significantly lower occipital alpha oscillation power compared to the healthy control group. Additionally, in negative contexts, frontal alpha lateralization was significantly lower in the social anxiety group compared to the healthy control group. This study employed similar techniques to FAA and ICA but had different subjects (university students) and utilized ERP as an additional method.

Heine[18] conducted a study with 35 university students, using techniques such as FAA, Infinite Impulse Response (IIR) filter, ICA-based Electrooculography (EOG) correction, and FFT. The study aimed to investigate the relationship between depression, depression scores measured by the PHQ-9, and alpha asymmetry. The results indicated no difference in alpha asymmetry between the two depression groups used in the study. Furthermore, there was no significant association between depression scores measured by the PHQ-9 and alpha asymmetry measured on different electrode pairs. This study employed similar techniques such as FAA, ICA, and FFT

but had a different method (IIR filter) and focused on the relationship between depression and alpha asymmetry.

These studies assist us in learning how Frontal Alpha Asymmetry (FAA) connects to mental feelings such as worry and sadness. Methods like FAA, Fast Fourier Transform (FFT), Independent Component Analysis (ICA) and others are used to study how different things affect the relationship between them. While the studies share similarities in methods used, they differ in terms of subjects, experimental paradigms, and specific variables investigated, providing a diverse perspective on the brain-behavior mechanisms associated with FAA. ICA is used for data preprocessing and FFT is for computing power spectral for computation of FAA in most past research papers.

In summary, these studies show the importance of FAA as a biomarker for anxiety studies. These studies have investigated the role of FAA in diverse psychological phenomena, such as state anxiety, social anxiety, and depression. Although these studies share methods like FAA, FFT, ICA, and ERP, they vary in subjects, experimental paradigms, and specific variables explored, providing a diverse perspective on the brain-behavior mechanisms associated with frontal alpha asymmetry. ICA is commonly used for data preprocessing, and FFT is employed for computing power spectral, similar to the current study.

2.6   Delta-Beta Correlation (DBC)

DBC is a phenomenon that arises from the interaction between neural oscillations in the beta and delta frequency bands[20]. Delta-beta correlation is a measure of the linear relationship between the power or amplitude of the delta and beta frequency bands in the brain. It has been studied concerning various cognitive processes and

psychiatric disorders. Beta waves (13 to 30 Hz) are rapid oscillations that originate in the cortex and are associated with higher-level cognitive processes such as attentional control. On the other hand, delta waves (0 to 4 Hz) are slower oscillations that reflect subcortical brain activity and are related to more fundamental affective processes. DBC, therefore, represents the coupling or coordination between these different frequency bands, reflecting the integration of various cognitive processes and bottom-up affective processes in the brain including attention, memory, and executive functions. In this literature review, we aim to explore the current state of knowledge regarding the delta-beta correlation and its implications. The summaries of the DBC's past research study is summarized in Appendix C.

In a study conducted by Harrwijn[10], 113 participants were involved, consisting of 18 individuals diagnosed with social anxiety disorder (SAD), 25 individuals diagnosed with subclinical SAD, and 43 individuals diagnosed with clinical SAD. The study employed the methods of DBC and FFT for analysis. The findings of the study revealed that participants with (sub)clinical SAD exhibited a higher negative delta-low beta correlation during anticipation compared to those without (sub)clinical SAD.

Myruski [20] studied 53 kids to see if DBC is linked with the use of emotion regulation (ER) methods that are changeable and active. He used both measures, DBC and Fourier transform (FFT), for this research. The findings showed that significant DBC were related to using more adaptive and active ways of thinking in the ER.

In a study conducted by Margaret[21], 184 adults with high social anxiety (SA) took part. The study used DBC and FFT to analyze things. The results showed no important change in the DBC among participants during the Writing Exercise

Condition. So, the study says this certain health problem did not really affect how closely related delta and beta were among adults with high SA.

Al-Ezzi [22] compared people with Social Anxiety Disorder (SAD) and healthy controls (HC) by analyzing DBC. The results showed a higher correlation between DBC during the baseline condition and SAD, whereas HC individuals exhibited a stronger correlation during the recovery state.

Poole[23] conducted a study that is participated by 67 children. The study utilized two methods of analysis: DFT and DBC. The study aimed to examine the relationship between shyness and brain activity. The findings revealed that positive shy children exhibited a higher frontal DBC compared to the other groups, namely non-positive shy and low shy children. This distinction in correlation patterns among the participant groups was observed through the application of DBC and DFT methods. These results suggest a potential link between shyness and specific brain activity patterns in children.

Poole [24] categorized salivary cortisol patterns and social anxiety into distinct classes in 50 children using DBC and DFT. It identified a high, stable class and a low, unstable class for both salivary cortisol patterns and social anxiety.

Poole [5] also studied DBC in children with behavioral inhibition (BI) by applying DBC and FFT another study. The findings indicated that children with high BI showed a higher DBC in frontal and central brain regions compared to children with low BI. There was also a marginal increase in delta-beta correlation in parietal brain regions for high BI children.

In the study conducted by Pascalis [25] involving 59 students, the researchers tried to figure out if DBC is connected with anxiety. FFT and Gratton's method were used

for this investigation. The findings showed the people who were anxious and didn't move had a big positive connection. This was more than what happened with the group that just relaxed. The research showed that DBC happened mostly when there was little delta activity in the Anxiety group, but not in the Relaxation one. These results show a connection between DBC and anxiety states, implying that maybe DBC can be an indicator of anxiety-related things happening when we are resting.

To summarize, the collective findings from these studies consistently point towards the potential utility of DBC as a valuable biomarker for anxiety-related processes. The investigations conducted by Harrwijn [10], Myruski [20], Margaret [21], Al-Ezzi [22], Poole [5], [23], [24], and Pascalis [25] collectively contribute to building a compelling case for the role of DBC in understanding and assessing anxiety-related phenomena across diverse populations. The observed correlations between DBC patterns and various factors such as social anxiety, emotion regulation strategies, shyness, salivary cortisol patterns, and behavioral inhibition in children, as well as anxiety states in students, underscore the potential of DBC as a sensitive indicator of underlying processes associated with anxiety. This emerging consensus across multiple studies underscores the importance of further exploration and validation of DBC as a robust and informative biomarker in the realm of anxiety research.

2.7   Machine Learning

The Logistic Regression and Random Forest machine learning algorithms are employed to develop the classification system for PSA.

2.7.1   Logistic Regression(LR)

LR is selected as the primary choice because it is one of the most commonly used multivariate regression algorithm for binary classification[26]. It used the dependent

variable as the logarithm of odds and predicted the occurrence of an event with a logit function. The working principle of the LR model is to find out the connection between the occurrence and non-occurrence of an event to explanatory inputs through the utilization of the maximum likelihood method. The binary outcome of the dataset used for the LR model is usually "Yes" as 1 and "No" as 0, which are discrete values instead of continuous values. LR model is flexible because the dependent data can be quantitative, qualitative and a combination of both quantitative and qualitative. These variables can be converted into binary numbers 1 and 0. The equation of logistic regression is:

$$P(z) = \frac{1}{1 + e^{-(C_0 + C_1 X_1 + C_2 X_2 + \cdots + C_n X_n)}}$$

where,

P(z) is the probability of PSA in this study,

C0 is the intercept of the model,

Ci represents the regression coefficient which is obtained from maximum likelihood in conjunction with their standard errors.

In this study, HPSA represents 1 and LPSA represents 0, and the threshold value taken for the probability of PSA is 0.5. (P(z)>0.5)

2.7.2 Random Forest (RF)

RF is a common ML algorithm used to solve the problem of classification and regression[27]. This algorithm is easy to use, flexible and can avoid overfitting if the tree assigned to the model is enough in the forest[28]. RF is a supervised learning

algorithm where it builds an ensemble of decision trees with the training method of bagging. The bagging method is just to merge all the decision trees to obtain a more accurate and stable predicted output. RF adds randomness to the model while growing the trees. This algorithm will search for the best feature among a random subset of features to make a better model that produces a wide diversity result.

2.7.3 Summary of Machine Learning Past Research Paper

Table 2.3: Summary of Machine Learning Past Research Paper

Paper	Sample Size	ML Algorithm Used	Accuracy (%)	Precision (%)	AUC score
Jennifer [27] /2021	275	LR	80.00	-	0.84
		RF	88.00	-	0.93
Kumar [26] /2022	490	LR	95.90	88.90	0.96
		RF	96.94	81.82	0.874
Nishi Yadav [29] /2022	101	LR	95.23	100	-
		RF	90.47	91.66	-
Bernard [30] /2022	3001	LR	73.70	-	0.684
		RF	77.70	-	-

From the summary, the observed accuracy, precision and AUC score are overall quite high (most of them more than 90%). They demonstrated strong performance, establishing LR and RF as robust machine learning models capable of making accurate predictions for the current dataset.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The methodology section provides an overview of the approach used to experiment. It describes the overall framework and procedures employed to achieve the study's objectives. By outlining the methodology, readers gain an understanding of how the research was carried out and can evaluate the reliability and validity of the results.

3.2 Project Flowcharts

The project flowchart is shown in Appendix D. The project flowchart represents the process flow of this thesis where the EEG and Behavioral data are collected, and the EEG data will undergo a preprocessing step where ICA is used to identify and reject the artifacts for the EEG data. The aritfacts are the eye blinks, muscle

movements and environmental interference. If the artifacts are not fully removed from the EEG data, the process needs to be repeated until all artifacts are removed and at this state, the EEG data is considered clean.

The cleaned EEG data will undergo FFT to obtain power spectrum for FAA and DBC (EEG Biomarkers) Analysis. The FAA will be analysed using Repeated measures ANOVA and DBC will be analysed using Pearson Correlation.

After that, The Logistic Regression and Random Forest machine learning algorithms are employed to develop the classification system for PSA based on the EEG and performance Biomarkers. Based on the interpreted data from FAA, DBC and classification results of machine learning algorithms, a comprehensive report is written which serves as the summary of findings, conclusions drawn from analysis and future work recommendations based on the insights gained.

3.3   Sample Description

The sample group of the experiment conducted is divided into two groups, which are LPSA and HPSA subjects. The experiment has specifically selected participants who meet certain criteria. These criteria include being undergraduate students at UTeM and falling within the age range of 18 to 26 years old. Individuals who have current substance abuse or dependency, major somatic or neurological disorders, colour blindness, or a history of reading disorder have been excluded from the initial stage of participant selection.

3.3.1   LPSA Subjects

A total of 12 participants with LPSA levels were included in the experiment. These individuals were selected based on their responses to a questionnaire that was

distributed at the start of the study. The questionnaire was distributed using electronic messaging platforms such as WhatsApp and Google Forms.

3.3.2   HPSA Subjects

An additional set of 12 participants with high PSA levels were included in the second sample group. These individuals were also selected based on their responses to the questionnaire distributed earlier. The questionnaire played a crucial role in identifying and categorizing the participants' conditions, enabling them to be placed in the appropriate group.

3.4   Experimental Paradigm

The Modified Stroop Word-Colour Test was utilized as the experimental paradigm in this study. This test builds upon the principles established by numerous researchers who have employed it to examine the brain's automatic processing capabilities. It is commonly employed in psychological research to assess the cognitive control exhibited by the brain [15].

3.4.1   Paradigm Arrangement

The stimuli are shown on a HP monitor and presented in a dark background. The Stroop task began with a fixation point, followed by congruent and incongruent conditions. The purpose of the fixation point was to encourage the subjects to focus on the upcoming stimuli position, which helped eliminate unnecessary delays in responding to the test. There were two conditions in this experimental paradigm, congruent and incongruent, which will be determined by the colour-printed word. The participants need to answer the colour of the printed word as fast and accurate as they can.
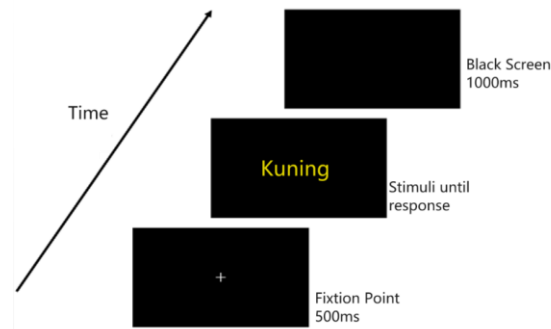
Figure 3.1: Modified Stroop Test Experiment Paradigm (Congruent)

The congruent condition of the Modified Stroop Test is shown in Figure 3.1. In the congruent state, the word printed will be 'Kuning,' which means yellow in English and will be printed with a yellow colour.
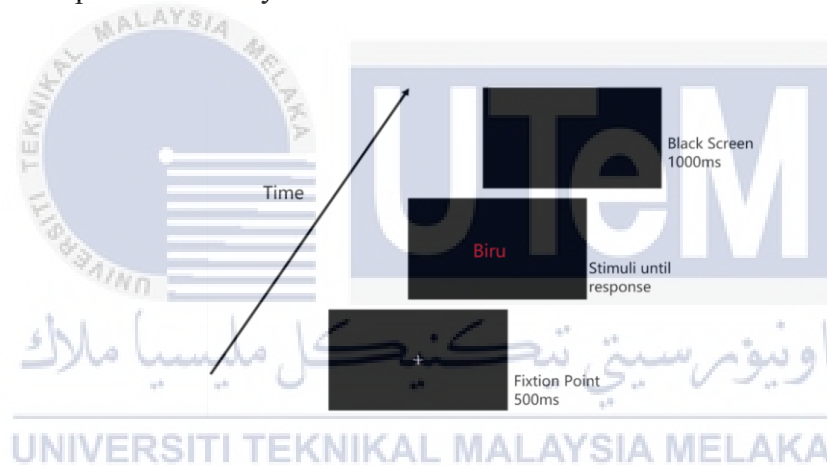
Figure 3.2: Modified Stroop Test Experiment Paradigm (Incongruent)

The incongruent state of the Modified Stroop Test is shown in Figure 3.2. For the incongruent state, it will be harder for participants to respond since the colour and meaning of the printed word are different. After responding to the stimuli, a blank screen was displayed for 1000ms, followed by the fixation point being displayed for 500ms. The process were repeated until all 120 stimuli have been answered. The time interval between the moment subjects read the word and the time they take to process the answer will be the reaction time window for this study.

3.5   Preprocessing of EEG data

The EEG data preprocessing pipeline for this study will involve several steps. Firstly, EEGLAB is used for filtering the raw EEG signals with a bandpass filter ranging from 0.3 to 30 Hz and setting the sampling rate to 256 Hz. Secondly, artifact rejection is performed to eliminate noise arising from muscle movements, eye movements and blinks by using ICA.
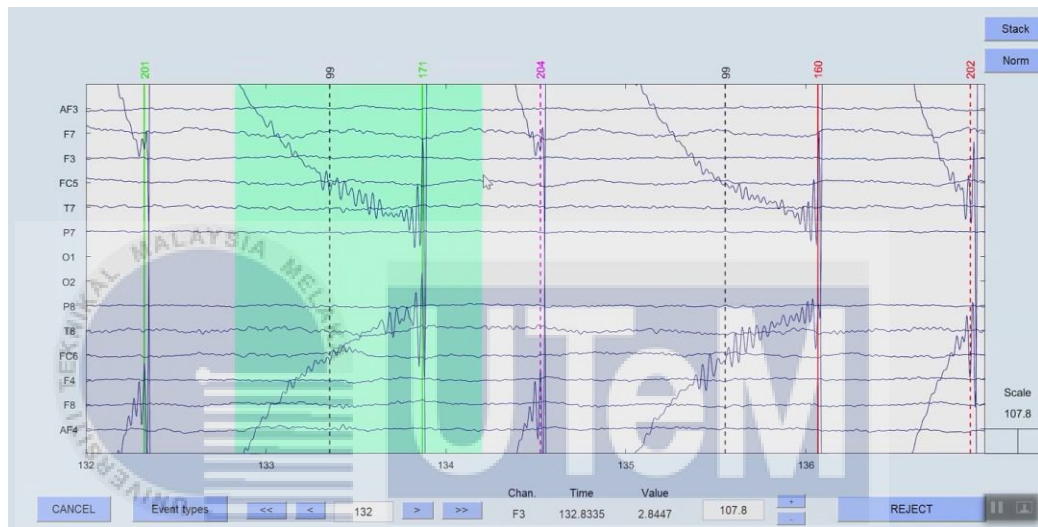
Figure 3.3: ICA Rejection

The ICA rejection is performed in EEGlab tools of MATLAB to remove the artifacts is shown in Figure 3.3. ICA is a powerful tool that can be used to separate EEG signals into independent components, making it easier to identify and remove sources of noise in the data.

Subsequently, we generated epochs of 1700ms for each condition, starting 200ms before the stimulus onset and ending 1500ms after the stimulus onset. These epochs were evaluated using ERPLAB for further analysis. To remove the effects of baseline variations, we performed baseline correction with a period of 150ms before stimulus onset.

Finally, the ERPs are averaged for each subject and condition to produce the averaged ERP waveforms, which were used for further statistical analysis. This comprehensive preprocessing pipeline ensured that the EEG data was cleaned and ready for analysis and free from artifacts, which will improve the accuracy and reliability of our results and allow us to draw robust conclusions from the study. After that, the Fast Fourier Transform is performed for computing the power spectrum of the cleaned EEG data as shown in Figure 3.4. The power spectrum obtained is for the computation of FAA and DBC.
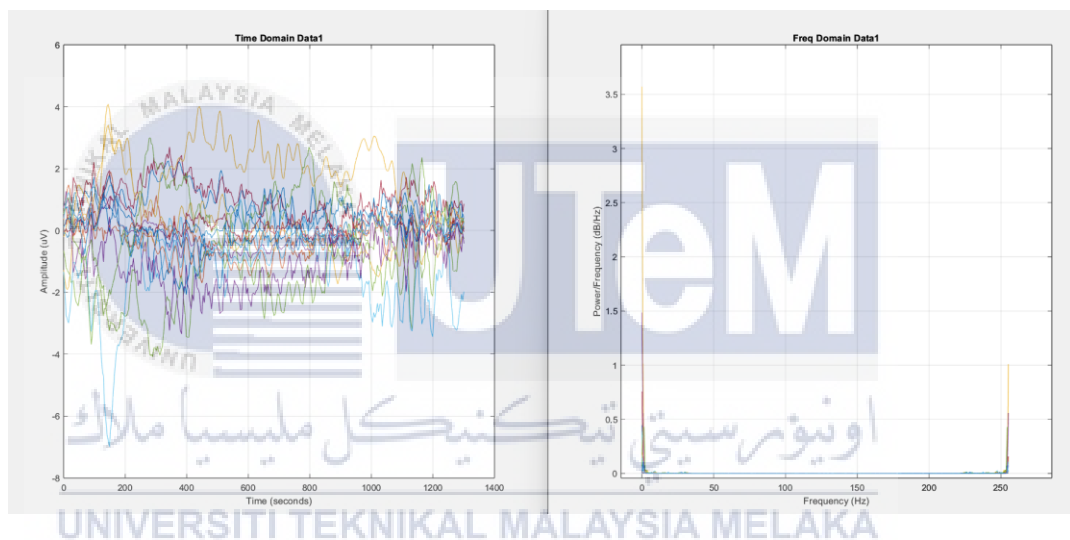
Figure 3.4: EEG Time Domain Data and Frequency Domain Data after performed FFT

3.6   Frontal Alpha Asymmetry (FAA)

To calculate FAA, the EEG data needed to be in power spectrum form. After the raw EEG data was preprocessed with ICA using EEGLAB and MATLAB software to obtain the cleaned EEG data, FFT algorithm was used to compute the power spectra. The alpha frequency band in the left and right frontal regions of the brain (8-13Hz) was filtered out from the power spectra obtained. FAA index was calculated by using the formula of the natural logarithm form which is FAA=ln(F4)-ln(F3) where F4 represents the power spectrum of the right frontal region, and F3 represents the power

spectrum of the left frontal region. To investigate the effects of the FAA on different experimental conditions, the FAA index was computed for both congruent and incongruent power spectra obtained from the FFT. This analysis was performed for each subject, including both HPSA and LPSA subjects, enabling a comprehensive exploration of the FAA phenomenon and its potential variations across individuals. After FAA computation, the FAA index values were to be used for the general linear model for HPSA and LPSA groups for further statistical analysis.

3.7   Delta-Beta Correlation (DBC)

To analyse DBC, the power spectrum of the EEG data collected during the Stroop task were utilized. The EEG data underwent a filtering process to extract the delta frequency band (0-4 Hz) and the beta frequency band (13-30 Hz). By isolating these specific frequency ranges, the analysis can focus on the neural activity associated with delta and beta oscillations.

Once the delta and beta frequency bands were extracted from the power spectrum, DBC was computed for each subject. This calculation was performed using the Pearson Correlation method, similar to the approach employed for calculating the FAA index. The Pearson Correlation coefficient measures the strength and direction of the relationship between the delta and beta oscillations within each subject. The DBC was computed for different electrode variables and is shown in Table 3.1.

Table 3.1: Electrode for DBC
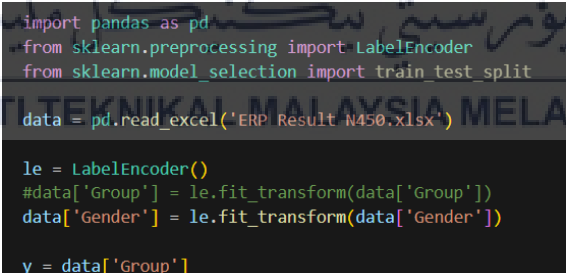
Electrode Variable for DBC	Electrode used
Frontal	F3, F4
Central	FC5, FC6
Parietal	P7, P8
Temporal	T7, T8

By examining the DBC concerning the experimental conditions (congruent and incongruent), insights can be gained into the neural dynamics and potential interactions between the delta and beta frequency bands during the Stroop task. This analysis will provide valuable information regarding cognitive processing and attentional mechanisms associated with these frequency ranges in the brain.

3.8 Machine learning

3.8.1 Data Preprocessing

Before any machine learning model is used for classification, the data of PSA subjects are tabulated into the excel file. The excel file is ensured not to have any missing values and outliers. The categorical variables in this study such as Group, are converted into binary values '1' and '0' by using Python programming language in Microsoft Visual Studio. The column "Group" is assigned to the dependent variable, y.

```python
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split

data = pd.read_excel('ERP Result N450.xlsx')

le = LabelEncoder()
#data['Group'] = le.fit_transform(data['Group'])
data['Gender'] = le.fit_transform(data['Gender'])

y = data['Group']
```

Figure 3.5: Data Cleaning for Machine Learning

3.8.2 Machine Learning Algorithms

Two machine learning algorithms are used, which are Logistic Regression and Random Forest. The features, denoted as X, for the Logistic Regression Model depicted in Figure 3.6 comprise features and outcomes related to the Incongruent condition. Specifically, these variables include Group, Subject, FC5 under Incongruent condition (FC5_Incon), Reaction Time under Incongruent condition

(RT_Incon), mu, tau, Gender, and the Number of Errors in answering the Stroop Task (Error_Con). In contrast, the Random Forest Model illustrated in Figure 3.7 employs the same set of features and outcomes as the Logistic Regression Model, with the additional inclusion of "Error_Con" as an independent variable within X. This extended set of features is intended to enhance the predictive capabilities of the Random Forest Model and is visually represented in Figure 3.7. The independent variables and dependent variables are split into 70% of the training dataset and 30% of the testing dataset.

```python
X = data.drop(columns=['Group','Subject','FC5_Incon','RT_Incon','mu','tau','Gender','Error_Con'])
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Figure 3.6: Variable Assign and Train Test Split for Logistic Regression Model

```python
X = data.drop(columns=['Group', 'Subject', 'FC5_Incon', 'RT_Incon', 'mu', 'tau', 'Gender'])
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Figure 3.7: Variable Assign and Train Test Split for Random Forest Model

3.8.2.1 Logistic Regression

The library of "statsmodel" and other necessary libraries is imported to instantiate the LR model. The model is fitted with the training dataset and the summary of the logistic regression model is displayed. K-fold validation and L1 regularization techniques are applied to the code to avoid overfitting and shown in Figure 3.8.

```python
kf = KFold(n_splits=num_folds, shuffle=True, random_state=42)
for train_index, val_index in kf.split(X_train):

    print('kth-fold=',k)
    X_train_fold, X_val_fold = X_train.iloc[train_index], X_train.iloc[val_index]
    y_train_fold, y_val_fold = y_train.iloc[train_index], y_train.iloc[val_index]

    # Create a logistic regression model with L1 regularization for each fold
    logit_model = sm.Logit(y_train_fold, X_train_fold)
    logit_result = logit_model.fit_regularized(method='l1', solver='liblinear', alpha=alpha)

    # Display summaries of the logistic regression model for each fold
    print(logit_result.summary())
```

Figure 3.8: K-Fold Validation

The performance of the logistic regression model is evaluated on the test dataset. The evaluation metrics are computed using the imported library and displayed such as confusion matrix, accuracy, precision, specificity, and sensitivity shown in Figure 3.9 and the mean of the evaluation metrics. The visualizations of the model are presented as ROC curves that shown in Figure 3.10.

```python
# Calculate evaluation metrics
#Confusion Matrix
confusion_matrix_result = confusion_matrix(y_test, y_pred)
#Plot confusion matrix as a heatmap with annotations
plt.figure(1, figsize=(5, 4))
heatmap = sns.heatmap(confusion_matrix_result, annot=True, fmt="d", cmap="Blues")
#Get axis information
ax=heatmap.axes
#Label TP, TN, FP, FN on the heatmap
ax.text(0.5, 0.2, f'True Negative', ha='center', va='center', color='black', fontsize=10)
ax.text(1.5, 0.2, f'False Positive', ha='center', va='center', color="black", fontsize=10)
ax.text(0.5, 1.2, f'False Negative', ha='center', va='center', color="black", fontsize=10)
ax.text(1.5, 1.2, F'True Positive', ha='center', va='center', color='black', fontsize=10)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title(f'Confusion Matrix of Fold {k}')
plt.show()

accuracy = accuracy_score(y_test, y_pred)
precision1 = precision_score(y_test, y_pred)
tn, fp, fn, tp = confusion_matrix_result.ravel()
specificity = tn / (tn + fp)
sensitivity = tp / (tp + fn)
```

Figure 3.9: Evaluation Metrics Code

```python
# Plot ROC curve
plt.figure(figsize=(8, 8))
plt.plot(fpr, tpr, color='darkorange', lw=2, label=f'ROC curve (area = {roc_auc:.2f})')
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title(f'ROC Curve - Fold {k}')
plt.legend(loc='lower right')
plt.show()
```

Figure 3.10: ROC Curve code

3.8.2.2 Random Forest

The library of "RandomForestClassifier" from sci-kit-learn and other necessary libraries is imported to instantiate the RF model. K-fold validation is applied to the code to avoid overfitting that is presented in Figure3.11.

```
#Random Forest
clf_k=RandomForestClassifier(max_depth=None,
                              min_samples_leaf=1,
                              min_samples_split=6,
                              n_estimators=20,
                              random_state=41)
num_folds=4
#Create a KFold object
kf=KFold(n_splits=num_folds, shuffle=True, random_state=41)
#Perform k-fold cross validation
cross_val_scores=cross_val_score(clf_k, X_train, y_train, cv=kf)
#cross_val_scores=cross_val_score(clf_k, X, y, cv=kf)
#Accuracy of All Fold
print("Cross-validation scores:", [round(score, 4) for score in cross_val_scores])
#Print the mean and standard deviation of the cross-validation scores
print(f"Mean accuracy: {cross_val_scores.mean():.2f}")
print (f"Standard deviation: {cross_val_scores.std():.2f}")
```

Figure 3.11: Random Forest Classifier and K-Fold Validation Code

The performance of the RF model is evaluated on the test dataset. The metrics are computed using the imported library and displayed such as the mean accuracy, standard deviation, confusion matrix, and classification report shown in Figure 3.12.

```
#Accuracy of All Fold
print("Cross-validation scores:", [round(score, 4) for score in cross_val_scores])
#Print the mean and standard deviation of the cross-validation scores
print(f"Mean accuracy: {cross_val_scores.mean():.2f}")
print (f"Standard deviation: {cross_val_scores.std():.2f}")

#Predict on the testing data for evaluation
clf_k.fit(X_train, y_train)
#Get predicted probabilities for class 1
y_pred_test=clf_k.predict_proba(X_test)[:, 1]
y_pred_test = (y_pred_test > 0.5).astype(int)
print("classification Report (Testing Data):")
print(classification_report(y_test, y_pred_test))

#Confusion matrix for testing data
cm_test=confusion_matrix(y_test, y_pred_test)
plt.figure(1, figsize=(5, 4))
heatmap=sns.heatmap(cm_test, annot=True, fmt="d", cmap="Blues")
#Get axis information
ax=heatmap.axes
#Label TP, TI, FP, FN on the heatmap
ax.text(0.5, 0.2, f'True Negative', ha='center', va='center', color='black', fontsize=10)
ax.text(1.5, 0.2, f'False Positive', ha='center', va='center', color="black", fontsize=10)
ax.text(0.5, 1.2, f'False Negative', ha='center', va='center', color="black", fontsize=10)
ax.text(1.5, 1.2, F'True Positive', ha='center', va='center', color='black', fontsize=10)
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title(f"Confusion Matrix of Testing Data")
plt.show()
```

Figure 3.12: Evaluation Metrics of RF

The visualizations of the model are presented as ROC curve, feature importance and learning curve as shown in Figure 3.13.

```
# Create a bar chart to visualize feature importances
plt.figure(figsize=(10, 6))
plt.barh(range(len(feature)), importances[sorted_indices])
plt.yticks(range(len(feature)), [feature[i] for i in sorted_indices])
plt.xlabel("Importance")
plt.ylabel("Feature")
plt.title("Feature Importances")
plt.show()

fpr, tpr, _ = roc_curve(y_train, y_pred_cv)
roc_auc = auc(fpr, tpr)
plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, color='darkorange', lw=2, label='ROC curve (AUC = {:.2f})'.format(roc_auc))
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic')
plt.legend(loc="lower right")
plt.show()


train_sizes, train_scores, test_scores = learning_curve(clf_k, X_train, y_train, cv=kf, train_sizes=np.linspace(0.1, 1.0, 5))
plt.figure(figsize=(8, 6))
plt.plot(train_sizes, np.mean(train_scores, axis=1), 'o-', label='Training score')
plt.plot(train_sizes, np.mean(test_scores, axis=1), 'o-', label='Validation score')
plt.xlabel('Training examples')
plt.ylabel('Score')
plt.title('Learning Curve')
plt.legend(loc='best')
plt.show()
```

Figure 3.13: Visualization of RF

The Random Forest model's decision tree can be visualized by exporting the tree structure into a DOT file within the Visual Studio Code IDE in Figure3.14. Subsequently, the DOT file can be converted into a PNG image using the Command Prompt displayed in Figure 3.15.

```
# Visualize one of the trees in the Random Forest
estimator = clf_k.estimators_[0]

# Export the tree to a DOT file
export_graphviz(estimator, out_file='tree.dot',
                feature_names=X_train.columns.tolist(),
                class_names=['Class 0', 'Class 1'],
                filled=True, rounded=True, special_characters=True)
```

Figure 3.14: Visualization of Tree of Random Forest

```
C:\Z- Jonathan\Utem Notes\Sem 7\PSM 2>dot -Tpng tree.dot -o tree.png
```

Figure 3.15: Conversion of DOT File into PNG File

3.9   Gantt Chart

Table 3.2: Gantt Chart
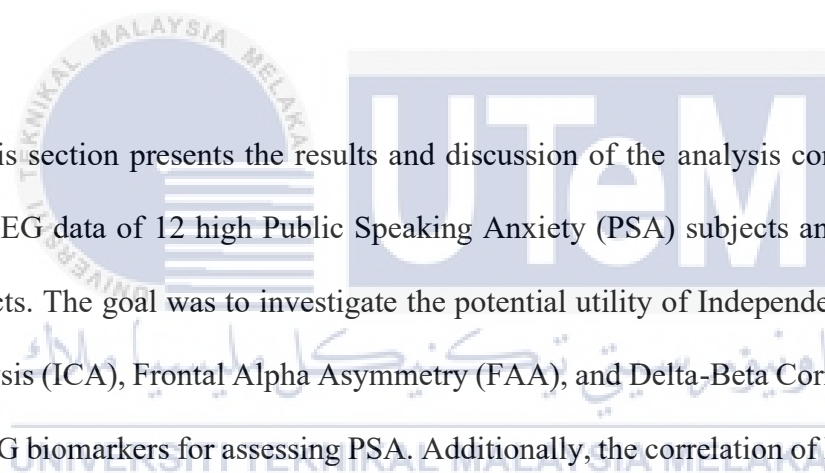
Aktiviti Projek / Project Activities	|				SEM I																SEM BREAK					SEM II														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Literature Review	X	X	X	X	X	X	X	X	X	X	X	X	X												X	X	X	X	X	X	X	X	X	X	X					
EEG data pre-processing (ICA)					X	X	X	X	X	X	X	X	X																											
Power Spectral Analysis								X	X	X	X																													
Frontal Alpha Asymmetry Analysis								X	X	X	X																													
Delta-Beta Correlation Analysis																									X	X	X	X	X	X										
Logistic Regression Machine Learning																												X	X	X	X	X	X	X	X					
Random Forest Machine Learning																												X	X	X	X	X	X	X	X					
Concluding Findings																																					X	X		
Publish Results																												X	X	X	X	X	X	X	X					
Thesis Writing					X	X	X	X	X	X	X	X	X							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	

(SEMINAR PSM I located at week 14 of SEM I; SEMINAR PSM II located at weeks 14–15 of SEM II)

The Gantt chart is divided into 10 activities and shown in Table 3.2. The Gantt chart is discussed in Appendix E.
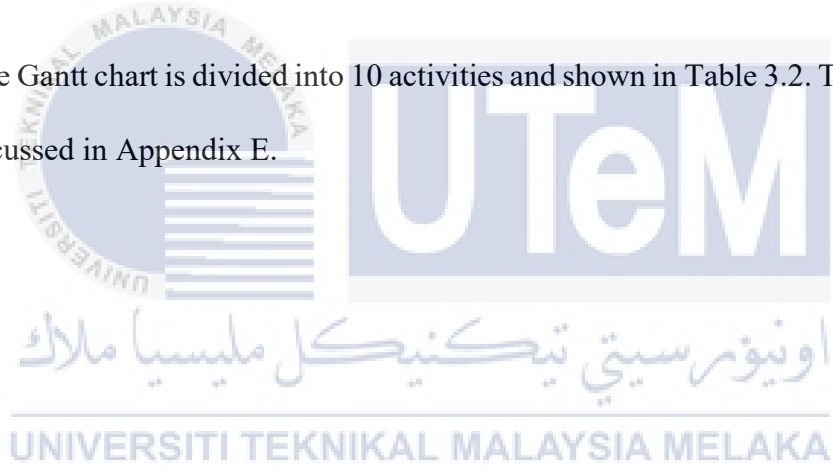
CHAPTER 4

RESULTS AND DISCUSSION

This section presents the results and discussion of the analysis conducted on the raw EEG data of 12 high Public Speaking Anxiety (PSA) subjects and 12 low PSA subjects. The goal was to investigate the potential utility of Independent Component Analysis (ICA), Frontal Alpha Asymmetry (FAA), and Delta-Beta Correlation (DBC) as EEG biomarkers for assessing PSA. Additionally, the correlation of FAA and DBC with reaction time was examined. The findings and implications of these analyses are discussed below.

4.1 Independent Component Analysis (ICA)

In this study, a total of 12 LPSA and 12 HPSA subjects' raw EEG data were utilized. To produce clean EEG data, Independent Component Analysis (ICA) was employed to remove artifacts. However, due to the potential lengthiness of describing the cleaning process for all raw EEG data in this thesis report, the focus will be on presenting the cleaning results for two LPSA subjects and two HPSA subjects.

Diagrams illustrating the EEG data before and after the application of ICA were included to visually demonstrate the cleaning process.

4.1.1 ICA of HPSA Subjects

The HPSA subjects presented for ICA in this thesis are subject 2 and subject 3.

Figure 4.1: EEG data of HPSA Subject 2 before ICA

The artifacts that exist in the F8 electrode of HPSA Subject 2 are shown in Figure 4.1.
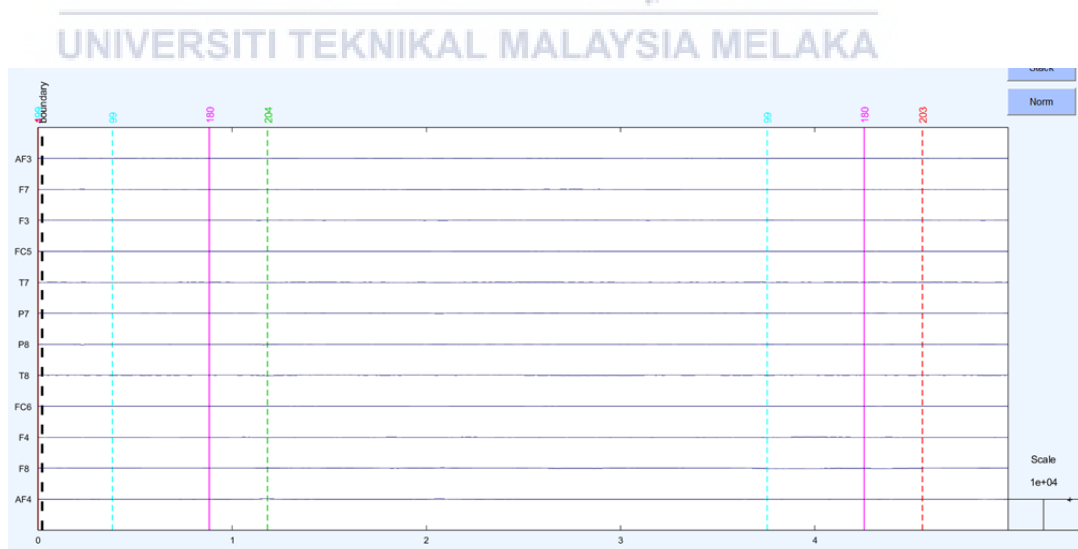
Figure 4.2: EEG data of HPSA Subject 2 after ICA

The artifacts were removed from the F8 electrode of HPSA Subject 2 and shown in Figure 4.2.
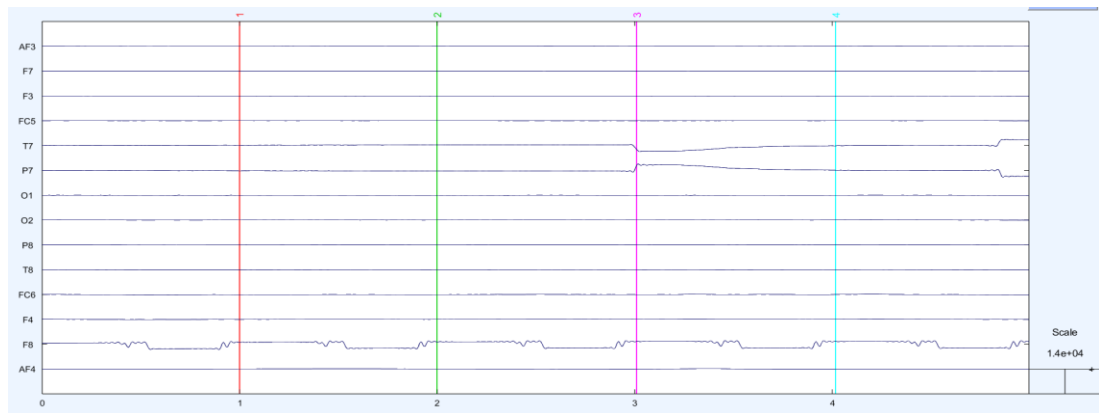
Figure 4.3: EEG data of HPSA Subject 3 before ICA

The artifacts that exist in the F8, T7 and P7 electrodes of HPSA Subject 3 are shown in Figure 4.3.
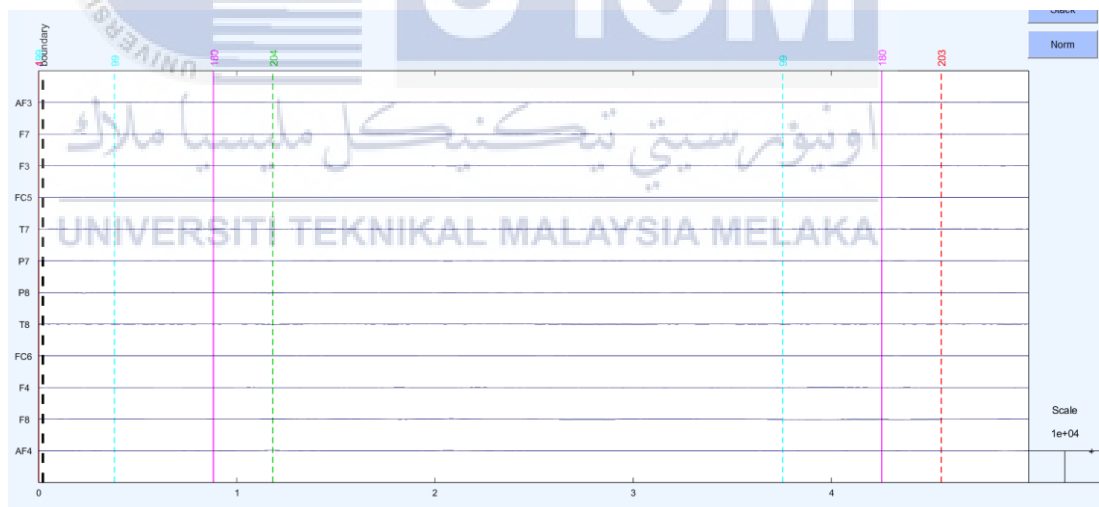
Figure 4.4: EEG data of HPSA Subject 3 after ICA

The artifacts were removed from F8, T7 and P7 electrodes of HPSA Subject 2 and shown in Figure 4.4.

4.1.2 ICA of LPSA Subjects

The LPSA subjects presented for ICA in this thesis are subject 4 and subject 10.
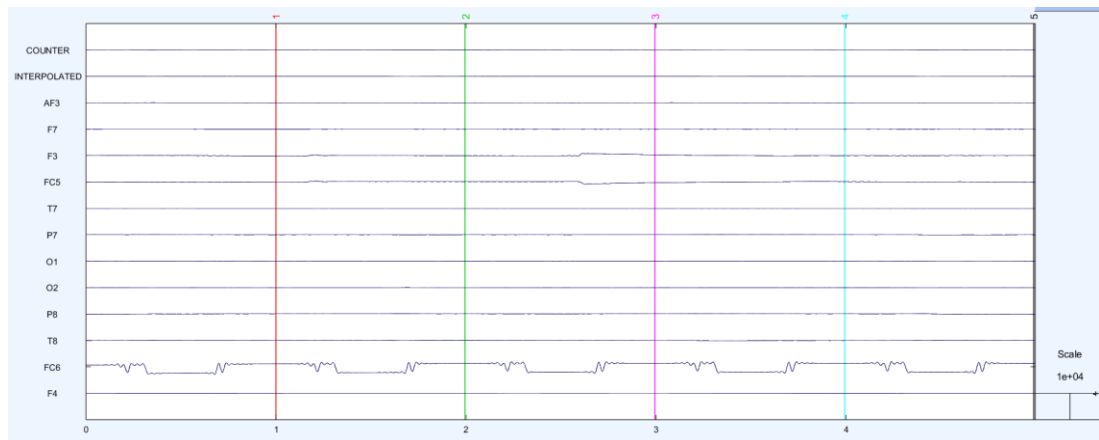
Figure 4.5: EEG data of LPSA Subject 4 before ICA

The artifacts that exist in the FC6 electrode of LPSA Subject 4 are shown in Figure 4.5.
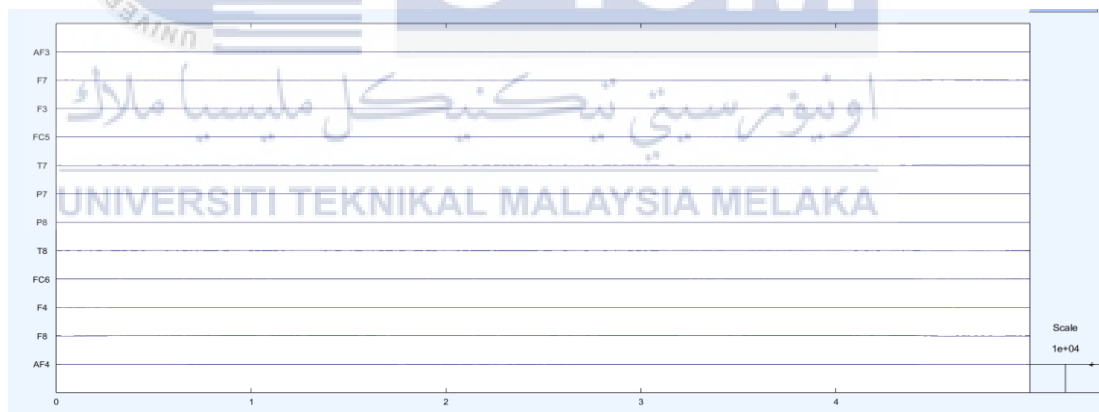
Figure 4.6: EEG data of LPSA Subject 4 after ICA

The artifacts were removed from the FC6 electrode of LPSA Subject 4 and shown in Figure 4.6.
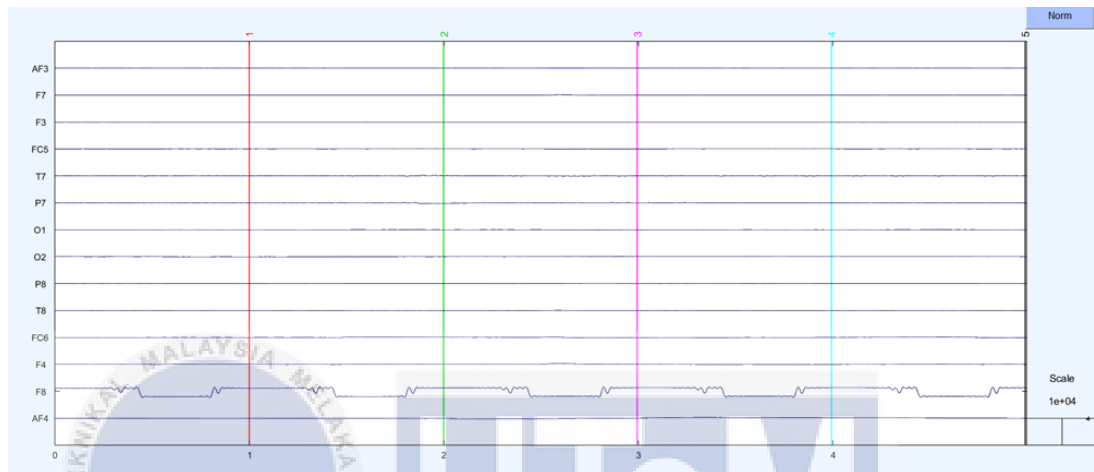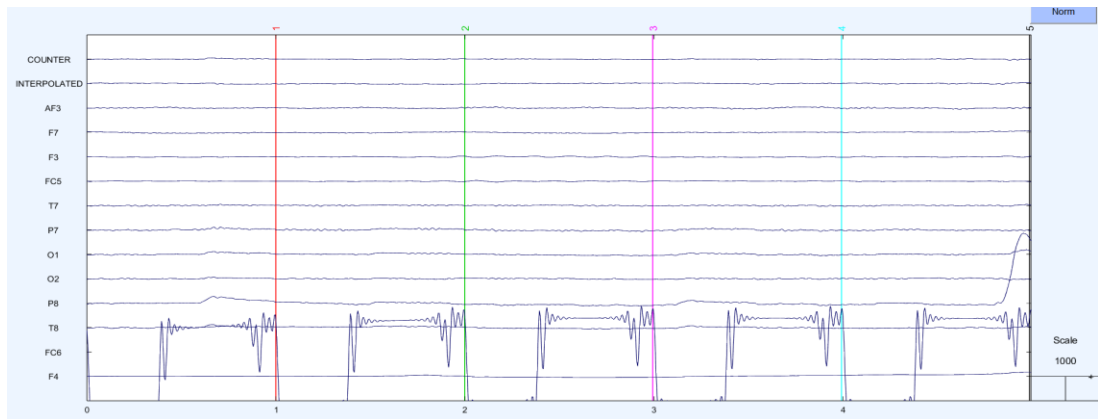
Figure 4.7: EEG data of LPSA Subject 10 before ICA

The artifacts that exist in the FC6 and P8 electrodes of LPSA Subject 10 are shown in Figure 4.7.
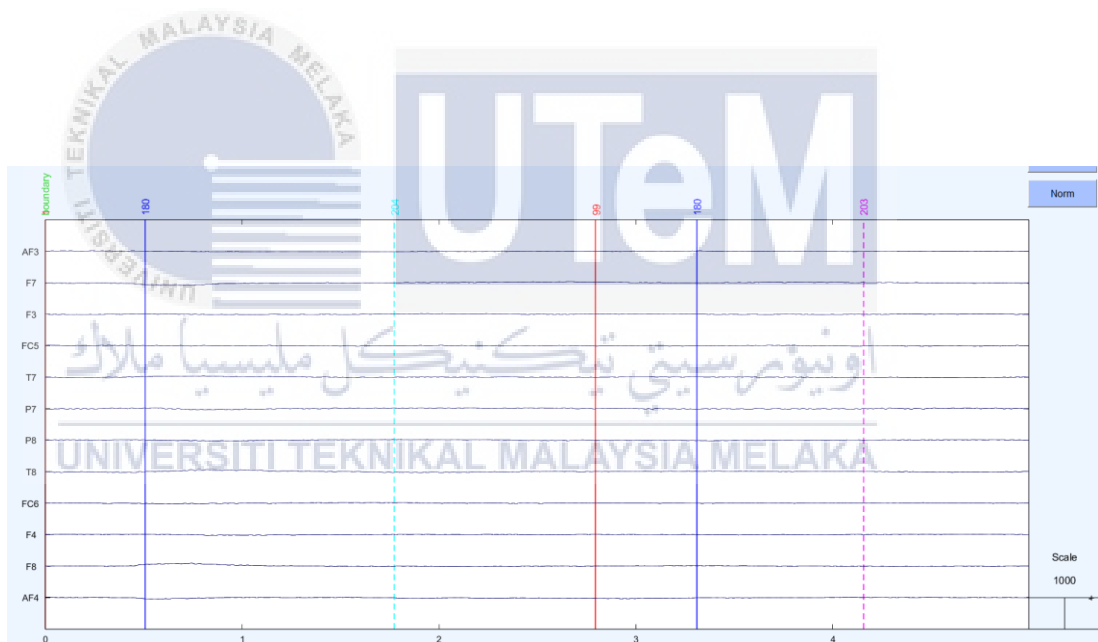
Figure 4.8: EEG data of LPSA Subject 10 after ICA

The artifacts were removed from FC6 and P8 electrodes of LPSA Subject 10 and shown in Figure 4.8.

4.2   Frontal Alpha Asymmetry (FAA)

In MATLAB, the FAA index of 12 HPSA and 12 LPSA subjects was computed and shown in Appendix F. The FAA index computed in Appendix F was further analysed using the Repeated Measures Analysis of Variance (ANOVA) test and shown in Figure 4.9.
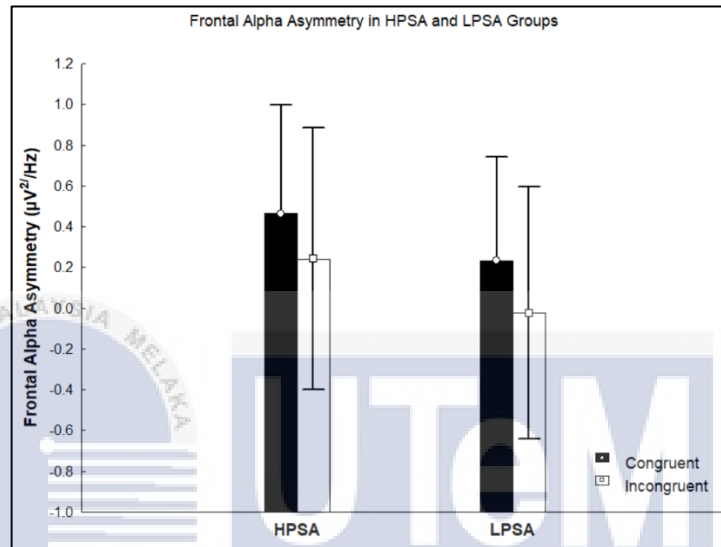
Figure 4.9: Bar Graph of FAA in HPSA and LPSA Groups

Table 4.1: Repeated Measures Analysis of Variance with Effects Sizes and Power (FAA)

Effect	SS	Degr. of Freedom	MS	F	p	Partial eta-squared	Non-centrality	Observed power (alpha=0.05)
	Repeated Measures Analysis of Variance with Effect Sizes and Powers (FAA) Sigma-restricted parameterization Effective hypothesis decomposition							
Intercept	2.62137	1	2.621369	1.940417	0.176942	0.077802	1.940417	0.266536
Group	0.76770	1	0.767701	0.568275	0.458594	0.024112	0.568275	0.111660
Error	31.07140	23	1.350931					
CON	0.71825	1	0.718250	1.207825	0.283135	0.049894	1.207825	0.183648
CON*Group	0.00309	1	0.003092	0.005200	0.943135	0.000226	0.005200	0.050548
Error	13.67727	23	0.594664					

The right FAA value of the HPSA group (Congruent Condition M= 0.465244 $\mu v^2/Hz$ SE= 0.256728 $\mu v^2/Hz$, Incongruent Condition M= 0.241085$\mu v^2/Hz$, SE= 0.310199 $\mu v^2/Hz$) is on average higher than the LPSA group (Congruent Condition M=

$0.232964 \; \mu v^2/Hz$ SE= $0.246656 \; \mu v^2/Hz$, Incongruent Condition M= -0.022678 $\mu v^2/Hz$, SE= $0.298030 \; \mu v^2/Hz$) in both congruent and incongruent conditions for the Stroop Task. A higher right FAA indicates heightened negative emotions, such as PSA in this context [31]. The present findings seem to be consistent with another research which found higher right FAA at the F3 and F4 electrodes of patients with schizophrenia compared to healthy controls [32]. However, the mixed-design ANOVA for FAA shown in Table 4.1 revealed no significant differences for any main effects or interactions across conditions [F (1, 23) = 0.005200, partial η^2 = 0.000226, p = 0.943135]. The interpretation of this finding is possibly due to the limited sample size of PSA individuals.

4.3   Delta-Beta Correlation (DBC)

The delta and beta band power were used to conduct the DBC test for the Frontal, Central, Parietal and Temporal electrodes for all participants in this study. The Delta-Beta Band values for distinct electrodes were delineated in the appendices of this study. Specifically, Appendix G detailed the Delta-Beta Band value associated with the Frontal Electrode, while Appendix H presented the corresponding value for the Central Electrode. Furthermore, Appendices I and J respectively illustrated the Delta-Beta Band values pertaining to the Parietal and Temporal Electrode. The study employed Pearson correlation analysis to investigate the correlation between delta (0-4 Hz) and beta (13-30 Hz) waves. Notably, a total of five positive and statistically significant delta-beta correlations (DBC) were identified across frontal, parietal, and central electrodes.

4.3.1   DBC of HPSA within the Frontal region in the Incongruent Condition

Figure 4.10: Scatterplot of Delta-Beta Correlation in the Frontal Region for HPSA Individuals under Incongruent Conditions

There was a significant positive linear DBC (r = 0.6931, p = 0.12(uncorrected)) in the frontal region (F3 and F4 electrodes) of the brain in the incongruent condition for HPSA subjects. The scatterplot of DBC in the frontal region for HPSA individuals under incongruent conditions is shown in Figure 4.10. Interestingly, the effect was not found in LPSA subjects (r=-.0127, p=.967). Comparing both the results of the HPSA and LPSA groups, there may be differential patterns of brainwave activity among the two groups. It can be seen that DBC in LPSA is indicated by a very weak negative correlation as the correlation coefficient, *r* is between 0.0 to -0.2 and the p-value is more than 0.5, meanwhile, the significant positive DBC observed in the frontal region of HPSA subjects during the incongruent condition may imply a specific neural processing mechanism related to PSA responses in these individuals. The present findings seem to be consistent with other research which found positive DBC in the frontal region in anxious subjects [24]. Heightened delta-beta correlation has been conceptualized as reflecting exaggerated neural regulation and has been implicated in anxiety[5].

4.3.2   DBC of HPSA in Incongruent Condition at the Parietal region
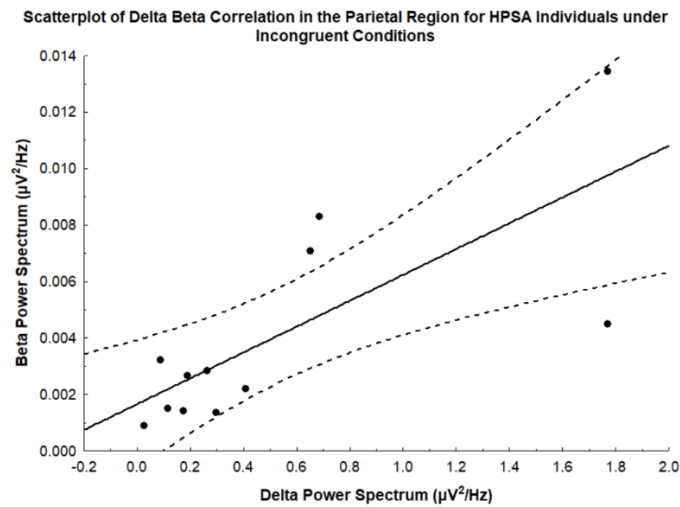
Figure 4.11: Scatterplot of Delta-Beta Correlation in the Parietal Region for HPSA Individuals under Incongruent Conditions

The study revealed a noteworthy contrast in DBC patterns between HPSA and LPSA subjects in the incongruent condition. Specifically, in the parietal region (P7 and P8 electrodes), there was a significant positive linear DBC in HPSA subjects ($r= .7477$, $p=.005$ (uncorrected)), indicating a strong positive correlation between delta and beta waves. Intriguingly, this effect was not observed in LPSA subjects ($r= .1410$, $p=.646$), where only a very weak positive correlation in DBC was evident. This discrepancy in DBC patterns between the two groups suggests the existence of differential brainwave activity among individuals with varying levels of performance and PSA. The substantial positive DBC observed in the parietal region of HPSA subjects during incongruent conditions may point to a specific neural processing mechanism associated with their responses to PSA. These findings align with prior research that identified positive DBC in the parietal region in anxious subjects [5]. The concept of heightened delta-beta correlation, often interpreted as reflecting exaggerated neural regulation, lends support to the notion that the observed DBC patterns could be implicated in anxiety. This suggests a potential neural basis for the

heightened performance social anxiety experienced by HPSA subjects, as indicated by the significant positive DBC in the parietal region during incongruent conditions.

4.3.3 Positive DBC in HPSA and LPSA subjects in Congruent Condition at the Parietal brain region
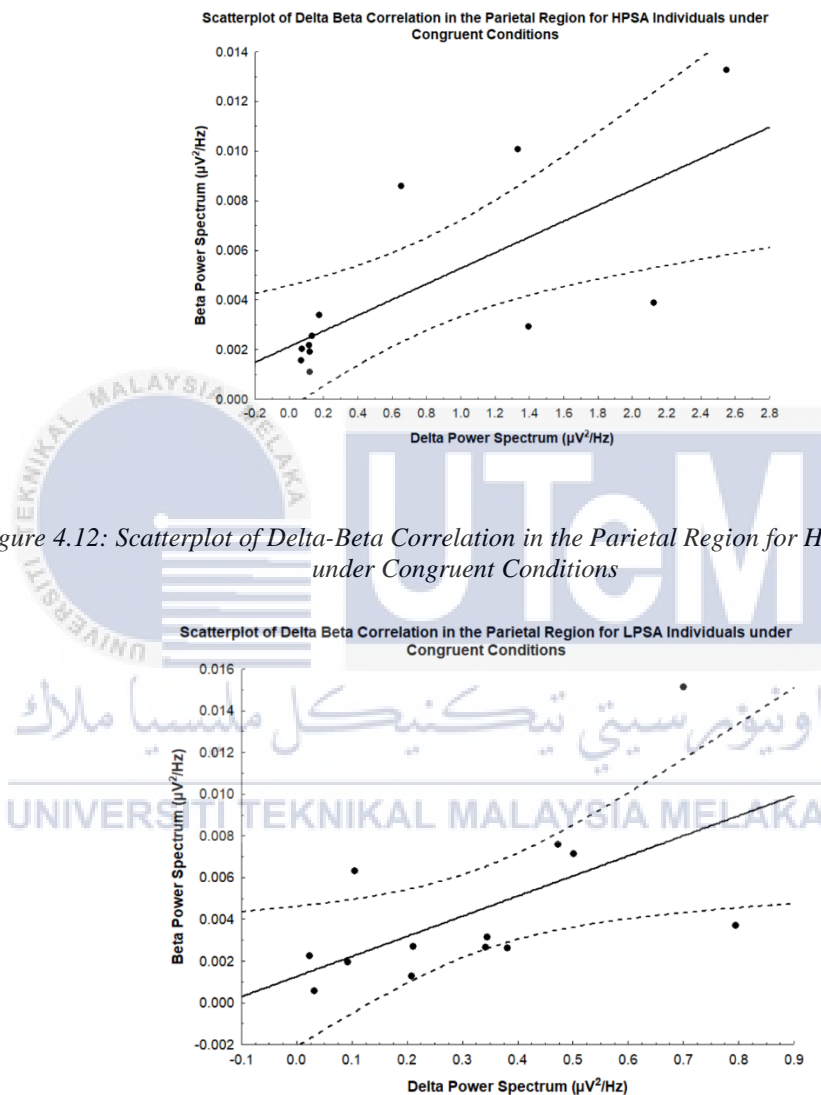
Figure 4.12: Scatterplot of Delta-Beta Correlation in the Parietal Region for HPSA Individuals under Congruent Conditions

Figure 4.13: Scatterplot of Delta-Beta Correlation in the Parietal Region for LPSA Individuals under Congruent Conditions

A significant positive linear DBC was identified in the parietal region (P7 and P8 electrodes) of the brain during the congruent condition for individuals with HPSA (r= .7130, p=.009 (uncorrected)). Interestingly, this effect was also present in individuals with LPSA (r= .6069, p=.028(uncorrected)). The comparison of these

results underscores a consistent positive DBC for both HPSA and LPSA groups, suggesting that the parietal region's activity during the congruent condition serves as a robust predictor of PSA outcomes, regardless of the individual's anxiety level.
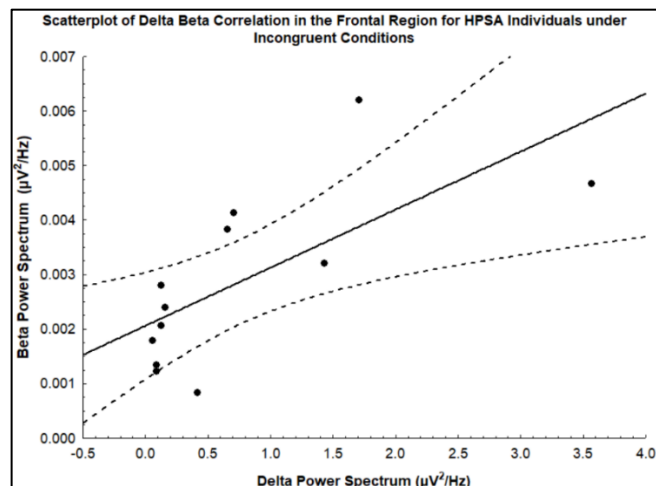
4.3.4 DBC of LPSA in Central brain region in the Incongruent Condition
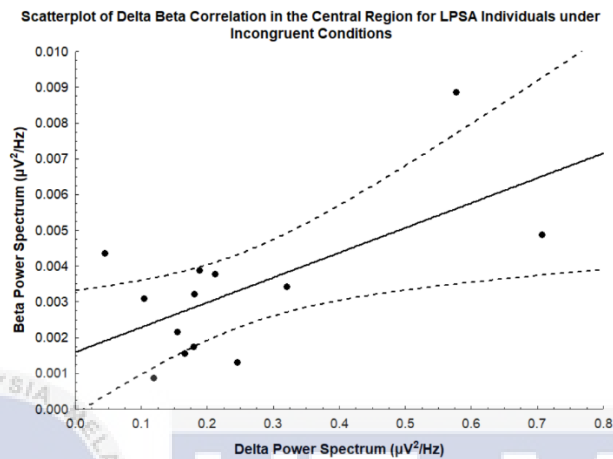
Figure 4.14: Scatterplot of Delta-Beta Correlation in the Central Region for LPSA Individuals under Incongruent Conditions

A noteworthy observation emerged as a significant positive linear DBC was detected in the central region (FC5 and FC6 electrodes) of the brain during the incongruent condition for individuals with LPSA (r= .6348, p=.020 (uncorrected)). Intriguingly, this effect was not present in individuals with HPSA (r= .1410, p=.646 (uncorrected)). A comparison of these results suggests potential differential patterns of brainwave activity between the two groups. Notably, the DBC in HPSA subjects is characterized by a very weak positive correlation, while the significant positive DBC observed in the central region of LPSA subjects during the incongruent condition may indicate a specific neural processing mechanism related to public speaking anxiety responses in individuals with low PSA.

4.3.5 Summary of DBC

The computed DBC is summarized in Appendix K.

Brain regions that have strong positive statistically significant DBC are:

- Frontal(F3 and F4) and Parietal (P7 and P8) regions in incongruent condition for HPSA subjects,

- Central region (FC5 and FC6) in incongruent condition for LPSA subjects,

- Parietal region (P7 and P8) in congruent condition for both HPSA and LPSA subjects,

4.4 Machine Learning

4.4.1 Logistic Regression

The reasons for incorporating K-fold cross-validation and L1 regularization process used for the LR model have been discussed in Chapter 3. The LR model was optimized on an L1 regularization basis in each of the four folds. Convergence is a stability indicator of the coefficient estimation process within the optimization process. The model includes prediction variables such as FC5_Con (Amplitude of FC5 electrode in Congruent condition), RT_Con (Reaction time in Congruent condition) and Error_Con (Error rate in Congruent condition), among others.

4.4.1.1 The Summary of the LR Model

The outlines for the LR models are presented in Figures 4.15–4.18.

Figure 4.15: Logistic Regression Model Summary of 1ˢᵗ Fold Cross Validation

Figure 4.16: Logistic Regression Model Summary of 2ⁿᵈ Fold Cross Validation

```
kth-fold= 3
Optimization terminated successfully    (Exit mode 0)
         Current function value: 0.421071497684464
         Iterations: 21
         Function evaluations: 31
         Gradient evaluations: 21
                      Logit Regression Results
==============================================================================
Dep. Variable:                  Group   No. Observations:                   12
Model:                          Logit   Df Residuals:                        6
Method:                           MLE   Df Model:                            5
Date:                Fri, 08 Dec 2023   Pseudo R-squ.:                  0.6517
Time:                        10:54:33   Log-Likelihood:                -2.8384
converged:                       True   LL-Null:                       -8.1503
Covariance Type:            nonrobust   LLR p-value:                   0.05937
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
FC5_Con             0        nan        nan        nan         nan         nan
RT_Con         0.0006      0.004      0.157      0.875      -0.006       0.007
Error_Incon   -0.3388      0.673     -0.503      0.615      -1.658       0.981
State          0.3526      0.244      1.445      0.148      -0.126       0.831
Trait          0.0550      0.154      0.357      0.721      -0.247       0.357
Age           -0.3510      0.271     -1.296      0.195      -0.882       0.180
sigma         -0.0093      0.031     -0.299      0.765      -0.070       0.052
==============================================================================
```

Figure 4.17: Logistic Regression Model Summary of 3ʳᵈ Fold Cross Validation

```
kth-fold= 4
Optimization terminated successfully    (Exit mode 0)
         Current function value: 0.4342224170197067
         Iterations: 18
         Function evaluations: 29
         Gradient evaluations: 18
                      Logit Regression Results
==============================================================================
Dep. Variable:                  Group   No. Observations:                   12
Model:                          Logit   Df Residuals:                        6
Method:                           MLE   Df Model:                            5
Date:                Fri, 08 Dec 2023   Pseudo R-squ.:                  0.6629
Time:                        10:54:33   Log-Likelihood:                -2.7476
converged:                       True   LL-Null:                       -8.1503
Covariance Type:            nonrobust   LLR p-value:                   0.05538
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
FC5_Con             0        nan        nan        nan         nan         nan
RT_Con         0.0056      0.004      1.244      0.214      -0.003       0.014
Error_Incon   -0.5440      0.533     -1.020      0.308      -1.589       0.501
State          0.2910      0.250      1.163      0.245      -0.200       0.782
Trait         -0.0953      0.192     -0.497      0.619      -0.471       0.281
Age           -0.2676      0.281     -0.952      0.341      -0.819       0.284
sigma         -0.0281      0.040     -0.704      0.481      -0.106       0.050
==============================================================================
```

Figure 4.18: Logistic Regression Model Summary of 4ᵗʰ Fold Cross Validation

Each coefficient sheds some light on the extent to which the independent variable (Group) influences the dependent variable's log odds. Certainly, for example, some coefficients, among them "Error_Con" and "FC5_Con" reach zero. The fact that the coefficient converges to zero means that for these specific cases, the associated variable has little to no impact on determining the outcome. However, while this may help to simplify the model and improve its interpretability, one must understand that

such variables may not be that explainable in those particular folds. Knowing how particular parameters affected the predictor in each fold increases our knowledge about the meaning of logistic regression outcomes. The information concerning the impact of individual parameters on the predictor in each cycle enlarges our understanding of the sense of logistic regression results.

4.4.1.2 Measured Results of LR Model
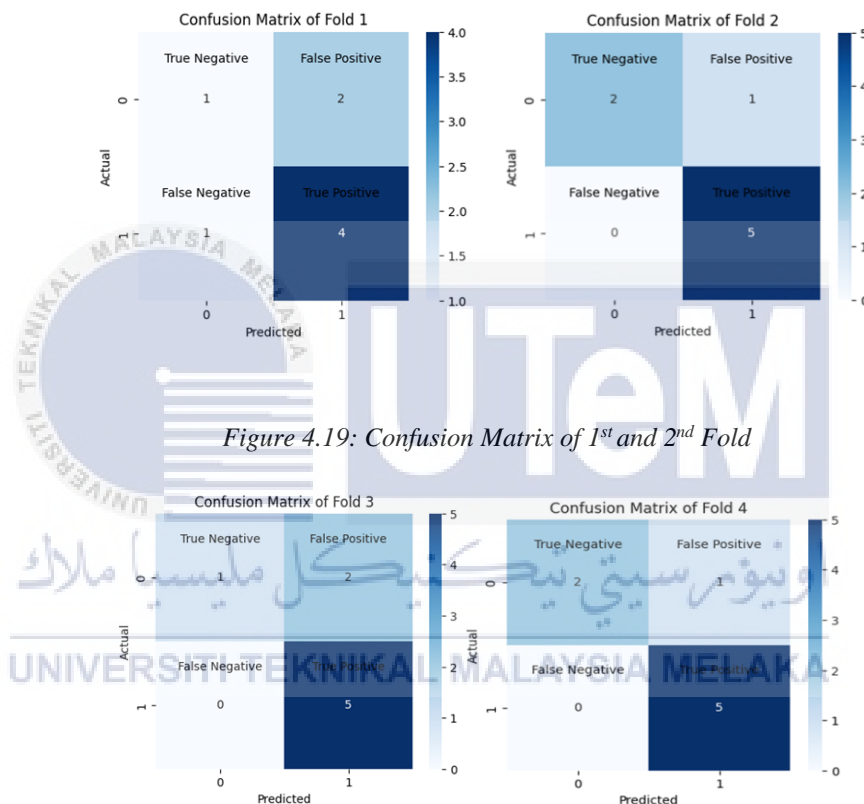
Figure 4.19: Confusion Matrix of 1st and 2nd Fold
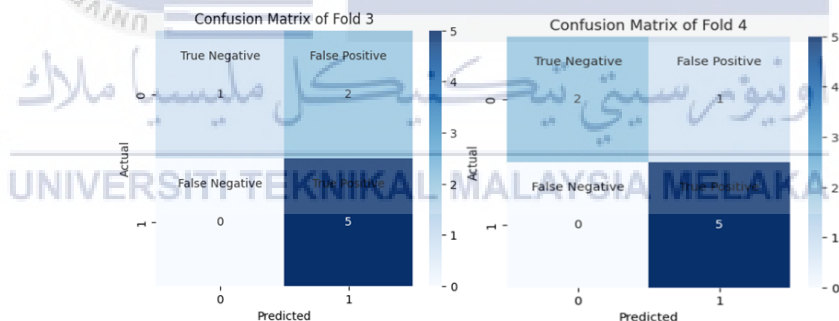
Figure 4.20: Confusion Matrix of 3rd and 4th Fold

Table 4.2: Summaries of the Confusion Matrix

Fold	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
1	4	1	2	1
2	5	2	1	0
3	5	1	2	0
4	5	2	1	0

Detailed results of the confusion matrix are depicted in Table 4.2. TP are quite high across all folds (1 through 4) for every fold ensuring that the model has identified accurately points there were present with PSA. Values of 1 or 2 are exhibited by TN in different folds. This implies that the model's success in recognizing individuals with LPSA fluctuates between folds. It is consistent across folds in the number of FPs and FNs as well. In every fold, 1-2 FP and 0-1 FN suggest some level of stability with respect to misclassification errors.

Figure 4.21: Accuracy, Precision, Specificity, Sensitivity and their average value of Every Fold

The Accuracy, Precision, Specificity, Sensitivity and average value of every fold cross-validation information is summarized in Table 4.3 below.

Table 4.3: Summary Result of Every Fold

Fold	Accuracy	Precision	Specificity	Sensitivity (Recall)
1	0.6250	0.6667	0.333	0.80
2	0.8750	0.8333	0.667	1.00
3	0.7500	0.7143	0.333	1.00
4	0.8750	0.8333	0.667	1.00
Average	0.7812	0.7619	0.500	0.95

Fold-level metrics created from various performance measures make an understanding of the category model in a particular set of data. The total prediction accuracy across the folds differs, however. In fold-2 and fold-4, the best accuracy achieved was 0.875 while it was 0.625 for fold-1 and it was 0.75 for fold-3. The precision measures in positive predictive values are also roughly fair and constant across folds, between 0.667-0.833.

For the negatives in folds 1 and 2, the specificity difference to the model was 0.333%. Although every fold is sensitive to a different extent, the general sensitivity lies anywhere between 0.80 and 1.00. The consistency shown in this regard also demonstrates that the model can identify instances of the positive set across any partitioning of the data. Such variables show how well the modelling process identifies true positives as well as true negatives. Differences in specificity and accuracies across folds may well serve as signs of parameters that would help improve the overall model performance.

4.4.1.3 Visualized Result

The ROC Curves for each fold are shown from Figure 4.22 to Figure 4.25.
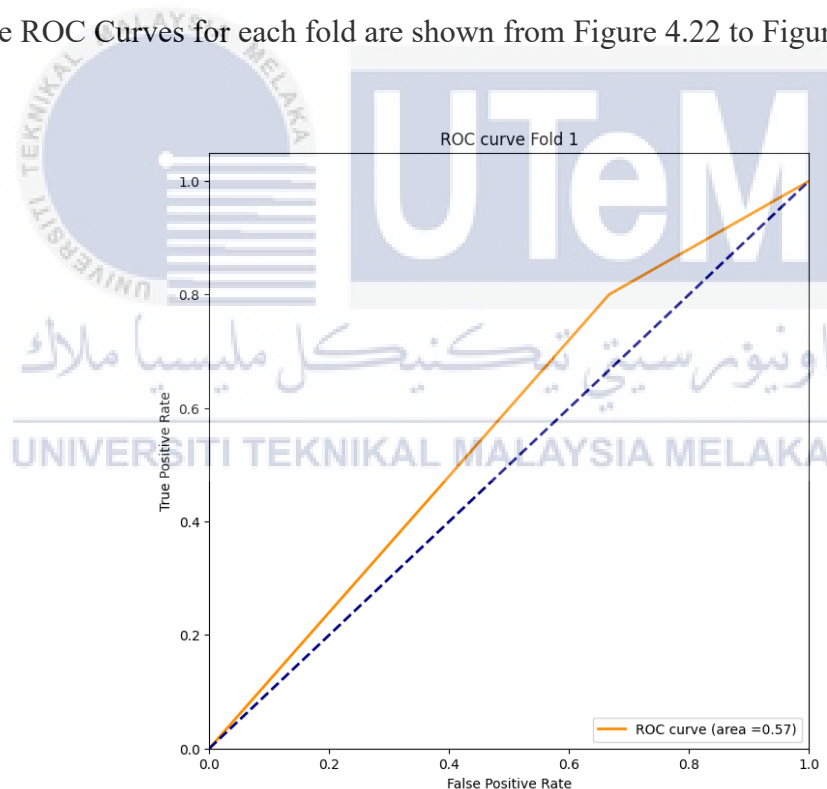
Figure 4.22: ROC Curve 1st Fold
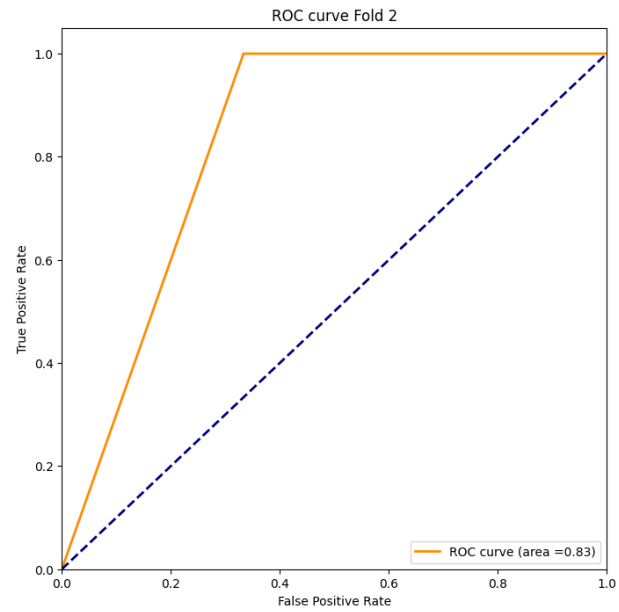
Figure 4.23: ROC Curve 2nd Fold
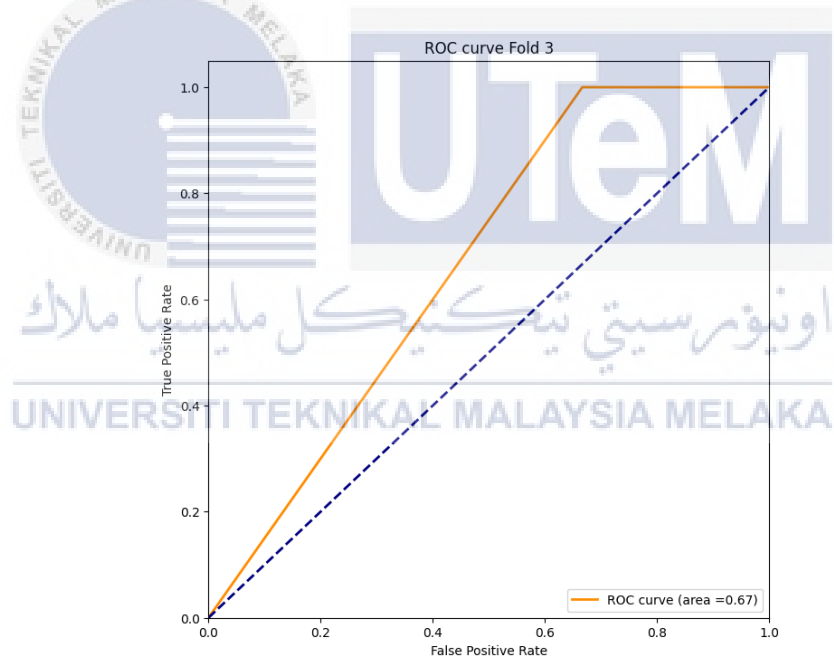
Figure 4.24: ROC Curve 3rd Fold
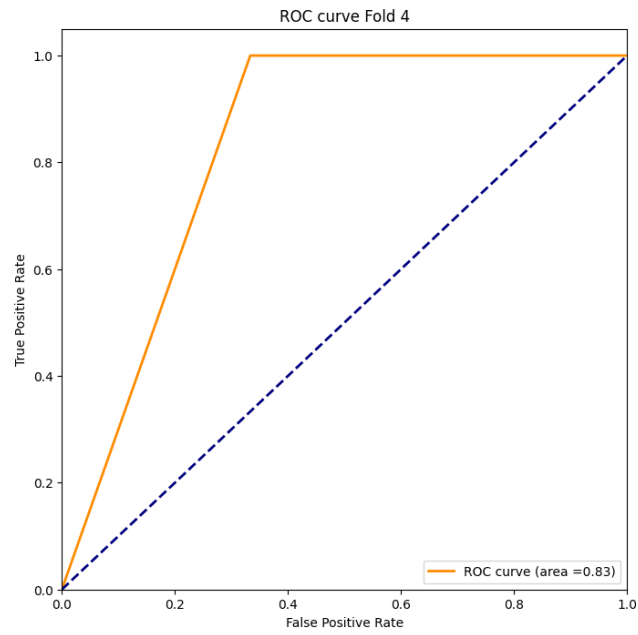
Figure 4.25: ROC Curve 4ᵗʰ Fold
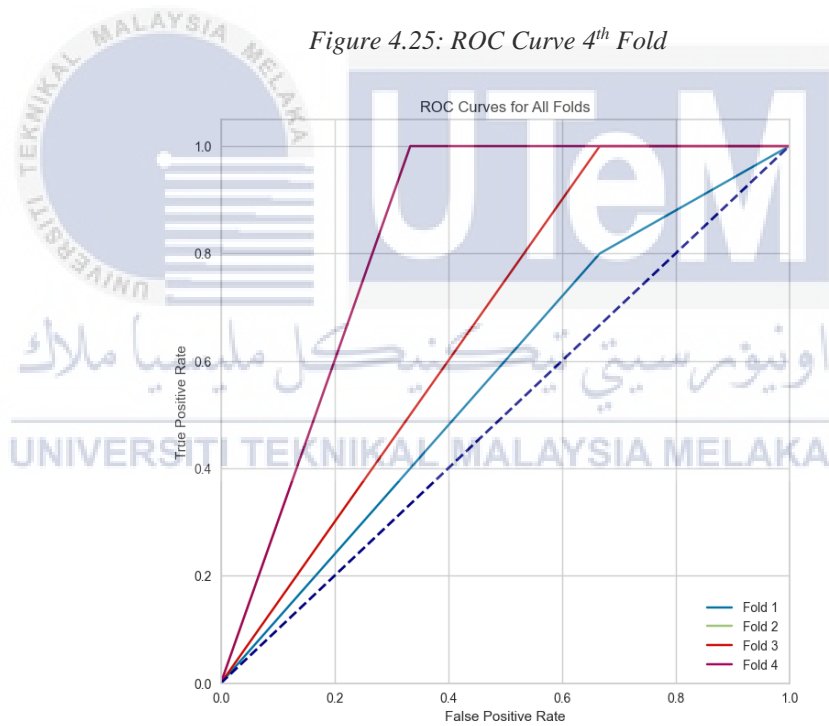
Figure 4.26: ROC curve of every fold

Table 4.4: Summary of ROC Curve

Fold	ROC Curve Value /Area Under ROC Curve (AUC)	Interpretation
1	0.57	The AUC of 0.57 suggests a relatively poor discrimination performance.
2	0.83	The AUC of 0.80 suggests a good discrimination performance.
3	0.67	The AUC of 0.67 suggests a reasonably good discrimination performance.
4	0.83	The AUC of 0.80 suggests a good discrimination performance

The ROC curve is summarized and analysed in Table 4.4. Folds 2 and 4 from Figure 4.23 and Figure 4.25 were particularly strong, with each having an AUC of 0.83, which is exceptionally strong and consistent in the ability to differentiate among the classes. These are solid folds that indicate the performance of the model being reliable and effective in all layers.

Fold 3 (refer Figure 4.24), though slightly lower AUC of 0.67, has only shown a slight decrease wherein it still exhibits a good discrimination capacity and ability. The performance as well is good in this fold and thus the observed AUC in this case concludes the fold present determines that the model has the power to determine underlying patterns present with the data.

In sharp contrast, Fold 1 (refer Figure 4.22) emerges with an AUC of 0.57 signalling a relatively lower discrimination capability. Such kind of a discrepancy in
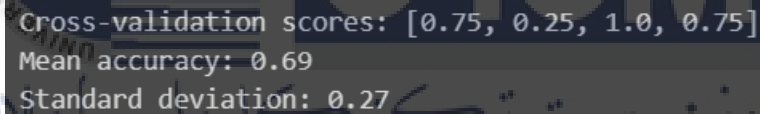
performance could mean the data within this particular partition has presented unique characteristics that set it apart or challenge the predictive model. It will improve the discriminatory power of features, model responses, or potential sources of variation in this particular fold that could be further explored are identified. Typically gaining a nuanced understanding would allow getting a well-rounded view of how the model performs and generalizes across folds from correlated features.

4.4.2   Random Forest (RF)

As discussed in Chapter 3, the RF model is optimized using 4-fold cross-validation.

4.4.2.1 Metrics Computed From 4-Fold Cross-Validation of RF

The cross-validation scores (accuracy), mean of accuracy and standard deviation are computed using the RF model were displayed in Figure 4.27.

```
Cross-validation scores: [0.75, 0.25, 1.0, 0.75]
Mean accuracy: 0.69
Standard deviation: 0.27
```

Figure 4.27: Metrics Computed from RF Model

The cross-validation scores, [0.75, 0.25, 1.0, 0.75], underscore the model's consistent performance across diverse folds, indicating a high level of stability. This stability implies that the model generalizes effectively, exhibiting resilience to variations in training and validation subsets. The reported mean accuracy of 0.69 reflects the model's overall ability to correctly classify diverse subsets of the data.

The accompanying standard deviation of 0.27 provides valuable insights into the variance of cross-validated scores around the mean accuracy. A lower standard

deviation, as observed here, signifies greater consistency across different folds, a positive characteristic for model generalizability. In contrast, a higher standard deviation would suggest heightened sensitivity to specific instances in the training dataset, potentially compromising the model's stability.

To assess the model's robustness, it is crucial to scrutinize whether the observed standard deviation aligns with expectations for reliable generalization. In this case, the standard deviation of 0.27 prompts a thoughtful examination of the model's ability to consistently perform well across varying data distributions. Hence, while the model exhibits commendable stability, further consideration may be warranted to determine if the standard deviation meets acceptable thresholds for robust generalization.

4.4.2.2 Confusion Matrix of Testing Data
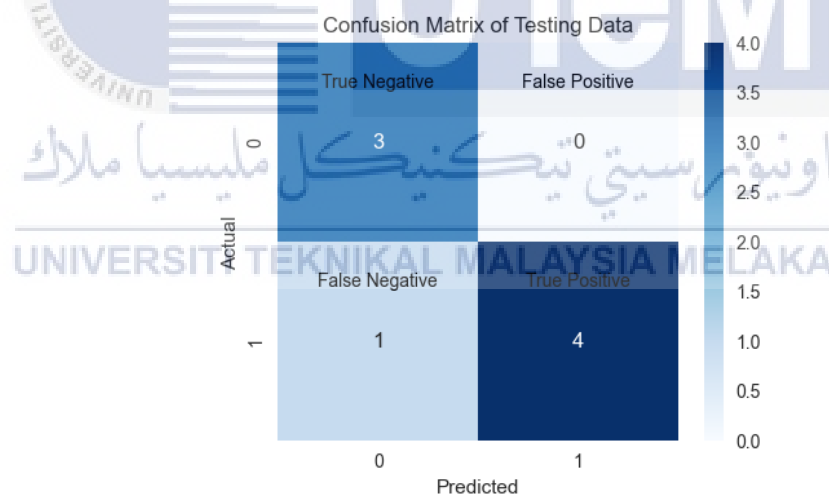
Figure 4.28: Confusion Matrix of Random Forest Testing Data

The confusion matrix presented in Figure 4.28 illustrates the performance of the RF model as follows: there are 3 instances correctly classified as negative (True Negatives), 1 instance misclassified as negative when it was actually positive (False Positive), and 4 instances correctly identified as positive (True Positives). In summary,

the model demonstrates 3 true negative predictions, 4 true positive predictions, and 1 false positive prediction.

4.4.2.3 Classification Report of Testing Data

The training data is fitted into the RF model without the k-fold cross-validation and the testing data was utilized for the prediction as unseen data of the model. The classification report is shown in Figure 4.29.


```
classification Report (Testing Data):
              precision    recall  f1-score   support

           0       0.75      1.00      0.86         3
           1       1.00      0.80      0.89         5

    accuracy                           0.88         8
   macro avg       0.88      0.90      0.87         8
weighted avg       0.91      0.88      0.88         8
```

Figure 4.29: Classification Report of Testing Data

The classification report offers valuable insights into the model's performance on individual classes, as well as general metrics for the test data. For Class 0, the precision stands at 0.75, indicating that when the model predicts this class, it is correct 75% of the time. The recall for Class 0 is a perfect 1.00, signifying that the RF model successfully captures all actual negative instances. The corresponding F1-score is 0.86.

On the other hand, Class 1 exhibits more favourable metrics, with a precision of 1.00, denoting that all positive predictions are accurate. The recall for Class 1 is 0.80, signifying that the model correctly identifies 80% of actual instances of this class. The F1-score for Class 1 is 0.89.

The overall accuracy of the model is reported as 0.88, implying that it correctly classifies 88% of instances in the testing data. The reported metrics collectively showcase the model's effectiveness and reliability in making accurate predictions across diverse instances in the testing data.

4.4.2.4 Visualization of Training Data and Testing Data

Receiver Operating Characteristics and Learning Curve are plotted to visualize the RF model that is shown in Figure 4.30 and Figure 4.31.
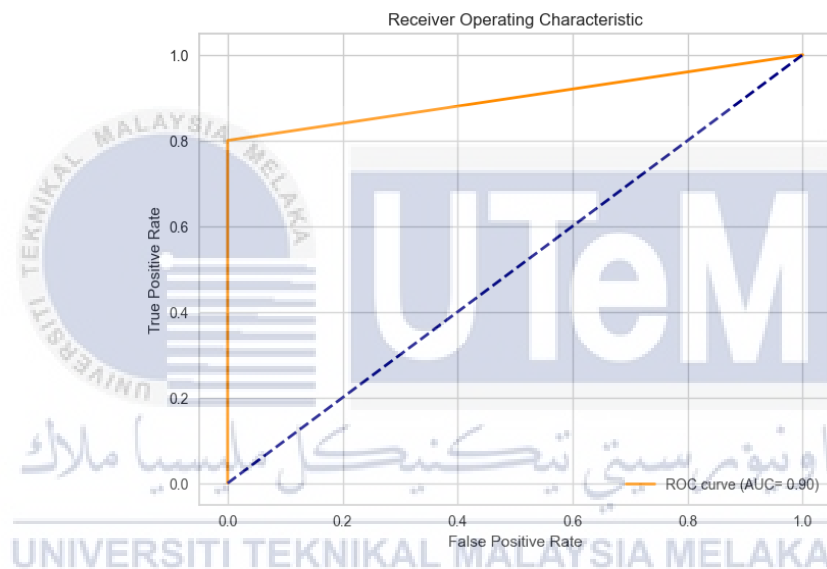
Figure 4.30: ROC Curve of RF model

ROC analysis of the RF model resulted in a noteworthy AUC-ROC value of 0.9, indicating its excellent ability to discriminate between positive and negative instances. The straightforward trajectory from (0,0) to (0,0.8) to (1,1) on the ROC curve underscores its strong performance and reliability in accurately classifying true positives while minimizing false positives. This signifies the model's potential as a dependable tool in practical applications.
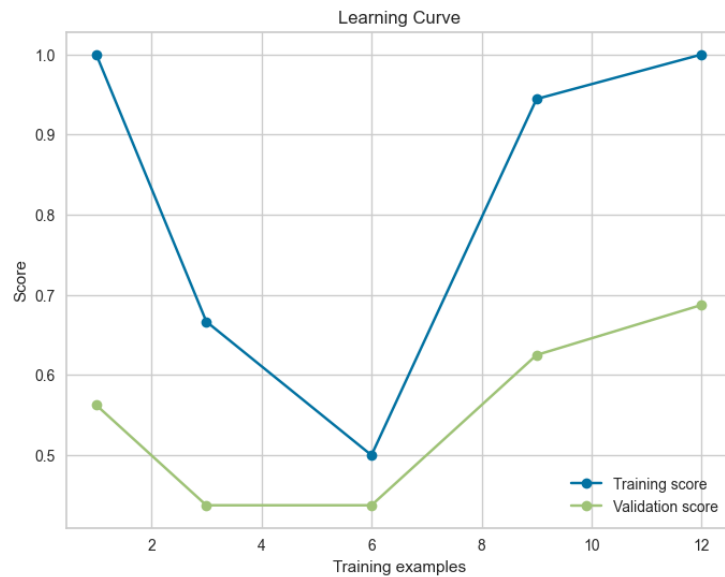
Figure 4.31: Learning Curve

The training score of the RF model demonstrates a distinctive pattern, initially decreasing from an initial value and then steadily increasing over successive training periods. This initial decrease suggests an early phase of learning where the model refines its predictions and adjusts to the complexities of the data. Subsequently, the score shows a consistent upward trajectory, indicating continuous improvement and effective learning and adaptation by the RF model.

In contrast, the validation score of the RF model exhibits fluctuations across different training intervals. Starting at a certain level, there is a subsequent decline, followed by variations and an eventual upward trend. These patterns in the validation score reflect the RF model's varying degrees of generalization during different phases of training. The fluctuations underscore the model's sensitivity to different training intervals, pointing towards potential opportunities for refinement to enhance its overall generalization across diverse datasets.

4.4.2.5 Feature Importance of RF

There is a total of 8 independent variables were used for the RF model training, which are amplitude of FC5 under congruent conditions (FC5_Con), Reaction Time under congruent conditions (RT_Con), Error Rate of the Stroop Task under congruent condition (Error_Con), Error rate of the Stroop Task in the incongruent condition (Error_Incon), State Anxiety, Trait Anxiety, Age and Sigma (of the Reaction Time obtained from ex-Gaussian Analysis). The feature importance is computed in Figure 4.32 and plotted as a bar chart in Figure 4.33. The feature importance is summarized in Appendix M.


```
Feature: ['FC5_Con', 'RT_Con', 'Error_Con', 'Error_Incon', 'State', 'Trait', 'Age', 'sigma']
State: 0.23441485068603712
Error_Con: 0.17296586044452875
Error_Incon: 0.12113539891788479
RT_Con: 0.10528830963665085
Age: 0.1022010174986039
FC5_Con: 0.09774407174746158
Trait: 0.09291715773548317
sigma: 0.07333333333333333
```
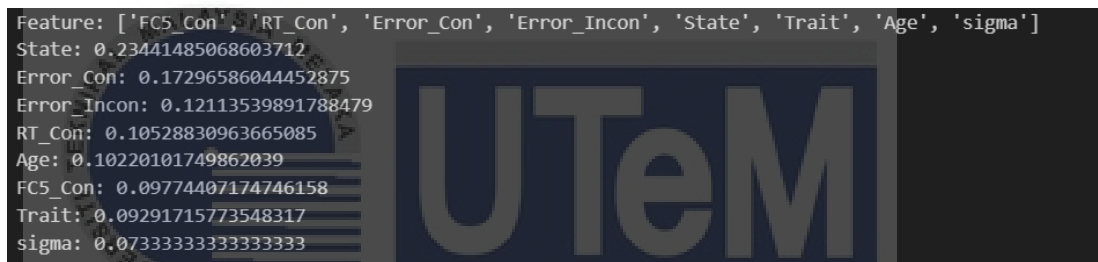
Figure 4.32: Feature Importance of RF Model
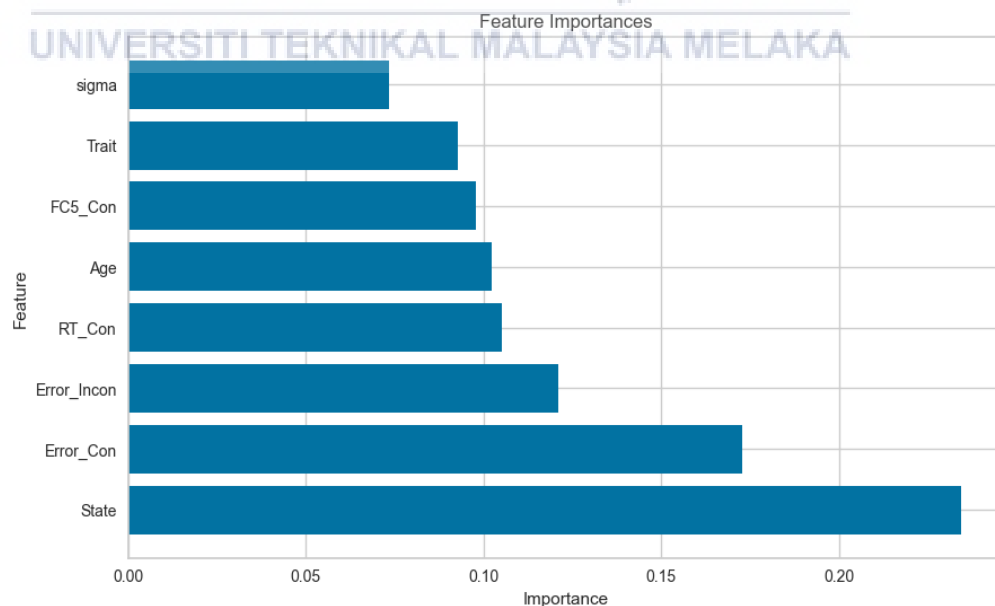
Figure 4.33: Bar Chart of Feature Importance of RF Model

These importance scores offer a quantitative measure of each feature's impact on the model's decision-making process. Understanding feature importance is crucial for interpreting the model's behavior and may guide further investigations or refinements in the model, such as feature engineering or selection, to optimize predictive performance. It's worth noting that the interpretation of feature importance is context-dependent, and domain knowledge should be considered for a comprehensive understanding of the model's behavior.

4.4.2.6 Visualization of Tree in RF model
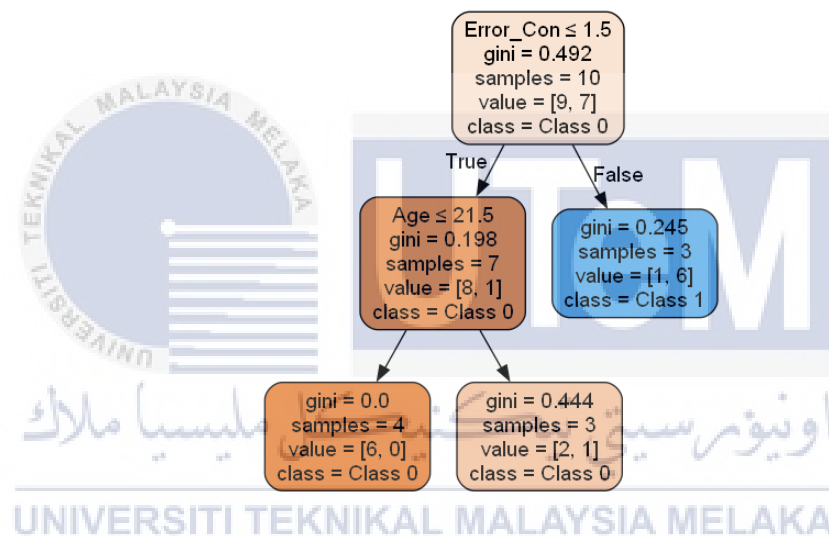
Figure 4.34: One of the trees in the RF model

The orange colour indicates the majority prediction of Class 0 (LPSA individuals) and the blue colour indicates the majority prediction of Class 1(HPSA individuals). Gini impurity is a metric that quantifies the likelihood of misclassifying an instance randomly chosen from the dataset. A more intense colouring designates a lower Gini impurity, which in turn indicates a lower likelihood of misclassification. In the RF model, the root node serves as the starting point for the decision-making process. This node is determined by the condition "Error_Con $\leq$ 1.5." The Gini impurity for this

node is 0.492, indicating the level of impurity in the node's samples. There are a total of 10 samples in this node, with a class distribution of [9 Class 0, 7 Class 1]. As the Gini impurity is lower, the predicted class for this node is Class 0. Moving to the branches, the "True" branch emanates from the root node and is based on the condition "Age $\leqslant$ 21.5." This branch exhibits a reduced Gini impurity of 0.198 with 7 samples, primarily consisting of 8 instances of Class 0 and 1 instance of Class 1. The decision at this point is to predict Class 0 (LPSA).

Further exploration along the "True" branch reveals another decision node with the condition "None," signifying a leaf node where no further branching occurs. In this leaf node, the Gini impurity is 0.0, implying a pure node with 4 samples, all belonging to Class 0 (LPSA). Therefore, the final prediction for this path is Class 0(LPSA). Contrarily, the "False" branch from the second node leads to another leaf node. Here, the Gini impurity is 0.444, indicating some impurity in the 3 samples. Returning to the root node, the "False" branch leads to a leaf node with a Gini impurity of 0.245. The 3 samples in this node have a class distribution of [Class 0, Class 1] resulting in the final prediction of Class 1.

In summary, this decision tree structure outlines the sequential conditions and decisions that the model employs to predict the classes of samples. It navigates through the features and partitions the dataset based on specific conditions, ultimately arriving at predictions for each path in the tree. Understanding the nodes, conditions, and predicted classes facilitates a comprehensive interpretation of the decision-making process within the tree.

4.4.3 Comparison of Logistic Regression (LR) and Random Forest (RF) model

In terms of performance stability, LR exhibited varying performance across folds during cross-validation, suggesting sensitivity to data partitioning. In contrast, RF demonstrated consistent performance across diverse folds, indicating a higher level of stability in handling different subsets of the data.

Random Forest outperformed Logistic Regression in terms of discrimination, as evidenced by higher Area Under ROC Curve (AUC-ROC) values. This implies that RF had a superior ability to differentiate between positive and negative instances compared to LR.

Logistic Regression provided coefficients, offering insights into the impact of individual predictors on the outcome. This contributes to the interpretability of the model. On the other hand, Random Forest offered feature importance scores, guiding the understanding of variable contributions to the overall model. The choice between coefficients and feature importance depends on the specific interpretative needs of the analysis.

Logistic Regression showed some variability in performance metrics across folds, indicating potential sensitivity to different data subsets. In contrast, Random Forest demonstrated stability in cross-validation and robust performance in the testing phase, suggesting a more reliable generalization to unseen data.

The choice between LR and RF models depends on the specific goals and characteristics of the data. LR provides straightforward interpretability through coefficients, making it suitable for scenarios with simpler relationships. In contrast, RF excels in predictive performance, particularly when dealing with complex, non-

linear data patterns, and offers stability across diverse subsets. The decision hinges on the trade-off between interpretability and performance, with LR favored for its clarity in understanding individual predictors, and RF preferred for robustness in handling intricate datasets and delivering superior predictive capabilities. Ultimately, the selection is context-dependent, necessitating a careful evaluation of model strengths against specific objectives and constraints.

4.5   Environment and Sustainability

4.5.1   SDG3 Good Health and Well-being

The study aims to investigate the brain-behavior mechanisms underlying PSA through the analysis of EEG data, which could contribute to improving mental health outcomes for individuals with this condition. In this study, the research of FAA and DBC as the EEG biomarkers allow early detection for individual with PSA.

4.5.2   SDG9 Industry, Innovation, and Infrastructure

The study's use of biomedical engineering techniques (FAA and DBC) to analyze EEG data may have broader applications in the development of medical technologies and devices to support the diagnosis and treatment of a range of health conditions.

4.5.3   SDG 17 Partnerships for the Goals

The study's findings (FAA, DBC and the classification of individual with PSA) could contribute to the development of effective interventions for individuals with PSA, which could require collaboration and partnership across different sectors and stakeholders to implement.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

In this study, the general aim is to explore the brain-behaviour mechanisms responsible for PSA through EEG data processing. After the raw EEG has been confirmed that the artifacts are fully rejected, the EEG is then transformed into a frequency domain through FFT to make comparative evaluations of Frontal Alpha Asymmetry (FAA) and Delta-Beta Correlation (DBC) in individuals with HPSA and LPSA. Therefore, the hypothesis is supported by HPSA subjects FAA being higher than LPSA subjects consistently. The analysis revealed that the DBC of HPSA is significant and strong for the Frontal region (F3, F4) in incongruent conditions for HPSA subjects, Central region (FC5, FC6) in incongruent conditions for LPSA subjects, Parietal region (P7, P8) both congruently and incongruently paired were found signifiers with a threshold $p \leqslant 0.05$.

Apart from that also, the EEG and performance biomarkers are used for classifying machine algorithm Logistic Regression (LR) and Random Forest 15(RF). LR's mean accuracy is at 78.12%, and RF's mean accuracy is at 69%. There is an improvement needed in both the classification algorithm as well because of a restriction to limit the subjects (individuals with PSA). Therefore, the obtained results contribute to the development of knowledge regarding neurobiological processes that are behind PSA and its prediction models.

5.2   Future Work

There is no doubt that this study has shed light on PSA neurobiology by finding the EEG biomarkers as well as the PSA classification. On the contrary, there are areas in the future research of this study that need improvement. The brain dynamics related to PSA may be further elucidated using other neuroimaging methods like fMRI and MEG. Regarding the approach, it would be worthwhile to explore some complex machine learning methods as well as broaden the sample size of PSA subjects. Hence this is likely to enhance the generalizability of the results as well as further improve the classification model to yield a better classification accuracy.

This finally leads the study to understand how the brain impacts behaviour via public speaking anxiety. These results will form the basis of our further study on PSA, and we shall endeavor to provide more concrete interventions.

REFERENCES

[1]   Valentine Okoronkwo, "29+ Important Public Speaking Statistics To Help You In 2022." Accessed: Apr. 15, 2023. [Online]. Available: https://passivesecrets.com/fear-of-public-speaking-statistics/

[2]   M. Arlin Cuncic, "Speech Anxiety: Public Speaking With Social Anxiety." Accessed: May 27, 2023. [Online]. Available: https://www.verywellmind.com/tips-for-managing-public-speaking-anxiety-3024336

[3]   N. Ibrahim, N. A. Khairol Anuar, M. I. Mokhtar, N. Zakaria, N. H. Jasman, and N. N. Rasdi, "Exploring Fear of Public Speaking Through Social Cognitive Theory," *International Journal of Academic Research in Business and Social Sciences*, vol. 12, no. 1, Jan. 2022, doi: 10.6007/ijarbss/v12-i1/11320.

[4]   S. Wise, C. Huang-Pollock, and K. Pérez-Edgar, "Frontal alpha asymmetry in anxious school-aged children during completion of a threat identification task," *Biol Psychol*, vol. 179, Apr. 2023, doi: 10.1016/j.biopsycho.2023.108550.

[5]   K. L. Poole, B. Anaya, and K. E. Pérez-Edgar, "Behavioral inhibition and EEG delta-beta correlation in early childhood: Comparing a between-subjects and within-subjects approach," *Biol Psychol*, vol. 149, Jan. 2020, doi: 10.1016/j.biopsycho.2019.107785.

[6]   F. S. Feroz, A. R. Salman, M. H. M. Ali, A. I. Ismail, S. Indra Devi, and S. K. Subramaniam, "Attentional bias during public speaking anxiety revealed using event-related potentials," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, pp. 253–259, Oct. 2021, doi: 10.11591/ijeecs.v24.i1.pp253-259.

[7]   Farah Shahnaz Feroz, Muhammad Hairulnizam Mat Ali, Afiq Idzudden Ismail, Ahmad Rifhan Salman, and Faaizah Shahbodin, "EVENT-RELATED POTENTIALS REVEAL IMPAIRED EMOTION-COGNITION INTERACTION IN INDIVIDUALS WITH PUBLIC SPEAKING ANXIETY," 2021.

[8]   V. Flasbeck, J. Engelmann, B. Klostermann, G. Juckel, and P. Mavrogiorgou, "Relationships between fear of flying, loudness dependence of auditory evoked potentials and frontal alpha asymmetry," *J Psychiatr Res*, vol. 159, pp. 145–152, Mar. 2023, doi: 10.1016/j.jpsychires.2023.01.031.

[9]   N. Schürmann, "Enabling #EEGManyLabs: Quality in Automatically Preprocessed EEG Data and Psychopathological Associations of Frontal Alpha Asymmetry," Jan. 2022.

[10]   A. Harrewijn, M. J. W. van der Molen, I. M. van Vliet, J. J. Houwing-Duistermaat, and P. M. Westenberg, "Delta-beta correlation as a candidate endophenotype of social anxiety: A two-generation family study," *J Affect Disord*, vol. 227, pp. 398–405, Feb. 2018, doi: 10.1016/j.jad.2017.11.019.

[11]   C. Barros, A. R. Pereira, A. Sampaio, A. Buján, and D. Pinal, "Frontal Alpha Asymmetry and Negative Mood: A Cross-Sectional Study in Older and Younger Adults," *Symmetry (Basel)*, vol. 14, no. 8, Aug. 2022, doi: 10.3390/sym14081579.

[12]   S. Glier, A. Campbell, R. Corr, A. Pelletier-Baldelli, and A. Belger, "Individual differences in frontal alpha asymmetry moderate the relationship between acute stress responsivity and state and trait anxiety in adolescents," *Biol Psychol*, vol. 172, Jul. 2022, doi: 10.1016/j.biopsycho.2022.108357.

[13]   O. A. David, R. Predatu, and A. Maffei, "REThink Online Video Game for Children and Adolescents: Effects on State Anxiety and Frontal Alpha Asymmetry," *Int J Cogn Ther*, vol. 14, no. 2, pp. 399–416, Jun. 2021, doi: 10.1007/s41811-020-00077-4.

[14]   M. VickramTejwani, "PublicSpeakingAnxiety inGraduateMedical Education—AMatterof Interpersonaland CommunicationSkills?," 2016.

[15]   Charlotte Ruhl, "Stroop Effect Experiment in Psychology." Accessed: May 28, 2023. [Online]. Available: https://www.simplypsychology.org/stroop-effect.html

[16]   Enes Zvornicanin, "What Is Independent Component Analysis (ICA)? | Baeldung on Computer Science." Accessed: May 28, 2023. [Online]. Available: https://www.baeldung.com/cs/independent-component-analysis

[17]   A. Hyvrinen and E. Oja, "Independent Component Analysis: A Tutorial," 1999. [Online]. Available: http://www.cis.hut.fi/projects/ica/

[18]   F. Heine, J. Stekelenburg, and F. P. Hashemi, "The Relationship Between Depression and Alpha Asymmetry in a Non-Clinical Sample," 2022.

[19]   S. Song *et al.*, "Event-Related Alpha Oscillatory Response in Early Stage of Facial Expression Processing in Social Anxiety: Inuence of Language Context," 2023, doi: 10.21203/rs.3.rs-1887631/v2.

[20]   S. Myruski, R. Bagrodia, and T. Dennis-Tiwary, "Delta-beta correlation predicts adaptive child emotion regulation concurrently and two years later," *Biol Psychol*, vol. 167, Jan. 2022, doi: 10.1016/j.biopsycho.2021.108225.

[21]   B. B. Margaret Rose Tobias and J. Arch Tiffany Ito Roselinde Kaiser Erik Willcutt Allison Atteberry, "i A MULTIMETHOD INVESTIGATION OF SOCIAL ANXIETY MAINTENANCE," 2022.

[22]   A. Al-Ezzi, N. K. Selman, I. Faye, and E. Gunaseli, "Electrocortical brain oscillations and social anxiety disorder: A pilot study of frontal alpha asymmetry and delta-beta correlation," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1529/5/052037.

[23]   K. L. Poole and L. A. Schmidt, "Positive Shyness in the Brain: Frontal Electroencephalogram Alpha Asymmetry and Delta–Beta Correlation in Children," *Child Dev*, vol. 91, no. 5, pp. e1030–e1045, Sep. 2020, doi: 10.1111/cdev.13379.

[24]   K. L. Poole and L. A. Schmidt, "Frontal brain delta-beta correlation, salivary cortisol, and social anxiety in children," *J Child Psychol Psychiatry*, vol. 60, no. 6, pp. 646–654, Jun. 2019, doi: 10.1111/jcpp.13016.

[25]   V. De Pascalis, A. Vecchio, and G. Cirillo, "Resting anxiety increases EEG delta–beta correlation: Relationships with the Reinforcement Sensitivity Theory Personality traits," *Pers Individ Dif*, vol. 156, Apr. 2020, doi: 10.1016/j.paid.2019.109796.

[26]   N. Kumar Bhagat, A. K. Mishra, R. K. Singh, C. Sawmliana, and P. K. Singh, "Application of logistic regression, CART and random forest techniques in prediction of blast-induced slope failure during reconstruction of railway rock-cut slopes," *Eng Fail Anal*, vol. 137, Jul. 2022, doi: 10.1016/j.engfailanal.2022.106230.

[27]   J. Lee, J. Cai, F. Li, and Z. A. Vesoulis, "Predicting mortality risk for preterm infants using random forest," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-86748-4.

[28]   Niklas Donges, "What Is Random Forest? A Complete Guide | Built In." Accessed: Nov. 30, 2023. [Online]. Available: https://builtin.com/data-science/random-forest-algorithm

[29]   U. Uzma, S. Kanjilal, and N. Yadav, "Comparative Analysis of Stress among Undergraduate Students Using Logistic Regression and Random Forest Techniques," 2022.

[30]   B. X. W. Liew, F. M. Kovacs, D. Rügamer, and A. Royuela, "Machine learning versus logistic regression for prognostic modelling in individuals with non-specific neck pain," *European Spine Journal*, vol. 31, no. 8, pp. 2082–2091, Aug. 2022, doi: 10.1007/s00586-022-07188-w.

[31]   R. M. Millis, J. Arcaro, A. Palacios, and G. L. Millis, "Electroencephalographic Signature of Negative Self Perceptions in Medical Students," *Cureus*, Feb. 2022, doi: 10.7759/cureus.22675.

[32]   K. I. Jang, C. Lee, S. Lee, S. Huh, and J. H. Chae, "Comparison of frontal alpha asymmetry among schizophrenia patients, major depressive disorder patients, and healthy controls," *BMC Psychiatry*, vol. 20, no. 1, Dec. 2020, doi: 10.1186/s12888-020-02972-8.

APPENDICES

Appendix A

Literature Gap

Paper/Year	Subject	Experiment	Reaction Time study	Event Related Potential	FAA	DBC
Feroz[6]/2021	PSA	Stroop	Yes	Yes	Nil	Nil
Farah[7]/2021	PSA	Flanker	Yes	Yes	Nil	Nil
Flasbeck[8]/2023	Non-PSA	Beck Depression Inventory (BDI-II), State-trait Anxiety Inventory (STAI), Fear of Flying Scale (FFS)	Nil	Yes	Yes	Nil
Schumann[9]/2022	Non-PSA	EEG resting state paradigm	Nil	Nil	Yes	Nil
Harrewijn[10]/2018	Non-PSA	Social Performance Task	Nil	Nil	Nil	Yes
Barros[11]/2022	Non-PSA	Emotion Stimuli	Nil	Nil	Yes	Yes
Glier[12]/2022	Non-PSA	Trier Social Stress Test (TSST)	Nil	Nil	Yes	Nil
Wise[4]/2023	Non-PSA	Threatening Identification Task	Nil	Nil	Yes	Nil
David[13]/2021		REThink Game	Nil	Nil	Yes	Nil

Appendix B

Summary of FAA Past Research Paper

Year/Paper	Subjects	Method	Results	Similarities	Differences
Flasbeck [8] /2023	36 participants with self-reported fear of flying (FF) and 41 unaffected participants (NFF)	FAA, ICA, FFT	Participants with a fear of flying showed higher right FAA at F8-F7 electrodes compared to the other group, but no difference was found at F4-F3 electrodes.	Method: FAA, FFT, ICA	Different Experimental Paradigm (Resting-State)
Schumann [9] /2022	47 healthy adults	ICA, FAA, FFT	An association between relative rightward lateralization of alpha power at one electrode pair and depressive symptoms was found.	Method: ICA, FAA, FFT	Type of subjects
Barros [11] /2022	39 older adults ($\geq$ 60 years old) & 57 younger adults (between 18 and 35 years old)	ICA, FAA	Older adults had a higher FAA value than younger adult	Method: ICA, FAA	Type of subjects
Glier [12] /2022	145 adolescents	ICA, FAA, FFT	• Adolescents with rightward FAA activation and high trait anxiety showed blunted cortisol reactivities.	Method: ICA, FAA, FFT	Type of subjects

David [13] /2021	165 healthy children and adolescents (aged between 10 and 16 years)	FAA, FFT, ICA	Higher scores indicate more right-sided alpha activity, suggesting greater inhibition in the right hemisphere associated with negative affect. The study found a significant negative correlation between frontal alpha asymmetry and state anxiety after the REThink intervention. Specifically, the REThink group showed a significant increase in right hemisphere inhibition, supporting the efficacy of the intervention in reducing negative modulation.	Method: FAA, FFT, ICA	Different subjects and experimental paradigm
Heine [18] /2022	35 university students	FAA, Infinite Impulse Response (IIR) filter, ICA-based EOG correction, FFT	There was no difference in alpha asymmetry between the two depression groups used in this study, and there was no significant association between depression	Method: FAA, ICA, FFT	Method: IIR filter

			scores on the PHQ-9 and alpha asymmetry measured on different electrode pairs.		
Song [19] /2023	62 healthy university students	Event Related Potential (ERP), FAA, ICA	In individuals with social anxiety, there was a significant decrease in occipital alpha oscillation power compared to the healthy control group, particularly when the emotional context and expression were congruent. Additionally, in negative contexts, the social anxiety group exhibited significantly reduced frontal alpha lateralization compared to the healthy control group.	Method: FAA, ICA	Type of subject, Method: ERP

Appendix C

Summary of DBC Research Paper

Year/Paper	Subjects	Method	Results	Similarities	Differences
Harrwijn [10] /2018	113 participants (18 SAD subjects, 25 subclinical SAD subjects, 43 clinical SAD subjects)	DBC, FFT	Co-segregation analyses indicated a higher negative delta-beta correlation during anticipation in participants with (sub)clinical SAD compared to those without (sub)clinical SAD.	Method: DBC, FFT	Type of subjects
Myruski [20] /2022	53 children (23 female, 30 male)	DBC, FFT	A stronger direct brain connectivity (DBC) was found to be associated with a higher utilization of adaptive and relatively active emotion regulation (ER) strategies.	Method: DBC, FFT	Type of subjects
Margaret [21] /2022	184 high-SAD adults	DBC, FFT	No significant difference in the delta-beta correlation between the Writing Exercise Conditions	Method: DBC, FFT	Type of subjects
Al Ezzi [22] /2020	4 social anxiety disorder subjects and 4 HC subjects	DBC	In individuals with Social Anxiety Disorder (SAD), there was a higher	Method: DBC	Type of subjects

			correlation observed between DBC during the baseline condition (both eyes closed and eyes open) compared to the recovery task. In contrast, healthy control (HC) individuals exhibited a stronger correlation during the recovery state compared to the baseline state.		
Poole K [23] /2020	67 children	DBC, DFT	Positive shy children had a higher frontal delta–beta correlation compared to other groups. (Non-positive shy and low shy)	Method: DBC	Method: DFT
Poole K [24] /2019	50 children	DBC, DFT	The patterns of children's salivary cortisol across visits could be categorized into two distinct classes: a high, stable class comprising 53% of the sample and a low, unstable class comprising 47% of the	Method: DBC	Method: DFT

			sample. On the other hand, children's social anxiety exhibited two distinct classes: a high, stable class comprising 50% of the sample and a low, stable class also comprising 50% of the sample.		
Poole K [5] /2020	118 children.	DBC, FFT,	Children who scored high on behavioral inhibition (BI) demonstrated higher delta-beta correlation compared to children with low BI in frontal and central brain regions. Additionally, there was a marginal increase in delta-beta correlation in parietal brain regions for high BI children relative to low BI children.	Method: DBC, FFT	Type of subjects
Pascalis [25] /2020	59 students	DBC, FFT, Gratton	• The resting Anxiety group exhibited a significant positive between-subject	Method: DBC, FFT	Method: Gratton

			delta-beta correlation, which was significantly higher than the correlation observed in the Relaxation group. • DBC specifically for low delta activity in the Anxiety group, but not in the Relaxation group.		

Appendix D

Project Flowchart
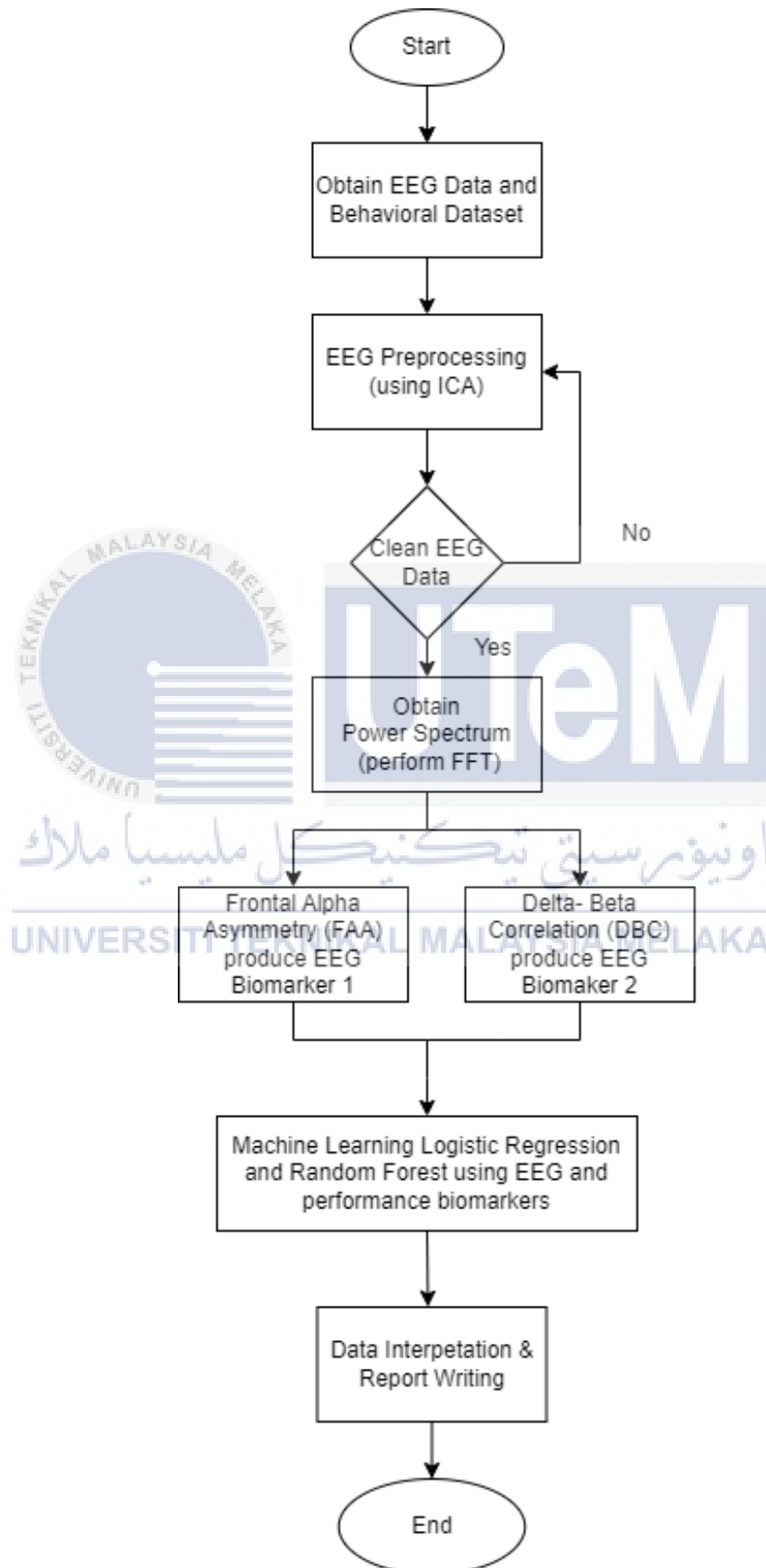
Appendix E

Gannt Chart Explanation

Activities	Explanation
Literature Review	• Duration: Week 1 of Semester 1 to Week 11 of Semester 2 • Description: Conduct a critical analysis and summary of existing research and scholarly articles on the topic to establish a comprehensive understanding of the subject.
EEG Data Preprocessing	• Duration: Week 5 of Semester 1 to Week 13 of Semester 1 • Description: Preprocess the raw EEG data using Independent Component Analysis (ICA) to remove artifacts and obtain clean EEG data suitable for accurate analysis.
Power Spectral Analysis	• Duration: Week 10 of Semester 1 to Week 13 of Semester 1 • Description: Perform power spectral analysis on the preprocessed EEG data to obtain the power spectrum, which will be used for further analysis, specifically for FAA and DBC analysis.
FAA Analysis	• Duration: Week 10 of Semester 1 to Week 13 of Semester 1 • Description: Analyze the new EEG biomarkers related to FAA using the power spectrum data obtained in the previous step.
DBC Analysis	• Duration: Week 1 of Semester 2 to Week 6 of Semester 2 • Description: Analyze the new EEG biomarkers related to DBC using the power spectrum data obtained in the previous step.
Machine Learning	• Duration: Week 6 to Week 13 of Semester 2 • Description: Apply the logistic regression and random forest algorithm to classify the PSA data, enabling the identification of patterns and trends in the EEG signals.
Concluding Findings	• Duration: Week 12 and Week 13 • Description: Summarize the research findings, draw conclusions based on the analysis conducted, and

	evaluate the implications of the results for the research topic.
Publish Results	• Duration: Week 7 to Week 13 of Semester 2 • Description: Prepare and publish the research results, including writing a research paper or preparing a presentation to disseminate the findings to the scientific community.
Thesis Writing	• Duration: Week 5 of Semester 1 to Week 13 of Semester 2 • Description: Complete any remaining tasks, finalize documentation, and wrap up the project activities.
Seminar	• Seminar 1 is placed in week 13 Semester 1. • Seminar 2 is placed in week 13 Semester 2.

Appendix F

FAA index of HPSA and LPSA Subjects

	HPSA			LPSA	
Subject ID	FAA Congruent ($\mu v^2/Hz$)	FAA Incongruent ($\mu v^2/Hz$)	Subject ID	FAA Congruent ($\mu v^2/Hz$)	FAA Incongruent ($\mu v^2/Hz$)
00002	1.913560829	-0.872445289	00004	-0.915559417	-1.679574557
00003	0.984342041	-0.10682081	00010	-0.749973927	0.986686376
00008	2.013738054	0.714541389	00011	0.298042323	0.643536293
00009	0.209379355	2.243509682	00013	0.818312511	-1.703405971
00019	-0.19812664	-0.814376253	00014	0.398481529	-1.311372544
00023	0.137737532	-0.399893528	00016	1.269352963	1.340380834
00026	-0.55389698	-0.028685738	00018	0.854864976	0.617230278
00030	0.031467873	-0.871097556	00021	0.072226261	0.047154243
00031	0.389115946	1.832702756	00022	0.063087941	-0.330518844
00035	0.145410643	0.074962248	00027	0.212838388	0.234623502
00039	0.020940256	0.914162706	00028	0.194046484	-0.757590564
00040	0.489253114	0.206459062	00034	2.643751006	1.824609133

Appendix G

Delta-Beta Band of Frontal Electrode

Subject	PSA Group	Delta-Beta Band of Frontal Electrode (F3, F4)			
		Congruent		Incongruent	
		Delta Band	Beta Band	Delta Band	Beta Band
00002	HPSA	0.143913212	0.003002985	0.158301189	0.00239913
00003	HPSA	0.129667477	0.005921813	0.657632985	0.003832944
00008	HPSA	2.317243866	0.005333956	3.566894396	0.004668984
00009	HPSA	2.96915688	0.002344195	1.432906363	0.003199687
00019	HPSA	0.068275781	0.001288341	0.057505279	0.00179124
00023	HPSA	0.252759448	0.008022118	0.706506947	0.004134565
00026	HPSA	0.542671915	0.009542996	1.71163987	0.006196544
00030	HPSA	0.182126077	0.001228762	0.091257563	0.00122733
00031	HPSA	0.847440048	0.002323072	0.127855258	0.002797011
00035	HPSA	0.280022656	0.001456297	0.414470462	0.000839663
00039	HPSA	0.108143765	0.001779057	0.12655659	0.002057001
00040	HPSA	0.401745402	0.000932914	0.091838911	0.001345144
00004	LPSA	0.015733503	0.000454499	0.037051664	0.000696081
00010	LPSA	0.028184964	0.001704607	0.109326277	0.001583384
00011	LPSA	0.376650416	0.001629949	0.712701254	0.001625066
00013	LPSA	0.371387693	0.001359823	0.648851451	0.001411304
00014	LPSA	0.113076486	0.002900179	0.078252811	0.001598261
00016	LPSA	0.271465869	0.003196422	0.621828375	0.001520073

00018	LPSA	0.108679247	0.001133484	0.055805553	0.000991276
00021	LPSA	0.081528955	0.003228288	0.030713855	0.00491885
00022	LPSA	0.042739617	0.001279323	0.109224771	0.002608506
00027	LPSA	0.310134487	0.003050174	0.954615745	0.003263333
00028	LPSA	0.148886876	0.00502375	0.239426321	0.002769657
00034	LPSA	0.705243989	0.000823739	0.508967308	0.001131569

Appendix H

Delta-Beta Band of Central Electrode

Subject	PSA Group	Delta-Beta Band of Central Electrode (FC5, FC6)			
		Congruent		Incongruent	
		Delta Band	Beta Band	Delta Band	Beta Band
00002	HPSA	0.148834507	0.002048379	0.472751956	0.001724733
00003	HPSA	0.189389976	0.004603348	0.152904451	0.004954717
00008	HPSA	0.026170351	0.003108526	0.092129787	0.003292064
00009	HPSA	1.96746023	0.002590976	1.953468973	0.004683448
00019	HPSA	0.839903451	0.00261786	0.151043479	0.002751307
00023	HPSA	0.055745083	0.002421609	0.021615457	0.001895168
00026	HPSA	1.677112786	0.005812988	0.668930304	0.011986215
00030	HPSA	0.375798165	0.001237622	0.333084748	0.0010486
00031	HPSA	0.28266474	0.004466466	0.194466276	0.00476498
00035	HPSA	0.094745131	0.002337458	0.030905229	0.00187479
00039	HPSA	0.07241626	0.002533529	0.310450517	0.001658008
00040	HPSA	0.047150275	0.000950085	0.043533854	0.001177487
00004	LPSA	0.470148268	0.004490366	0.211929316	0.003774993
00010	LPSA	0.078144943	0.002005724	0.181431855	0.003213225
00011	LPSA	0.101033505	0.000816274	0.119418102	0.000860148
00013	LPSA	0.250556649	0.002903069	0.246129648	0.001295708
00014	LPSA	0.047893257	0.003416815	0.155086961	0.002147759
00016	LPSA	0.879656779	0.006356852	0.57723969	0.008855639

00018	LPSA	0.117595295	0.002014072	0.17980335	0.001737505
00021	LPSA	0.040664579	0.003411126	0.045320085	0.004343132
00022	LPSA	0.852029506	0.002604434	0.188275744	0.003874726
00027	LPSA	0.115345892	0.002264182	0.166743617	0.001541121
00028	LPSA	0.420564028	0.006338828	0.320800217	0.003424749
00034	LPSA	0.622108763	0.003225878	0.104570655	0.003076211

Appendix I

Delta-Beta Band of Parietal Electrode

Subject	PSA Group	Delta-Beta Band of Parietal Electrode (P7, P8)			
		Congruent		Incongruent	
		Delta Band	Beta Band	Delta Band	Beta Band
00002	HPSA	0.177185162	0.003395027	0.087440019	0.003218071
00003	HPSA	0.069712601	0.001544989	0.025355559	0.000879779
00008	HPSA	0.654710279	0.008582221	0.65325542	0.00709008
00009	HPSA	2.126156949	0.003895325	1.770171306	0.004484665
00019	HPSA	0.120230574	0.001079442	0.296592618	0.001343042
00023	HPSA	1.332897094	0.010068069	0.687057284	0.008308575
00026	HPSA	1.396784363	0.002916479	0.190377602	0.002663107
00030	HPSA	0.076490202	0.002036578	0.116360501	0.00149984
00031	HPSA	2.548694493	0.013268577	1.770278254	0.01345413
00035	HPSA	0.13227847	0.002555959	0.175750291	0.001402166
00039	HPSA	0.116946969	0.002161231	0.408219959	0.002209689
00040	HPSA	0.122412855	0.001914244	0.263350899	0.002848912
00004	LPSA	0.032166725	0.000572604	0.026423916	0.000535255
00010	LPSA	0.211107053	0.002701157	0.216608493	0.003799396
00011	LPSA	0.381365596	0.002615944	0.434667559	0.001924738
00013	LPSA	0.342439294	0.002659268	0.708646315	0.001930405
00014	LPSA	0.023596931	0.002228532	0.036845644	0.001432966
00016	LPSA	0.105817259	0.00629893	0.104199245	0.008295717

00018	LPSA	0.092779853	0.001936936	0.050261889	0.002084774
00021	LPSA	0.344656121	0.003143018	0.144293341	0.002505291
00022	LPSA	0.208449902	0.001283811	0.480549699	0.001928015
00027	LPSA	0.473138417	0.007600657	0.53446337	0.004106584
00028	LPSA	0.700120477	0.015112667	0.53176208	0.010156132
00034	LPSA	0.795474305	0.003694686	0.319800288	0.0034275

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Appendix J

Delta-Beta Band of Temporal Electrode

Subject	PSA Group	Delta-Beta Band of Temporal Electrode (T7, T8)			
		Congruent		Incongruent	
		Delta Band	Beta Band	Delta Band	Beta Band
00002	HPSA	0.120857345	0.002574969	0.260193402	0.003744578
00003	HPSA	0.052027493	0.00716276	0.060832564	0.004852687
00008	HPSA	1.021415827	0.005879451	1.585840127	0.003918045
00009	HPSA	2.031930443	0.009879445	0.704312816	0.009203984
00019	HPSA	0.025560977	0.000803683	0.02226617	0.000773845
00023	HPSA	0.152335098	0.006947462	0.744438895	0.006354218
00026	HPSA	0.909034355	0.011710825	0.985746801	0.011718639
00030	HPSA	0.358499191	0.002521451	0.328023796	0.003257561
00031	HPSA	0.52248498	0.00918227	0.248144164	0.007364866
00035	HPSA	0.125287343	0.010757893	0.140062783	0.005261207
00039	HPSA	0.085907349	0.001750615	0.111966902	0.001157908
00040	HPSA	0.097668435	0.002531188	0.149010662	0.004343601
00004	LPSA	0.098029528	0.004921165	0.221416788	0.002244694
00010	LPSA	0.102924982	0.001771233	0.163587578	0.002407269
00011	LPSA	0.259827551	0.001746161	0.112436179	0.001165593
00013	LPSA	0.274096785	0.005391184	0.187040594	0.004510473
00014	LPSA	0.0587425	0.003208734	0.05018075	0.001759122
00016	LPSA	0.089590802	0.012509595	0.075999999	0.011784661

00018	LPSA	0.307696553	0.001731089	0.08301043	0.001296149
00021	LPSA	0.386751297	0.004917519	0.982622324	0.004554407
00022	LPSA	0.034898167	0.000810872	0.033691855	0.001183571
00027	LPSA	0.140542345	0.001060563	0.225798643	0.000872732
00028	LPSA	0.535675133	0.013009599	1.638939133	0.0108108
00034	LPSA	0.551660347	0.002308174	0.37123371	0.003074312

Appendix K

Summary of DBC

Brain Region (Electrode Used)	Stroop Task Conditions	Subject Group	r	p	Correlation
Frontal (F3, F4)	Congruent	HPSA	.0436	.893	Non-significant, weak, positive
		LPSA	-.1068	.728	Non-significant, weak, negative
	Incongruent	HPSA	.6931	.012	Significant, strong, positive
		LPSA	-.0127	.967	Non-significant, weak, negative
Central (FC5, FC6)	Congruent	HPSA	.3747	.230	Non-significant, moderate, positive
		LPSA	.3785	.202	Non-significant, moderate, positive
	Incongruent	HPSA	.3350	.287	Non-significant, moderate, positive
		LPSA	.6348	.020	Significant, strong, positive
Parietal (P7, P8)	Congruent	HPSA	.7130	.009	Significant, strong, positive
		LPSA	.6069	.028	Significant, strong, positive
	Incongruent	HPSA	.7477	.005	Significant, strong, positive
		LPSA	.1410	.646	Non-significant, weak, positive
Temporal (T7, T8)	Congruent	HPSA	.5135	.088	Non-significant, strong, positive
		LPSA	.1947	.524	Non-significant, weak, positive
	Incongruent	HPSA	.4344	.158	Non-significant, moderate, positive
		LPSA	.3868	.192	Non-significant, moderate, positive

Appendix L

Behavioural Data (Reaction Time)

Subject	Group	Reaction Time (ms)	
		Congruent	Incongruent
00002	HPSA	1703.856667	1771.022034
00003	HPSA	615.2339286	689.3186441
00008	HPSA	1150.060345	1163.176667
00009	HPSA	1086.927119	1162.123333
00019	HPSA	629.5644068	742.955
00023	HPSA	665.9916667	896.6728814
00026	HPSA	1331.393103	1635.737931
00030	HPSA	1762.288333	1679.596667
00031	HPSA	779.1206897	781.8072727
00035	HPSA	1124.813333	1466.035593
00039	HPSA	970.1218182	1023.557895
00040	HPSA	1082.175	1189.057627
00004	LPSA	1109.956667	1117.583333
00010	LPSA	759.2711864	779.3685185
00011	LPSA	863.6491525	957.0706897
00012	LPSA	854.5166667	969.6649123
00013	LPSA	1466.366667	1739.930508
00014	LPSA	748.7839286	842.15
00021	LPSA	1078.961667	1131.082456

00022	LPSA	800.2283333	922.9175439
00027	LPSA	821.1152542	823.9694915
00028	LPSA	946.1932203	1250.823636
00034	LPSA	678.7116667	756.085
00036	LPSA	1015.383333	1295.267797

Appendix M

Summarization of Feature Importance of RF Model

Feature	Importance Value	Interpretation
State	0.2344	State Anxiety has the highest importance among features, suggesting participants' level of anxiety during task has a significant impact on the model's predictions.
Error_Con	0.1730	Error_Con follows closely, indicating its substantial contribution to the model's decision-making process..
Error_Incon	0.1211	Error_Incon holds notable importance, contributing significantly to the model's understanding of the data.
RT_Con	0.1053	RT_Con has a moderate level of importance, suggesting its relevance in influencing the model's outcomes.
Age	0.1022	Age is identified as a significant feature, implying its influence on the model's predictions.
FC5_Con	0.0977	FC5_Con contributes moderately to the model's decision-making, highlighting its role in shaping predictions.

| Trait | 0.0929 | Trait Anxiety is a relevant feature, indicating its impact on the model's understanding of the dataset. |
| sigma | 0.0733 | Sigma has the lowest importance among features, suggesting a relatively smaller influence on the model's predictions. |

Appendix N

Logistic Regression Python Code

```python
#Logistic Regression Classifier
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split, KFold, cross_val_score, cross_val_predict
from sklearn.metrics import accuracy_score, precision_score
from sklearn.metrics import confusion_matrix, classification_report, cohen_kappa_score
from sklearn.metrics import roc_curve, auc, precision_recall_curve
data = pd.read_excel('ERP Result N450.xlsx')
le=LabelEncoder()
data['Gender'] = le.fit_transform(data['Gender'])
y = data['Group']
X = data.drop(columns=['Group','Subject','FC5_Incon','RT_Incon','mu','tau','Gender'])
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
#All the metrics
all_accuracy = []
all_fpr = []
all_tpr = []
all_precision = []
all_recall = []
all_precision1 = []
all_specificity=[]
all_sensitivity=[]

#Adjustment values
num_folds=4 #Number of Folds num folds
alpha=2 #1 Regularization alpha value
k=1
kf = KFold(n_splits=num_folds, shuffle=True, random_state=42)
for train_index, val_index in kf.split(X_train):

    print('kth-fold=',k)
    X_train_fold, X_val_fold = X_train.iloc[train_index], X_train.iloc[val_index]
    y_train_fold, y_val_fold = y_train.iloc[train_index], y_train.iloc[val_index]

    # Create a logistic regression model with L1 regularization for each fold
    logit_model = sm.Logit(y_train_fold, X_train_fold)
    logit_result1 = logit_model.fit_regularized(method='l1', solver='liblinear', alpha=alpha)

    # Display summaries of the logistic regression model for each fold
    print(logit_result1.summary())
    y_pred_prob = logit_result1.predict(X_test)
    y_pred = (y_pred_prob > 0.5).astype(int)
    # Calculate evaluation metrics
    #Confusion Matrix
    confusion_matrix_result = confusion_matrix(y_test, y_pred)
    #Plot confusion matrix as a heatmap with annotations
    plt.figure(1, figsize=(5, 4))
    heatmap = sns.heatmap(confusion_matrix_result, annot=True, fmt="d", cmap="Blues")
    #Get axis information
    ax=heatmap.axes
    #Label TP, TN, FP, FN on the heatmap
    ax.text(0.5, 0.2, f'True Negative', ha='center', va='center', color='black', fontsize=10)
    ax.text(1.5, 0.2, f'False Positive', ha='center', va='center', color="black", fontsize=10)
    ax.text(0.5, 1.2, f'False Negative', ha='center', va='center', color="black", fontsize=10)
    ax.text(1.5, 1.2, f'True Positive', ha='center', va='center', color='black', fontsize=10)
    plt.xlabel('Predicted')
    plt.ylabel('Actual')
    plt.title(f'Confusion Matrix of Fold {k}')
    plt.show()
```

```python
    accuracy = accuracy_score(y_test, y_pred)
    precision1 = precision_score(y_test, y_pred)
    tn, fp, fn, tp = confusion_matrix_result.ravel()
    specificity = tn / (tn + fp)
    sensitivity = tp / (tp + fn) #tpr
    kappa = cohen_kappa_score(y_test, y_pred)

    fpr, tpr, thresholds = roc_curve(y_test, y_pred_prob)
    roc_auc = auc(fpr, tpr)

    total = len(y_test)
    tp_percentage = (tp / total) * 100
    tn_percentage = (tn / total) * 100
    fp_percentage = (fp / total) * 100
    fn_percentage = (fn / total) * 100

    # Plot ROC curve
    plt.figure(figsize=(8, 8))
    plt.plot(fpr, tpr, color='darkorange', lw=2, label=f'ROC curve (area = {roc_auc:.2f})')
    plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
    plt.xlim([0.0, 1.0])
    plt.ylim([0.0, 1.05])
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title(f'ROC Curve - Fold {k}')
    plt.legend(loc='lower right')
    plt.show()

    precision, recall, _ = precision_recall_curve(y_test, y_pred_prob)
    all_accuracy.append(accuracy)
    all_fpr.append(fpr)#
    all_tpr.append(tpr)#sensitivity
    all_precision.append(precision)
    all_recall.append(recall)
    all_precision1.append(precision1)
    all_specificity.append(specificity)
    all_sensitivity.append(sensitivity) # Recall, TPR
    k=k+1

print("\n")
print("Measured Result of Every Fold")
print("Accuracy    :", [round(acc, 4) for acc in all_accuracy])
print("Precision   :", [round(prec, 4) for prec in all_precision1])
print("Specificity:", [round(spec, 4) for spec in all_specificity])
print("Sensitivity:", [round(sens, 4) for sens in all_sensitivity])
print("Kappa       :", [round(kap, 4) for kap in all_kappa])

avg_accuracy = np.mean(all_accuracy)
avg_precision = np.mean(all_precision1)
avg_specificity = np.mean(all_specificity)
avg_sensitivity = np.mean(all_sensitivity)
avg_kappa = np.mean(all_kappa)
print("\n")
# Print average values
print("Average Accuracy    :", round(avg_accuracy, 4),"= {:.2%}".format(avg_accuracy))
print("Average Precision   :", round(avg_precision, 4),"= {:.2%}".format(avg_precision))
print("Average Specificity:", round(avg_specificity, 4),"  = {:.2%}".format(avg_specificity))
print("Average Sensitivity:", round(avg_sensitivity, 4),"  = {:.2%}".format(avg_sensitivity))
print("Average Kappa       :", round(avg_kappa, 4)," = {:.2%}".format(avg_kappa))
```

Appendix O

Random Forest Python Code

```python
#Random Forest Classifier
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, KFold, cross_val_score
from sklearn.model_selection import cross_val_predict,learning_curve
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.metrics import precision_score, recall_score, f1_score, roc_curve, auc
from sklearn.metrics import classification_report, cohen_kappa_score
from yellowbrick.model_selection import FeatureImportances
from sklearn.tree import export_graphviz
data = pd.read_excel('ERP Result N450.xlsx')
le=LabelEncoder()
data['Gender'] = le.fit_transform(data['Gender'])
y = data['Group']
X = data.drop(columns=['Group','Subject','FC5_Incon','RT_Incon','mu','tau','Gender'])
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

#Random Forest
clf_k=RandomForestClassifier(max_depth=None,
                             min_samples_leaf=1,
                             min_samples_split=6,
                             n_estimators=20,
                             random_state=41)
num_folds=4
#Create a KFold object
kf=KFold(n_splits=num_folds, shuffle=True, random_state=41)
#Perform k-fold cross validation
cross_val_scores=cross_val_score(clf_k, X_train, y_train, cv=kf)
#cross_val_scores=cross_val_score(clf_k, X, y, cv=kf)
#Accuracy of All Fold
print("Cross-validation scores:", [round(score, 4) for score in cross_val_scores])
#Print the mean and standard deviation of the cross-validation scores
print(f"Mean accuracy: {cross_val_scores.mean():.2f}")
print (f"Standard deviation: {cross_val_scores.std():.2f}")

#Predict on the testing data for evaluation
clf_k.fit(X_train, y_train)
#Get predicted probabilities for class 1
y_pred_test=clf_k.predict_proba(X_test)[:, 1]
y_pred_test = (y_pred_test > 0.5).astype(int)
print("Classification Report (Testing Data):")
print(classification_report(y_test, y_pred_test))

#Confusion matrix for testing data
cm_test=confusion_matrix(y_test, y_pred_test)
plt.figure(1, figsize=(5, 4))
heatmap=sns.heatmap(cm_test, annot=True, fmt="d", cmap="Blues")
#Get axis information
ax=heatmap.axes
#Label TP, TI, FP, FN on the heatmap
ax.text(0.5, 0.2, f'True Negative', ha='center', va='center', color='black', fontsize=10)
ax.text(1.5, 0.2, f'False Positive', ha='center', va='center', color="black", fontsize=10)
ax.text(0.5, 1.2, f'False Negative', ha='center', va='center', color="black", fontsize=10)
ax.text(1.5, 1.2, f'True Positive', ha='center', va='center', color='black', fontsize=10)
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title(f"Confusion Matrix of Testing Data")
plt.show()
```

```python
#Feature Importances
importances = clf_k.feature_importances_
feature=X_train.columns.tolist()
print("Feature:",feature)
sorted_indices = importances.argsort()[::-1]

# Print feature importances and names
for i in range(len(feature)):

    print(f"{feature[sorted_indices[i]]}: {importances[sorted_indices[i]]}")

# Create a bar chart to visualize feature importances
plt.figure(figsize=(10, 6))
plt.barh(range(len(feature)), importances[sorted_indices])
plt.yticks(range(len(feature)), [feature[i] for i in sorted_indices])
plt.ylabel("Feature")
plt.xlabel("Importance")
plt.title("Feature Importances")
plt.show()

#ROC Curve
fpr, tpr, _= roc_curve(y_test, y_pred_test)
roc_auc=auc(fpr, tpr)
plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, color="darkorange", lw=2, label='ROC curve (AUC= {:.2f})'.format(roc_auc))
plt.plot([0, 1], [0, 1], color="navy", lw=2, linestyle='--')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic')
plt.legend(loc="lower right")
plt.show()
```

```python
#Learning Curve
train_sizes, train_scores, test_scores=learning_curve(clf_k, X_train, y_train, cv=kf, train_sizes=np.linspace(0.1, 1.0, 5))
plt.figure(figsize=(8, 6))
plt.plot(train_sizes, np.mean(train_scores, axis=1), 'o-', label='Training score')
plt.plot(train_sizes, np.mean(test_scores, axis=1), 'o-', label='Validation score')
plt.xlabel('Training examples')
plt.ylabel('Score')
plt.title('Learning Curve')
plt.legend(loc='best')
plt.show()
```

```python
# Visualize one of the trees in the Random Forest
estimator = clf_k.estimators_[0]

# Export the tree to a DOT file
export_graphviz(estimator, out_file='tree.dot',
                feature_names=X_train.columns.tolist(),
                class_names=['Class 0', 'Class 1'],
                filled=True, rounded=True, special_characters=True)
```