DEMONSTRATING THE PROSPECTIVE COMPONENT OF SENSE OF AGENCY USING MACHINE LEARNING

B. Tech. Capstone Project Report

by

Anway Pimpalkar

111907066

Under the guidance of

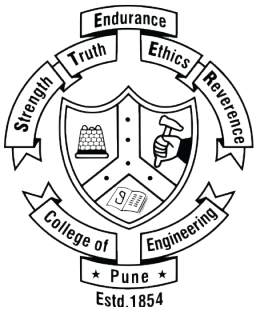
Dr. Nivethida Thirugnanasambandam

Indian Institute of Technology Bombay

and

Dr. Rashmika Patole

College of Engineering Pune



DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION

COLLEGE OF ENGINEERING PUNE - 05

May 2023

Abstract

A normal human exhibits fluent control over the events in the outside world through their actions. This control is linked with an innate and subjective experience, referred to as *Sense of Agency*. Though the exact neural mechanisms leading to the experience remain unknown, the source of this experience can be attributed to two components - retrospective and prospective Sense of Agency. The retrospective component banks on the inference that the brain creates a Sense of Agency based on inferences of its own actions after the completion of it. The prospective component, on the other hand, suggests that the brain's Sense of Agency is also associated with the prediction of outcomes, and not only the retrospective inferences. This project aims to strengthen the literature supporting the prospective component of Sense of Agency, and use a machine learning approach to show differences in neural activity before the onset of movement. Using various feature extraction and machine learning techniques, the research conducted for this project leads to the conclusion that the prospective component plays a fundamental role in generating a Sense of Agency.

Acknowledgements

I would like to express my sincere gratitude to my mentors and peers who have played a critical role by supporting me throughout this project.

Firstly, I extend my heartfelt thanks to **Dr. Nivethida Thirugnanasambandam** from the Indian Institute of Technology Bombay and **Dr. Rashmika Patole** from College of Engineering Pune for their guidance and mentorship throughout the project. Their expertise in neuroscience and signal processing has been instrumental in shaping the direction of this research endeavor. I also owe a debt of gratitude to **Dr. Raghavendran Lakshminarayanan** for his invaluable technical guidance in machine learning and statistics, and **Dr. Shilpa Metkar** for her insights on neural signal processing techniques.

I am also deeply grateful to my colleagues from the Human Motor Neurophysiology and Neuromodulation Laboratory at the Indian Institute of Technology Bombay, including **Sainath Murali**, **Rajat Joshi**, **Arkoprovo Sarkar**, **Rashi Sharma**, and **Garima Gupta** for their unwavering support and encouragement. Their foundational work and multidisciplinary knowledge provided a solid foundation for me to build upon.

Furthermore, I would like to acknowledge the **Center of Excellence in Signal & Image Processing (COE-SIP)** at College of Engineering Pune for providing access to their state-of-the-art computing facilities. Their resources were invaluable in enabling me to complete computationally intensive HCTSA iterations in a timely manner.

Lastly, I would like to express my appreciation to **Prof. Mukul Sutaone**, **Dr. Srinivas Mahajan**, and **Dr. Vibha Vyas** for their support and encouragement in pursuing a project in the domain of my interests. This experience has been a defining moment in my academic journey at College of Engineering Pune, and I am truly grateful for their guidance and mentorship.

Contents

List of Tables **ii**

List of Figures **iii**

1 Introduction **1**

1.1 Localization of Sense of Agency 1

1.2 Components of Sense of Agency 2

1.2.1 Retrospective Component of Sense of Agency 2

1.2.2 Prospective Component of Sense of Agency 4

1.2.3 Measuring Sense of Agency - Intentional Binding 4

1.3 Bereitschaftspotential or Readiness Potential 5

1.4 EEG Signal Frequency Bands 7

1.5 Project Aim and Objectives 7

2 Literature Review **9**

2.1 Background and Theories 9

2.2 Measuring Sense of Agency Implicitly 10

2.3 Importance of Sense of Agency 10

2.3.1 Schizophrenia and Other Neurological Disorders 10

2.3.2 Healthy Aging 12

2.3.3 Human-Computer Interaction 12

2.3.4 Freedom and Responsibility 13

3 Experimental Setup and Data Collection **14**

3.1 Research Paradigm 14

3.1.1 Procedure 14

3.1.2 Probability Blocks 15

3.2 Cleaning and Preprocessing 16

3.3 Data Verification 16

4 Methodology I: Highly Comparative Time Series Analysis 18

4.1 Feature Extraction Overview 18

4.2 HCTSA 18

4.3 Pearson's Correlation Coefficient 19

4.4 Methodology 20

4.4.1 Approach 1: Features Extracted from Averaged 9 Electrodes 20

4.4.2 Approach 2: Features Extracted from Individual 9 Electrodes 21

5 Methodology II: Spectral Density Estimation 24

5.1 Overall Objective 24

5.2 Welch's Method 25

5.2.1 Steps for Calculating the PSD 25

5.2.2 Advantages and Precautions 26

5.3 Recursive Feature Elimination 27

5.4 Support Vector Classifiers 28

5.5 Random Forest Classifiers 28

5.6 Differences between SVC and RFC 29

5.7 Methodology 30

5.7.1 Subjectwise Averaging 30

5.7.2 Extracting PSD and Time Domain Features 30

5.7.3 Approach 1: PSD Features with SVC, Validated with Data from Same Standard Distribution 31

5.7.4 Approach 2: PSD Features with SVC, Validated with Data from Different Standard Distribution 32

5.7.5 Approach 3: Adding Time Domain Features to Approach 2 32

5.7.6 Approach 4: PSD Features with RFC 32

5.7.7 Approach 5: PSD Features and Time Domain Features with RFC . . 35

5.7.8 Approach 6: K-Fold Nested LeaveOneOut SVC (Approach 3 Extension) 35

5.7.9 Approach 7: K-Fold Nested LeaveOneOut RFC (Approach 5 Extension) 35

6 Implementation, Results and Discussions 38

6.1 Methodology I: HCTSA 38

6.1.1 Implementation 38

6.1.2 Results 38
6.2 Methodology II: Spectral Density Estimation 39
6.2.1 Implementation 39
6.2.2 Results 39
6.2.3 Feature Analysis 41
6.3 Discussions 41
6.3.1 HCTSA vs Welch's Method 44
6.3.2 Probability Block Classification using PSD 44
6.4 Future Directions 46

7 Summary and Conclusions **47**

List of Tables

3.1 Number of Individual Trials per Probability Block Across All Subjects. . . . 15

6.1 Summary of Results for Methodology II. 40

6.2 Summary of Features Extracted for SVC Binary Classifiers. 42

6.3 Summary of Features Extracted for RFC Binary Classifiers. 43

List of Figures

1.1 Pre-Supplementary Motor Area (Pre-SMA) of the Brain. 2
1.2 Position of Anterior Insula Region in the Brain. 3
1.3 Position of Cerebellum in the Brain. 3
1.4 Prospective and Retrospective Sense of Agency. 4
1.5 Visual Representation of Intentional Binding Effect. 5
1.6 Early and Late Readiness Potentials. 6

3.1 Research Paradigm for the Experiment. 15
3.2 Grand Averages for Electrodes Around the Pre-SMA Region. 17

4.1 Approach 1: HCTSA Methodology for Averaged Electrode Data. 22
4.2 Approach 2: HCTSA Methodology for Individual Electrode Data. 23

5.1 Steps for Calculation of PSD using Welch's Method. 25
5.2 Methodology 2, Approach 1: PSD Features with SVC, Validated with Data from Different Standard Distribution. 33
5.3 Methodology 2, Approach 2 and 3: PSD and Time Domain Features with SVC, Validated with Data from Different Standard Distribution. 34
5.4 Methodology 2, Approach 4 and 5: PSD Features and Time Domain Features with RFC. 36
5.5 Methodology 2, Approach 6 and 7: K-Fold Nested LeaveOneOut Approach. . 37

6.1 Visualization of the Electrodes Used to Train the SVC Model 45

Chapter 1

Introduction

Humans are used to feeling as though they are in charge of their actions. This *Sense of Agency* [1] is largely influenced by the consequences observed due to carrying out a particular action. The experience of control is heightened if people act in accordance with their intentions, that is, a higher *fluency of agency* is exhibited [2]. Sense of Agency is central to the experience of being an autonomous and active agent in the world. Understanding how the brain generates this Sense of Agency is a central question in neuroscience, with implications for fields ranging from philosophy to psychology to robotics.

The fact that the Sense of Agency is a subjective sensation and difficult to quantify directly presents one of the major difficulties in studying it. To examine this phenomenon in the lab, researchers have created a number of experimental paradigms, such as the Libet's Clock task. These exercises often entail asking participants to evaluate the time or relationship between their actions and their results.

1.1 Localization of Sense of Agency

Numerous brain areas have been identified by neuroscientists as contributing to the production of the experience of agency. The pre-supplementary motor area (pre-SMA), which is engaged in organizing and carrying out movements, is one of the important areas. According to studies, the pre-SMA is associated with the Sense of Agency, and when this region is disrupted, one may feel less in control of their actions [3]. It is located as shown in Figure 1.1.

The anterior insula, which is a part of the insular cortex, has also been linked to the production of the sensation of agency, according to research [4]. It has been demonstrated that when people feel like they have control over their activities, the insula, a part of the

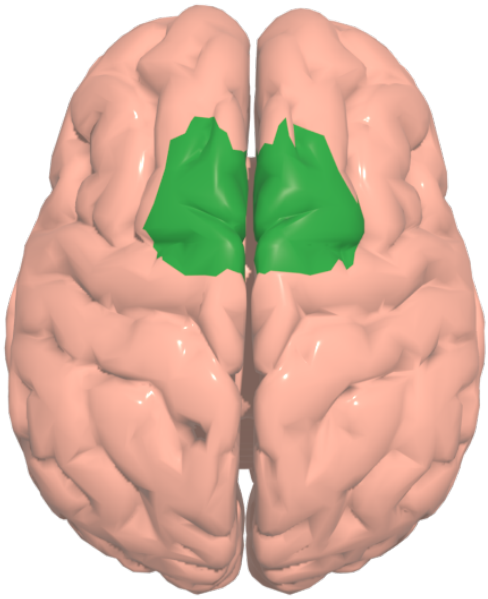


Figure 1.1: Pre-Supplementary Motor Area (Pre-SMA) of the Brain.

brain involved in processing bodily sensations and emotions, is active. Reduced sensation of control over one's body and behaviours might result from anterior insula disruptions.

In the posterior fossa, the cerebellum has also been implicated in generating the Sense of Agency [5]. Recent findings showing that the cerebellum participates in building the Sense of Agency, specifically the sensory consequences of our actions. This input allows us to adjust our actions in response to the changes in the environment.

1.2 Components of Sense of Agency

Cognitive processes can access various signals produced by the brain that contribute to a Sense of Agency at various moments. While sensory feedback signals are active after the action has been completed, premotor signalling occurs before the action begins. These two components are called *retrospective* and *prospective* respectively.

1.2.1 Retrospective Component of Sense of Agency

Feedback signals which arise due to the perception of a causal outcome are the source of the retrospective component of Sense of Agency. Any comparator models aiming to distinguish between these two account for the temporal misalignment since these signals are delayed due to the causal chain connecting the action and outcome.

This retrospective component is hypothesised to be triggered by external influencing

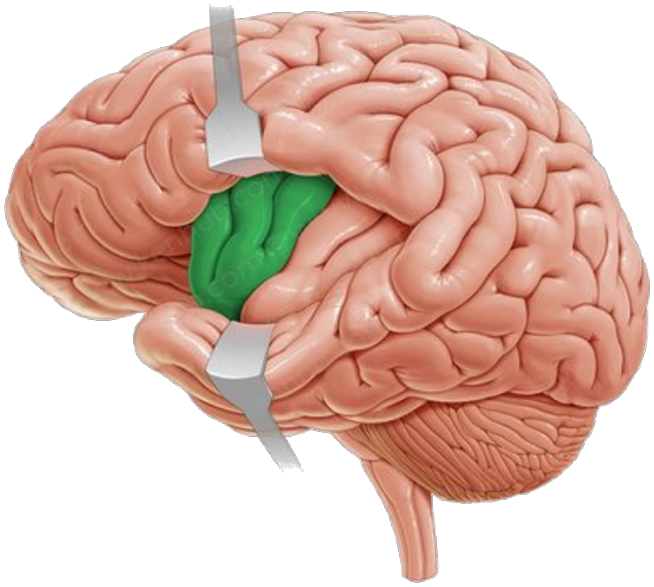


Figure 1.2: Anterior Insula Region Position in the Brain *(Courtesy: www.kenhub.com)*.

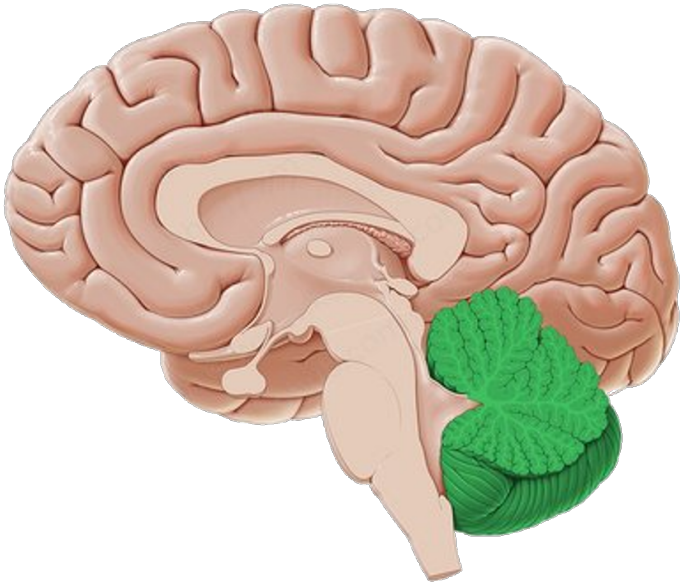


Figure 1.3: Position of Cerebellum in the Brain *(Courtesy: www.kenhub.com)*.

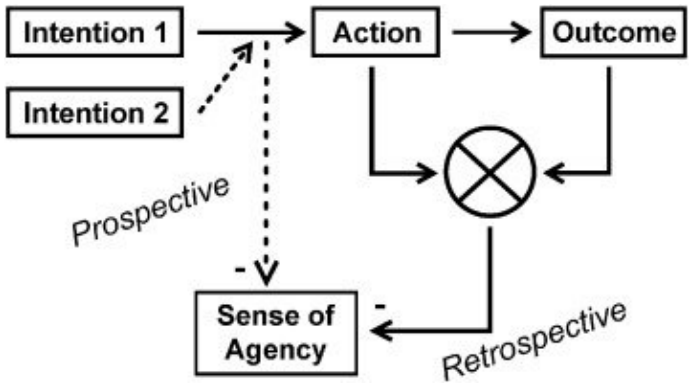


Figure 1.4: Prospective and Retrospective Sense of Agency. *(Courtesy: Sidarus et al. [6])*

factors such as the environment, which operate completely independently from human cognition. As a result, it may be stated that the brain develops a feeling of agency based on conclusions about its own activities that it draws after carrying them out.

1.2.2 Prospective Component of Sense of Agency

Along with the post-action component of Sense of Agency, literature also suggests the existence of a prospective component which is seen before the onset of the movement. Studies have shown that varying the probability that an outcome will occur, distinguishable blocks of behavior changes can be seen. For instance, when there is a 50% chance that a tone would play after a button click, it will result in a larger sensation of agency than trials when no tone is produced.

This is an interesting phenomenon, as it implies that the brain's Sense of Agency is also associated with the prediction of outcomes, and not only the retrospective inferences as claimed by literature. The prospective component is the crux of this project, and will be explored further in the forthcoming chapters.

1.2.3 Measuring Sense of Agency - Intentional Binding

Previous experiments have shown that Sense of Agency can be studied through *intentional binding* [2]. It refers to the subjective experience of time between an action and its outcome, where the interval between the action and the outcome is perceived as shorter than the interval between two similar events that are not causally related. The concept of intentional binding was first introduced by Haggard et al. in 2002, who observed that when individuals press a button and a tone follows immediately after, they perceive the interval between the

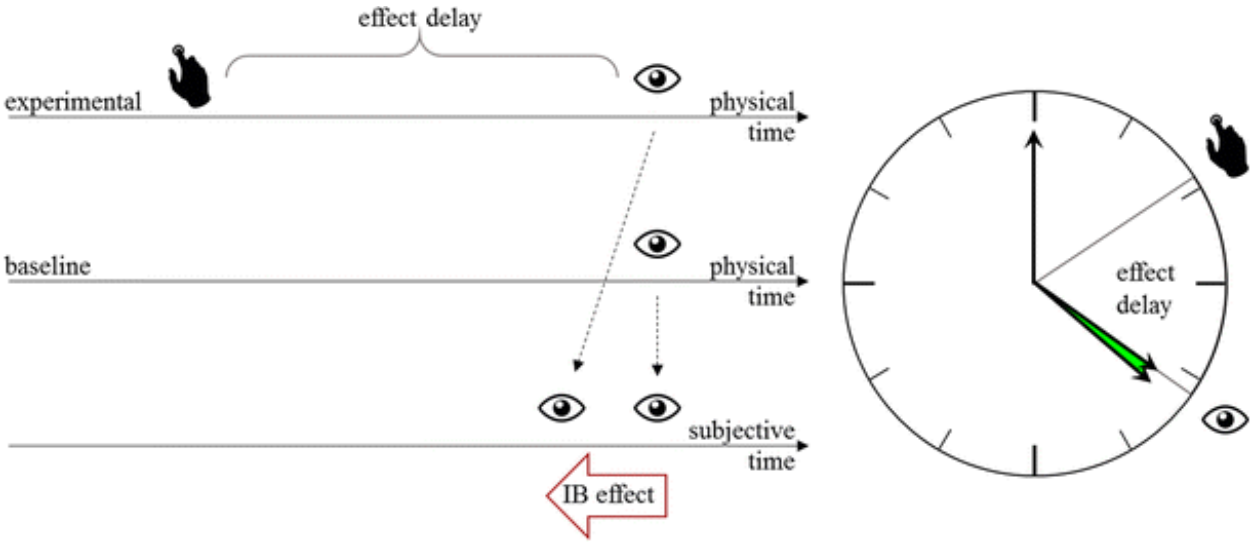


Figure 1.5: Visual Representation of Intentional Binding Effect. *(Courtesy: Ruess et al. [8])*

button press and the tone as shorter than when the tone is delayed [7].

This effect is not observed when the tone is played at random intervals, suggesting that intentional binding occurs only when there is a clear causal relationship between the action and the outcome. It is an implicit measure of the control which avoids any cognitive biases towards overestimating one's own agency.

Studies have reported that intentional actions are associated with an anticipated time compression between the corresponding action and its effect, such as an audio feedback tone. This compression appears to be sensitive to variables including reward, action-outcome intervals, and the degree of real control and is not substantially associated with people. Hence, shifts in intentional binding may not directly be an appropriate diagnostic measure of Sense of Agency. However, under identical experimental conditions, a difference in intentional binding can be interpreted as a difference in Sense of Agency [9].

1.3 Bereitschaftspotential or Readiness Potential

The onset of movement is characterized by a *Bereitschaftspotential* or *Readiness Potential (RP)*, which reflects the preconscious brain activity preceding a voluntary action [10]. This shift is not directly visible from trial to trial, rather becomes prominent when it multiple trials are averages together. Based on the distribution of the potential along the scalp and the slope of negative potential, the activity can be divided into two components: *early* and *late RP*. The early RP starts around 2 seconds before the commencement of a voluntary

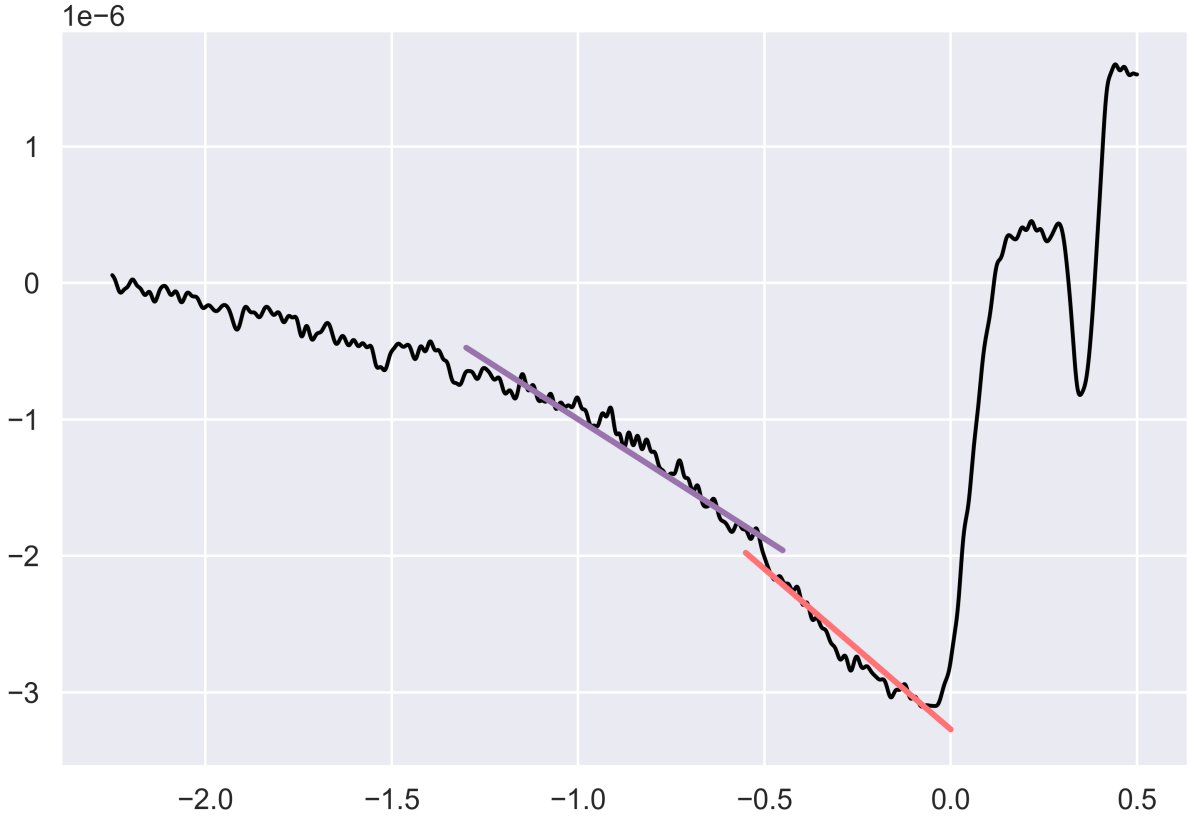


Figure 1.6: Early (Lavender) and Late (Light Red) Readiness Potentials.

movement, and has an overall consistent negative slope, although some gradual decrease (increase in the negative slope) is observed. Past studies have shown that the maximal early RP activity occurs in the pre-SMA. On the other hand, the late RP is characterized by a steeper gradient arising approximately 0.5 seconds before the movement onset. This has been localized to the contralateral premotor cortex.

Many studies have established a link between the onset of a RP and the preparation of a movement, and that a transient negativity of the slow cortical potential fluctuations facilitate the initiation. As mentioned, this pattern is observed visually only after averaging many single trials and cannot be observed directly, which means that trial-to-trial variations are not always caught onto fast. Therefore, we aim to study these individual variations in the RP which can lead to a subjective evaluation of the intentional binding and Sense of Agency shown by the subject [11].

1.4 EEG Signal Frequency Bands

The electrical activity of the brain can be classified into different frequency bands, each representing a different aspect of neural function. These frequency bands can provide valuable information about brain activity across the spatiotemporal domain. There are five main frequency bands that are commonly used in EEG analysis: delta, theta, alpha, beta, and gamma.

The delta band, which normally ranges between 0.5 and 4 Hz, has the lowest frequency. Delta waves are connected to deep sleep and are assumed to have a role in the upkeep of physiological processes including breathing and heartbeat.

Theta band activity occurs between 4 and 8 Hz and is connected to sleep and tiredness. Deep relaxation and meditation are known to produce theta waves. Attention deficit hyperactivity disorder (ADHD) or other cognitive deficits may be indicated by an overabundance of theta waves in specific regions of the brain.

With the eyes closed and in a calm or awake state, the alpha band, which ranges from 8 to 13 Hz, is frequently observed. The visual cortex of the brain is also connected to alpha waves, which are assumed to be important in sensory processing. Alpha waves can be absent when there is brain injury or a neurological disorder.

The beta band, with a frequency range between 13 and 30 Hz, is typically detected in an alert, focused condition. Beta waves, which are connected to cognitive processes, can be used to measure focus and attention. An excess of beta waves could be a sign of stress or anxiety.

The gamma band, which ranges in frequency from 30 to 100 Hz, is linked to complex cognitive functions including memory and learning. Gamma waves can be used to measure the brain's reaction to sensory stimuli because they are also involved in the processing of sensory information.

1.5 Project Aim and Objectives

The aim of the project is to elucidate the neural correlates between the prospective Sense of Agency and intentional binding experimentally. We aim to localize the prospective component to different regions of the brain and also understand its correlated signal frequency spectrum of it.

The objectives of the project can be summarized as the following:

1. To demonstrate the role of the prospective component of Sense of Agency using machine learning classifiers.

2. To build a classifier model relying upon specific features of a given premotor EEG signal as input to categorize data into various probability blocks.

3. To attempt to localize the effects of intentional binding and Sense of Agency in the brain, narrowing down the source locations and frequency bands.

.

Chapter 2

Literature Review

Over the last twenty years, there have been a significant increase in the number of scientific studies on Sense of Agency. This chapter discusses the background and measurement techniques of Sense of Agency, and also presents an extensive review on the importance of studying this subject.

2.1 Background and Theories

According to one of the earliest theories set forward by Wolpert et al. in 2000, predictive motor control mechanisms are the source of action awareness. In this position, we might be aware of the motor predictions made by the motor control system's forward model component [12]. This position is supported by a variety of sources. For instance, as part of an exploratory technique before neurosurgery in 1991, Fried et al. activated the supplementary motor region in humans. The need to move was noted by patients at modest electrical stimulation intensities. When the level of stimulation was raised, they actually carried out the suggested movement [13]. According to Wolpret et al., this phenomena occurs because conscious knowledge of action is based not on the actual position of the limbs but rather on the predicted position of the limbs as determined by efferent orders. According to this perspective, we have internal motor knowledge about our movements that is, at least in part, consciously aware.

Even though experiments on Sense of Agency have some methodological issues, this increase still exists. One significant one is that there is little phenomenological Sense of Agency [14]. That is to say, our awareness of our agentic experiences when taking acts is often rather limited. This is very different from conscious experience in other manners such as vision, where our perceptions are robust and steady. This implies that Sense of Agency

might be challenging to quantify. Because of this, experimenters have had to use a lot of creativity to create paradigms that accurately capture this elusive feeling.

2.2 Measuring Sense of Agency Implicitly

Haggard et al. used the Libet Clock methodology in 2002 to prove that when we voluntarily act, the perceived timings of the activity and its result are moved closer together. This shift in how you see time is thought to be an implicit indicator of your feeling of agency [7]. This formed the most popular implicit measure of Sense of Agency, termed intentional binding.

In a second experiment, the authors changed the timing of the tone following intentional activity as well as its predictability. According to the study's findings, temporal continuity and predictability have a significant role in determining binding. Based on their research, the authors hypothesised that the central nervous system has a distinct cognitive function that involves tying together significant sensory events that surround intentional action. They theorised that this function may be essential for the typical Sense of Agency.

The intentional binding effect was used by Moore et al. in 2008 to quantify the degree to which sensations of action are impacted by subsequent effects. In this paradigm, participants hit a key at a predetermined moment, which results in a tone being played 250 milliseconds later. Because the participant causes the tone, this results in an operant context. In successive blocks of trials, participants evaluate the timing of their keypress or the timing of the tone they heard [15].

2.3 Importance of Sense of Agency

An overview of Sense of Agency theory and research is presented in the preceding section. From this perspective, it is not quite evident why any of this is important, especially in terms of impact. Many studies have also emphasized on the importance of studying Sense of Agency, summarized in this section.

2.3.1 Schizophrenia and Other Neurological Disorders

The most agency research has been done on schizophrenia than any other condition, making it the quintessential Sense of Agency disorder. Positive symptoms and negative symptoms are the two categories into which the symptoms of schizophrenia are divided. Negative

symptoms are those in which a normal function is absent (such as 'alogia' or diminished speech). On the other side, positive symptoms are those that include anomalous perceptions (hallucinations) or beliefs (delusions). The category of positive symptoms includes abnormal perceptions of agency. Although these strange experiences can take many different forms, passivity symptoms are the most typical. A patient who exhibits signs of passivity will believe that they have no control over their activities [16].

Research on schizophrenia patients has shown that these people struggle with agency processing. Simple hand motions were done by healthy controls and patients with schizophrenia by Daprati et al. in 1997 [17]. They weren't able to watch their own motions up close. Instead, a video link allowed them to see visual feedback of the movement on the screen. These motions were either (a) their own real motions, (b) motions performed by another experimenter in a separate room, or (c) motions created by that experimenter executing a different motion. To avoid any visible identifying cues, both the experimenter and the volunteers wore gloves. After each trial, the individual was merely asked to say whether the movement on the screen was their own or the experimenter's. When the researcher made the identical motions as the patients, patients - particularly those who were experiencing signs of passivity - made more mistakes than controls in identifying the source of the action.

Since then, a number of further investigations have supported these action recognition issues. Since then, a number of further investigations have supported these action recognition issues. For instance, Franck et al. in 2001 evaluated patients and healthy controls on an action recognition test [18]. In this experiment, participants moved while only seeing video feedback of their moves. Different degrees of spatial and temporal distortion were introduced. Participants were required to comment on whether the hand motions seen on the screen resembled their own after each experiment. In all scenarios, healthy people tended to decline earlier than ill who took significantly longer to notice these inconsistencies. Once more, this points to aberrant action awareness in patients.

In addition to schizophrenia, a number of other illnesses are starting to get agency researchers' attention. For instance, Gentsch et al. have demonstrated that people with obsessive compulsive disorder exhibit deficiencies in sensorimotor prediction, which reduces sensory suppression. This discovery supports the earlier findings from schizophrenic patients [19]. Additionally, Oren et al. demonstrated that people with high obsessive-compulsive tendencies frequently leave out the word "agency" when speaking, which may be a sign of diminished agency in these people [20].

2.3.2 Healthy Aging

There has been a big drive to study cognitive changes across the lifetime due to an ageing population. This is done in an effort to lessen some of the detrimental consequences that these changes may have on older persons. Old age appears to be accompanied with changes in feeling of agency, which calls for more research. The wellbeing of older adults may be improved by recognising these changes and creating treatments to address them [16].

Most studies on the subject have focused on how general changes in Sense of Agency (based on self-reports) and ageing relate to various measures of health and well-being. It is clear from this research that ageing is associated with a loss of agency. For instance, Lachman and Firth found that roughly 80% of young individuals (25–39 years old) disagreed with the statement "What happens in my life is beyond my control" but only 62% of elderly persons did [21].

Few studies have particularly examined this from a cognitive neuroscience or experimental psychology perspective. The research conducted by Metcalfe et al. is one of the few that has. They found that older people responded less readily to three external performance manipulations (such as adding a temporal delay between their movement and a pointer moving on the screen) than a control group of younger adults. Intriguingly, the data' pattern points to a decreased sensitivity to external sensory signals to agency in older persons [22]. This should be further investigated in future studies. It will then be feasible to start creating therapies intended to address these anomalous agency processing patterns in older persons.

2.3.3 Human-Computer Interaction

The Sense of Agency of the user is a crucial factor to take into account while creating new interfaces, as has long been understood. In actuality, the seventh of Shneiderman's Eight Golden Rules of Interface Design is to "support an internal locus of control" while designing interfaces [23]. According to him, users "strongly desire the sense that they are in control of the system and that the system responds to their actions." Given this, research on feeling of agency will be very helpful for interface design, both in terms of the measures that have been produced and in terms of understanding what neurocognitive processes affect Sense of Agency.

Researchers are now attempting to bridge the gap between HCI and agency. The impact of different user interfaces on the user's Sense of Agency has been studied in the past. Intentional binding for traditional keyboard input and intentional binding for a brand-new input

method called "skinput" (in which a user taps on their own skin to run the computer) were contrasted by Coyle et al. They discovered that skinput had stronger purposeful binding, indicating that this input method gives the user a greater sensation of control [24]. In 2014, Limerick et al. examined intentional binding for a voice interface in a different investigation. When compared to the keyboard interface, intentional binding was found to be significantly lower for the speech interface [25].

2.3.4 Freedom and Responsibility

Humans appear to value responsibility highly; most, if not all, civilizations demand that its members be held accountable for their actions. Haggard and Tsakiris have made a strong case for the importance of Sense of Agency in determining how culpability is assigned [26]. According to a manuscript by Frith, accepting responsibility for one's own acts serves a crucial societal purpose. It implies that individuals may be held accountable for their actions, enabling behaviour to be properly governed through punishment or reward. It is possible to execute this behavioural management in a way that will benefit the social group and foster social cohesiveness [27]. The legal system is a mechanism that promotes this type of behavioural regulation in many civilizations. The legal system is likely to be affected by this research given the significance of Sense of Agency for establishing culpability [16].

Nichols has highlighted an intriguing connection between Sense of Agency research and the free will debate. The conflict between our perception of ourselves as aware, logical free agents and our understanding that this is irreconcilable with determinism creates the free will dilemma.Since our experiences with agency that surround our voluntary activities help to create the perception that we are generally aware, logical free agents, the Sense of Agency is important in this context. According to Nichols, understanding the neurological underpinnings of free will beliefs would help us assess whether or not those beliefs are legitimate and accurate [28].

Chapter 3

Experimental Setup and Data Collection

This chapter details the experimental background of the project, including the research paradigm and the data that was generated through it. The experimental setups, data collection, cleaning, and preprocessing steps were planned and implemented by Sainath Murali[1] et al. at the National Brain Research Centre (Manesar, India).

3.1 Research Paradigm

As discussed in Chapter 1, the aim is to study the Sense of Agency experienced by a subject during a voluntary movement. The paradigm was conducted on 25 healthy adults who volunteered for the study (13 females, 12 males). 64-channel EEG data was recorded in an acoustically and electromagnetically attenuated room using Brain Vision actiCHamp Plus[2]. All electrophysiological data was recorded at a 1000 Hz sampling rate.

3.1.1 Procedure

The experiment uses a modified Libet clock method at it's crux. A Libet clock is a well-established procedure to explore subjective Sense of Agency and decision timings [29]. For the current application, to increase the chances complete voluntary movement of the subject without any pre-planned outcome, the numbers on a Libet clock were replaced with randomized alphabets, as shown in Figure 2.1. Within the alphabet clock, a red dot revolved around the center with a revolutional period of approximately 2500 milliseconds.

[1]Human Motor Neurophysiology and Neuromodulation Laboratory, IIT Bombay, India

[2]Brain Vision LLC., North Carolina, USA

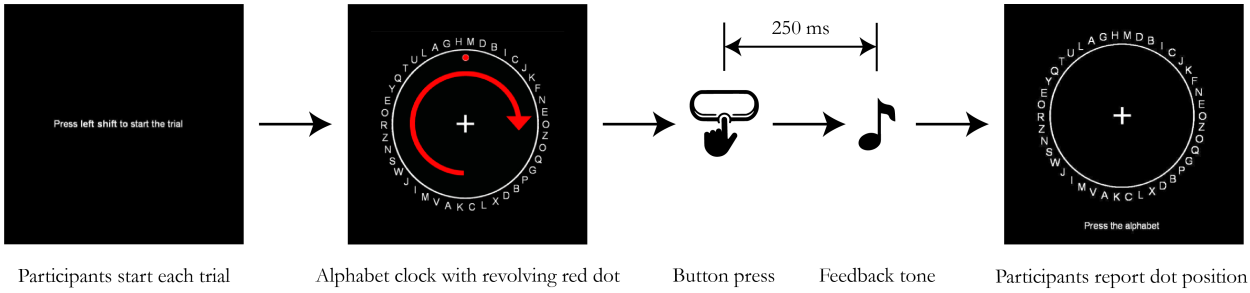


Figure 3.1: Research Paradigm for the Experiment.

The revolving alphabet clock was presented to the participants on a monitor screen. Participants were asked to perform a voluntary movement by pressing the spacebar key anytime according to their wish after one complete rotation. 250 milliseconds later, a tone was played in accordance to the probability function discussed below. Approximately one rotation later, the clock disappeared and then was shown again with the instructions to mark the position of the red dot at which they pressed the button. The time difference between the percepted button press and actual button press is calculated and saved as a measure of the *judgement error*.

3.1.2 Probability Blocks

A probability function was used to determine whether a tone is to be played as feedback to the voluntary movement or not. Probability functions of 0% (baseline), 50% and 75% were used to generate data in five blocks as shown in Table 3.1. All blocks were pooled for further steps.

Probability	Block	Description	Number of Single Trials
0%	Baseline No Feedback	Trials with no possibility of a tone following action.	800
50%	50% With Feedback	The trials where a tone followed the action, when the probability of it occuring was 50%.	800
	50% No Feedback	The trials where a tone did not follow the action, when the probability of it occuring was 50%.	800
75%	75% With Feedback	The trials where a tone followed the action, when the probability of it occuring was 75%.	1200
	75% No Feedback	The trials where a tone did not follow the action, when the probability of it occuring was 75%.	400

Table 3.1: Number of Individual Trials per Probability Block Across All Subjects.

3.2 Cleaning and Preprocessing

The raw EEG data collected was cleaned and preprocessed before analysis. After importing the EEG, anatomical and digitization files, a bandpass and notch filter was applied to them, followed by detrending. The EEGs for the first 15 subjects were referenced to the FCz electrode, and the POz electrode was used for the rest. Any bad channels and noisy segments were subsequently removed. Independent Component Analysis was used to remove any unwanted artifacts in the data such as eye blinks or movements. The bad channels were interpolated and marked. Lastly, the data was epoched at the the time of the button press, marking it to be at 0.0 seconds. Datapoints of 2.25 seconds before the event and 0.50 seconds after the event were retained for further analysis.

A total of 4,000 individual trials were collected across all subjects. Due to errors in the data collection process of Subject 06, the data had to be dropped. After removing the bad trials which were interpolated, a total of 3,015 single trials were available for the data verification.

3.3 Data Verification

The scope of the current project work commences with the EEG analysis performed in this section.

To verify the correctness of the data collected, the grand average of trials collected at the active electrodes around the central midline (Cz) were visualized and inspected. This was chosen as these electrodes collect electrophysiological activity from the pre-SMA region, which has been localized as the source of early RP in individuals. The highest potential was expected to be seen in this region.

Hence, based on the electrode positions, the Cz, FCz, CPz, FC1, FC2, C1, C2, CP1, and CP2 electrode averages were examined. The grand average plots for these nine electrodes are shown in Figure 3.2, with amplitude marked in millivolts on the y-axis and time in seconds on the x-axis. The maximal negative amplitude can be observed in the FCz electrode, indicating that the source of the potential is closer to the region captured by it. Further away from the FCz electrode both vertically and horizontally, the amplitude weakens. The data captured from this electrode also has a lower SNR than the rest.

Observing the Cz electrode amplitude data, the different components discussed prior can be identified. Starting from -1.3 seconds, a slow negative gradient can be traced till about

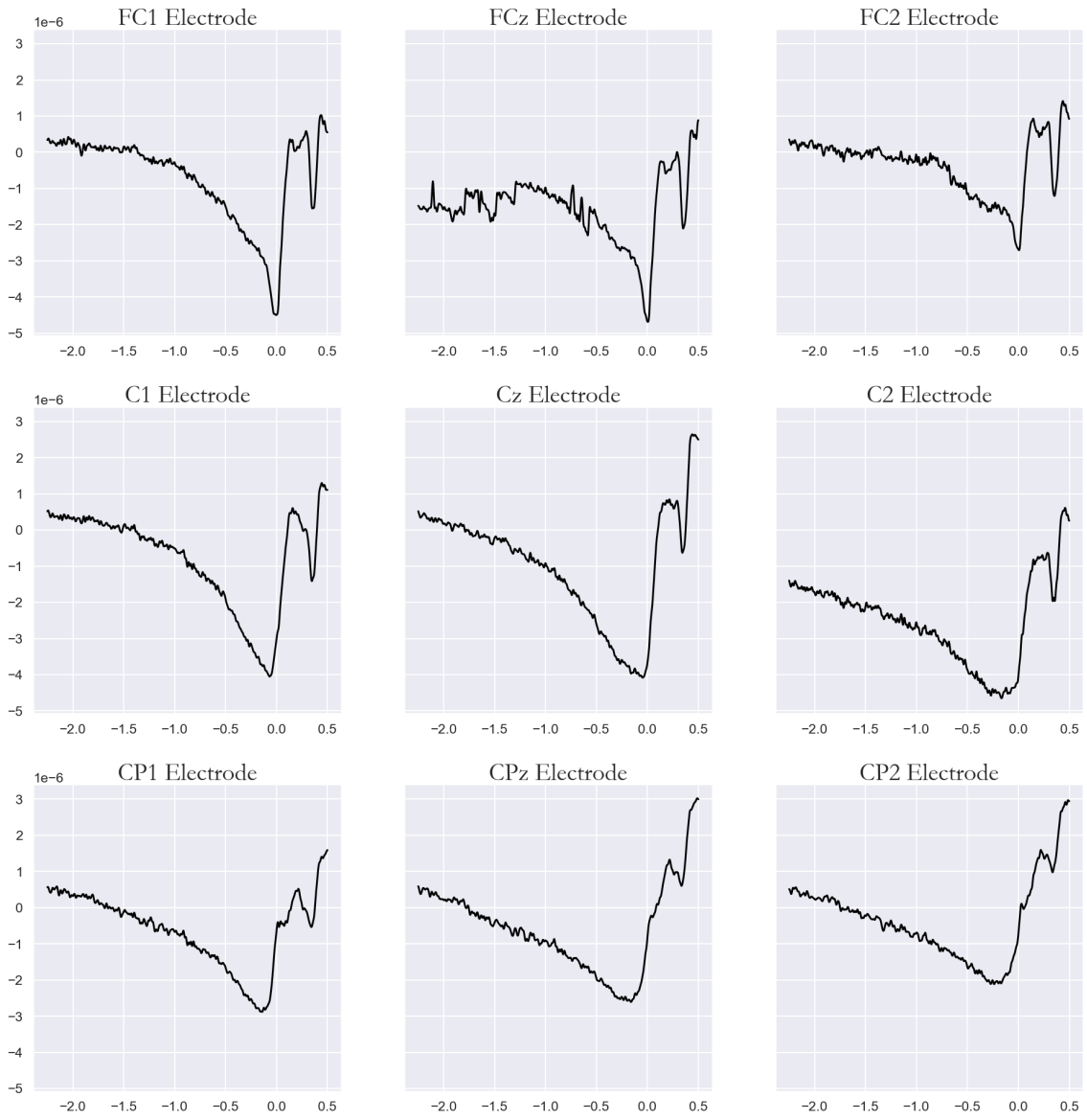


Figure 3.2: Grand Averages for Electrodes Around the Pre-SMA Region.

-0.5 seconds, after which the gradient increases negatively rapidly till the onset of the action at 0.0 seconds. The prior potential corresponds to the early RP, and the latter to the late RP generated to facilitate the onset of movement. This is in agreement to the RP characteristics discussed in previous literature.

Chapter 4

Methodology I: Highly Comparative Time Series Analysis

In the initial attempt to achieve the project objectives, the regression problem statement was approached by looking at each EEG single trial as a time series data stream with various types of extractable features. This was carried out by using the open source HCTSA package developed by Fulcher et al. in 2017 [30][31].

4.1 Feature Extraction Overview

As mentioned in Chapter 1, classifiers for electrophysiological data are a common methodology adopted for a wide range of studies. These classifiers are based on key feature extraction methods and implementations to guide input parameters for predicting classes. These features extracted are in different forms and types, such as time domain, frequency/spectral domain, statistical, decomposition, fractal scaling, and more [32].

Instead of focusing on a few hand-picked features, a mass feature extraction methodology was adopted, as detailed in the following section. Although feature extraction algorithms are common for EEG classification studies, almost none have used them to predict cognitive variables such as the one looked at in this present study (i.e., judgement error) [33].

4.2 HCTSA

Highly Comparative Time Series Analysis (HCTSA) is an approach demonstrated by Fulcher et al. in 2013 [30]. It is a computational method for the analysis of time series data that allows for the extraction of a large number of features or descriptors from the data. These

descriptors capture various aspects of the time series, such as its statistical properties, shape, and frequency content. Once these descriptors are extracted, machine learning techniques can be used to classify, cluster, or otherwise analyze the data. HCTSA has been used for a variety of purposes, including climate research, financial statistics, and neuroscientific data analysis. It is intended to be a general-purpose tool for the study of time series data.

The HCTSA package for MATLAB[1] extracts 7,749 features from a given time series data. These characteristics are based on the distribution, the properties of correlation, the information theory, model fitting, forecasting, stationarity, step detection, nonlinear analysis, fractal scaling, periodicity measures, symbolic transformations, fundamental statistics, and other characteristics like extreme values, visibility graphs, physics-based simulations, and dependence on preprocessings used to a time series. Each feature in the operation set is performed with a variety of input parameters, and each parameter set produces outputs that are distinctive. A complete list of the operations produced in HCTSA is listed at https://hctsa-users.gitbook.io/hctsa-manual/readme/list-of-included-code-files.

4.3 Pearson's Correlation Coefficient

The Pearson's correlation coefficient [34] is a gauge of the linear connection between two variables. The strength and direction of the association between two variables are frequently discovered using this method. Karl Pearson, who initially suggested the idea in 1895, was honoured with the coefficient's name. In many different disciplines, including psychology, biology, economics, and engineering, to mention a few, the Pearson's correlation coefficient is often utilised.

Pearson's correlation coefficient is given by Equation 4.1.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.1}$$

where:

r is the Pearson's correlation coefficient

x_i and y_i are the i^{th} values of the two variables being compared

$\bar{x}$ and $\bar{y}$ are the means of x and y, respectively

n is the number of observations in the data

[1] r2022b, The Mathworks Inc., Massachusetts, USA

The numerator of the formula calculates the covariance between the two variables, while the denominator calculates the standard deviation of both variables. It can be interpreted as follows:

- A value of 1 indicates a perfect positive correlation. This means that the two variables move in the same direction with the same magnitude.
- A value of -1 indicates a perfect negative correlation. This means that the two variables move in opposite directions with the same magnitude.
- A value of 0 indicates no correlation. This means that there is no linear relationship between the two variables.

In this particular project, the Pearson's correlation coefficient was used to check the efficacy of the features extracted, and their relevance to our application.

4.4 Methodology

As a part of the HCTSA feature experimentation, the data was extracted in two main methods:

1. Early RP, late RP and complete (early + late) RP averaged over nine grouped electrodes of interest.
2. Complete RP averaged over nine individual electrodes of interest.

The first method is further divided into two sub-methods, based on the handling of the special values in the feature matrix generated by HCTSA.

4.4.1 Approach 1: Features Extracted from Averaged 9 Electrodes

Attempt 1: Drop All Special Valued Features

The first approach implemented the feature extraction process in a series of steps as shown in Figure 4.1. The raw data is divided into the early, late and complete RP using the time intervals as discussed earlier (-2.00 to -0.50 seconds: early RP; -0.50 to -0.00 seconds: late RP; -2.00 to 0.00 seconds: complete RP). The data from the nine electrodes of interest was averaged for all the sets. HCTSA is run to extract 7,749 features from each of the three sets,

after which the feature matrix is z-score standardized. All features with any special values (infinities or NaNs) are dropped, which result in 1200 to 1500 features being dropped across all three sets. The three are pooled together and their Pearson Correlation Coefficient is calculated against the judgement error calculated from the behavioral data.

Attempt 2: Define a Tradeoff Between Features and Data Trials

Further, another approach was implemented to handle the special values as a trial attempt. It may be possible that due to the loss of features as handled in the previous attempt, valuable data may have been lost. Hence, to maximize the number of features available for correlation, any operations with more than 15% special values were dropped and then all time series with any special values were eliminated. This procedure allowed us to retain between 500 to 600 more features across the three sets at the cost of 1300 to 1400 time series.

4.4.2 Approach 2: Features Extracted from Individual 9 Electrodes

Attempt 1: Calculated over the Complete Readiness Potential

To assess individual channel contributions in the judgement error, the individual channels were averaged for the complete RP - creating nine sets of data to extract features from. A procedure similar to the one followed in the previous section was implemented for the sets, using a 10% feature drop method to handle the special values. This approach is depicted in Figure 4.2.

Attempt 2: Calculated over the Early and Late Readiness Potential Individually

To assess individual channel contributions in the judgement error, the individual channels were averaged for the complete RP - creating nine sets of data to extract features from. A procedure similar to the one followed in the previous section was implemented for the sets, using a 10% feature drop method to handle the special values.

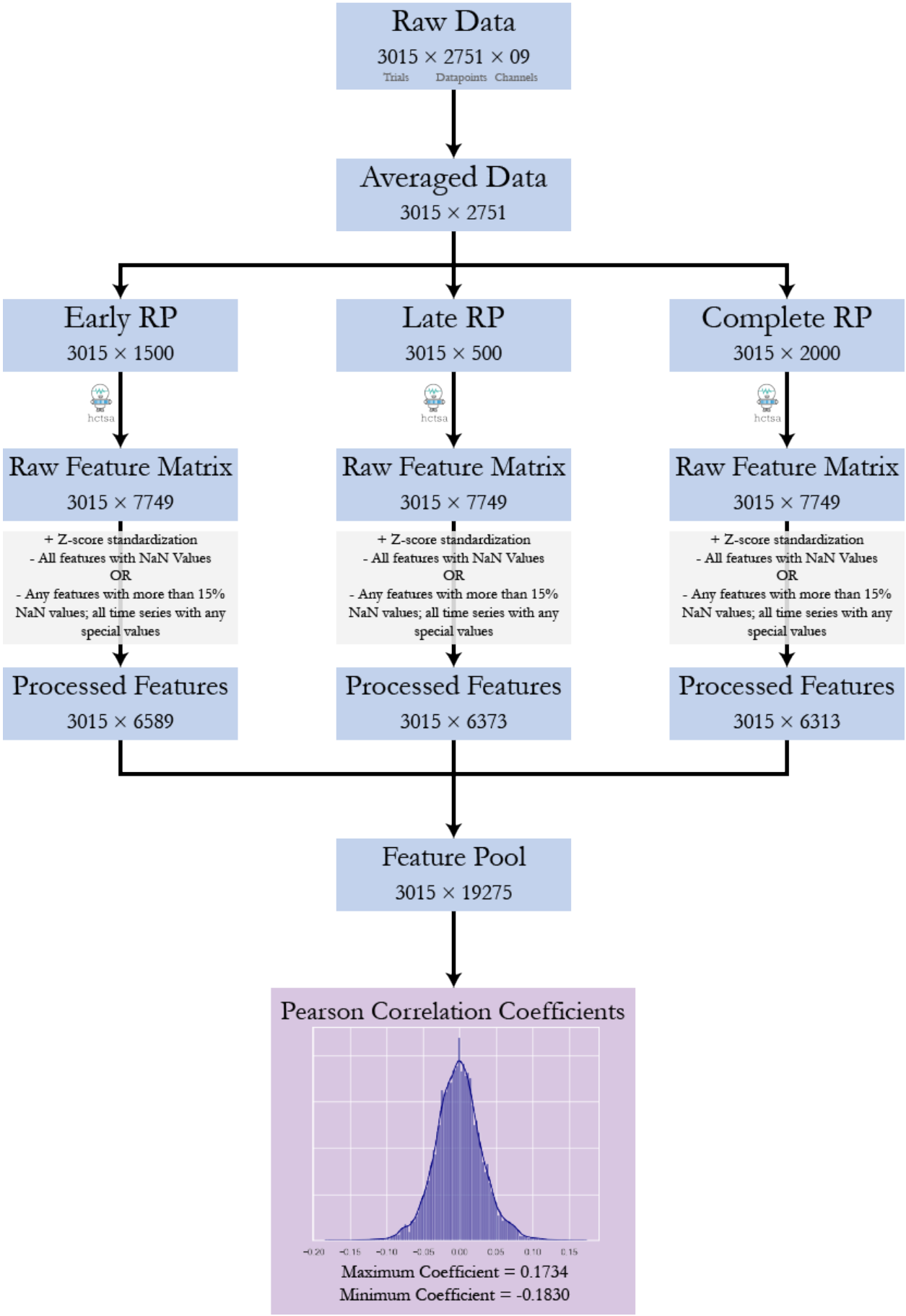


Figure 4.1: Approach 1: HCTSA Methodology for Averaged Electrode Data.

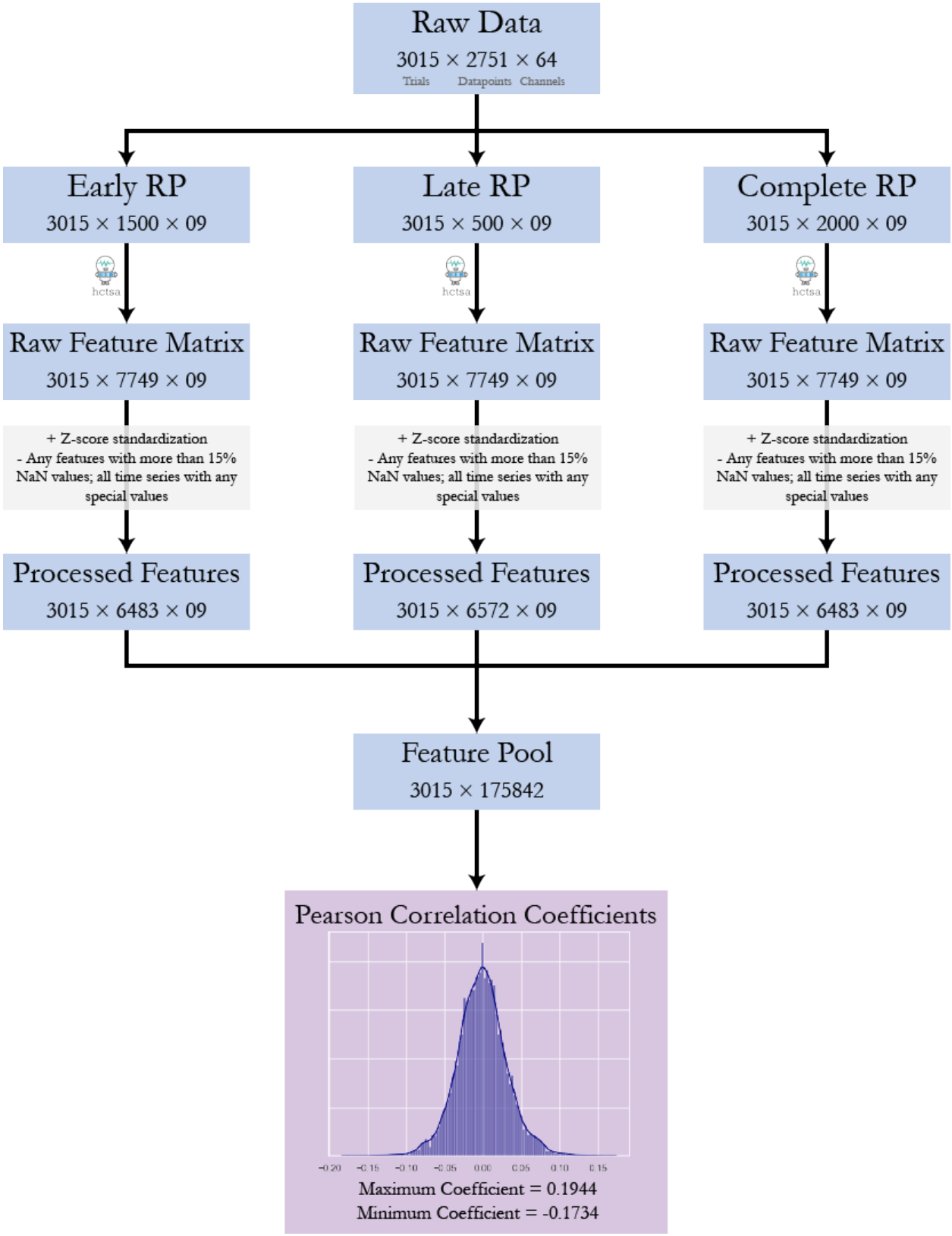


Figure 4.2: Approach 2: HCTSA Methodology for Individual Electrode Data.

Chapter 5

Methodology II: Spectral Density Estimation

A common approach to studying EEG data is to analyze it in the frequency domain. It is difficult to determine the underlying brain activity from EEG data since they are complicated and non-stationary. However, it is feasible to separate the signal into its individual frequency components by translating the signal into the frequency domain using methods like Fourier analysis or wavelet transformations. Different frequency bands of the EEG signal have been associated with specific neural processes, as mentioned in Chapter 1.

5.1 Overall Objective

The aim of this methodology is to be able to classify between different sets of probability blocks based on extracted frequency and time domain features. Choosing the correct antagonistic blocks for binary classification will be helpful to establish the existence of EEG signal markers which exist as differences between these two blocks.

For example, if a binary classifier is able to distinguish between the 75 No Feedback and 50 No Feedback blocks, it will prove that there are electrophysiological changes observed in the brain when there is a higher probability that an outcome is associated with an action, even when the outcome does not occur in a particular trial. This will strengthen the views on prospective Sense of Agency, and will provide direct proof of it's existence. Similarly, other blocks which can be compared are 75 Feedback versus 50 Feedback, Baseline No Feedback versus 75 Feedback, and Baseline No Feedback versus 50 Feedback. This is supported by a study by Haggard et al., who performed a similar paradigm to obtain these results [7].

The results for these probability blockwise comparisons are discussed in Chapter 5. Now,

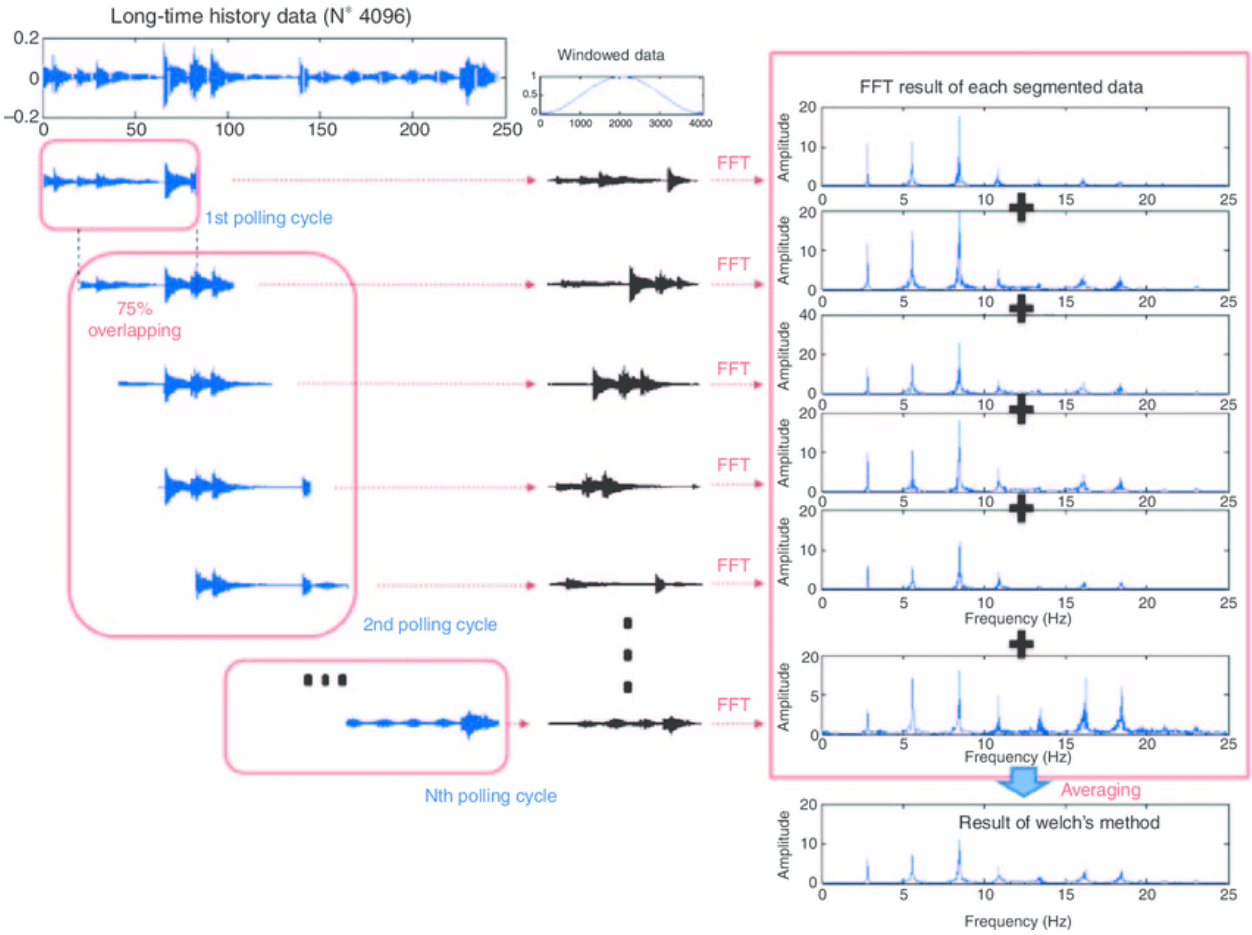


Figure 5.1: Steps for Calculation of PSD using Welch's Method. *(Courtesy: Cho et al. [36])*

the technical theoretical concepts underlying this methodology are discussed in depth.

5.2 Welch's Method

Welch's approach is a method for calculating a signal's power spectral density (PSD) [35]. It is a variant of the periodogram approach, which determines the PSD by performing the signal's autocorrelation function's Fourier transform. Welch's approach has an advantage over the periodogram in that it estimates the PSD with a smaller variance by averaging multiple periodograms of signal segments that overlap.

5.2.1 Steps for Calculating the PSD

The steps for using Welch's Method summarized in Figure 5.1:

1. Divide the EEG signal into evenly overlapping segments. The length of the signal must be longer than the overlapped length. These lengths determine the number of

segments in the length of the signal.

2. Apply a window function to each segment to reduce the spectral leakage caused by the discontinuities that arise due to the segmentation. A commonly used function is the Hamming Window function.

3. Compute the periodogram of each segment across the frequency range defined, and average the values across all segments to receive the final PSD estimate.

The formula for the Welch's Method of PSD extraction is given by Equation 5.1.

$$P_{welch}(f) = \frac{1}{K} \sum_{k=1}^{K} |X_k(f)|^2 \tag{5.1}$$

where,

$P_{welch}(f)$ is the estimated PSD at frequency f

K is the number of segments

$X_k(f)$ is the Fourier transform of the kth segment after applying the window function

The segments are assumed to have equal length L and overlap length M, and the total length of the signal is $N = KL - (K-1)M$.

5.2.2 Advantages and Precautions

Using the Welch's method has several advantages as compared to other spectral evaluation methods. They can be listed as:

- Reduced Variance: Welch's approach lowers the estimated PSD's variance by averaging many periodograms of the signal's overlapping segments. This lessens the impact of noise or other sources of variance and produces a smoother estimate of the PSD.

- Better Frequency Resolution: Welch's approach offers better frequency resolution than the periodogram method by employing overlapping segments and averaging them all together. This is especially helpful for examining narrowband signals or signals with various frequency components.

- Flexibility: The Welch's approach is a very adaptable method that may be used to a wide range of signal types, including non-stationary and irregularly sampled signals. It is also computationally effective and is implementable with straightforward techniques.

While the method has many advantages, there are a couple aspects that one should be cautious of.

- Smoothing: Although Welch's method reduces variance, it also smooths the PSD estimate. This can result in a loss of detail or sharpness in the PSD estimate, particularly for signals with rapid changes in frequency content.
- Parameter Tuning: The choice of segment length and overlap in Welch's method can have a significant impact on the resulting PSD estimate. If these parameters are not chosen carefully, they can result in biased or inaccurate estimates.
- Edge Effects: When applying a window function to each segment of the signal, Welch's method can introduce edge effects or spectral leakage. These effects can distort the PSD estimate, particularly at frequencies close to the edges of the frequency band of interest. Some techniques, such as zero-padding or tapering, can be used to mitigate these effects.

5.3 Recursive Feature Elimination

Recursive Feature Elimination (RFE) is a feature selection technique used in machine learning to select the most important features from a dataset [37]. RFE works by recursively removing features from the dataset and building a model on the remaining features until the desired number of features is reached. The importance of each feature is determined by the impact that its removal has on the performance of the model.

The basic algorithm for RFE can be described as follows:

1. Train a model on the entire dataset and determine the feature importance of each feature.
2. Remove the least important feature from the dataset.
3. Train a new model on the reduced dataset.
4. Repeat steps 2 and 3 until the desired number of features is reached.

A number of measures, such as accuracy, F1 score, or coefficient values in a linear model, may be used to assess the feature relevance. The particular situation at hand and the kind of model being used will determine the measure to be employed. The top features for

the final model can be chosen by ranking each feature's relevance. Any machine learning technique, including decision trees, logistic regression, and linear regression, that allows feature significance estimation can be used to perform RFE.

5.4 Support Vector Classifiers

Support Vector Classifiers (SVCs) [38]are a group of strong and popular machine learning algorithms that have proved effective in a variety of applications, including image classification, audio recognition, and natural language processing. SVCs are particularly effective for binary classification tasks, where the goal is to separate two classes of data points in a feature space.

SVCs are well-liked because of their capacity to handle high-dimensional datasets and resistance to noise. The data can also be mapped onto a higher-dimensional space where it becomes linearly separable in order to cope with datasets that are not linearly separable. The mathematical formulation of SVCs is based on the idea of finding the maximum margin hyperplane (MMH) that separates the two classes of data points. The MMH is unique and has the property that it maximizes the distance between the closest data points of the two classes, also known as the margin. By maximizing the margin, SVCs achieve a good generalization performance and are less prone to overfitting.

The hyperplane can be described by the equation $w \cdot x + b = 0$, where w is a vector of weights and b is a scalar bias term. The decision boundary is obtained by solving the optimization problem shown in Equation 5.2:

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \max(0, 1 - y_i(w^T \phi(x_i) + b)) \tag{5.2}$$

5.5 Random Forest Classifiers

Random Forest Classifiers are a popular machine learning algorithm that is used for classification and regression tasks. Multiple decision trees are combined using an ensemble learning technique to increase the model's generalizability and predictive accuracy [39]. Each decision tree produces a prediction, and the final prediction is obtained by taking the majority vote of all the trees in the forest. This technique is called bagging, and it helps to reduce the variance and overfitting of the model.

The decision trees in a Random Forest Classifier are constructed using a top-down ap-

proach called recursive partitioning. Recursive partitioning is a greedy algorithm that splits the data into smaller subsets based on the value of a selected feature. The goal of the algorithm is to find the feature and threshold that maximize the information gain or decrease in impurity at each split.

In addition to bagging, Random Forest Classifiers also use a technique called feature randomization. Feature randomization involves selecting a random subset of features at each split, rather than using all features. This technique helps to reduce the correlation between trees and improve the diversity of the forest. The performance of Random Forest Classifiers can be further improved by tuning hyperparameters such as the number of trees, the depth of the trees, and the number of features to be considered at each split. The optimal values of these hyperparameters can be found using techniques such as grid search or randomized search.

5.6 Differences between SVC and RFC

As discussed above, SVCs and RFCs are both powerful machine learning algorithms for classification tasks, but they are both optimal for different situations.

1. Architecture: SVCs are linear models which work by finding the best hyperplane separating the data points into different classes. On the other hand, RFCs are ensemble learning methods that combine multiple decision trees to make predictions.

2. Decision Boundaries: SVCs form linear decision boundaries, whereas RFCs define non-linear boundaries which can be more complex.

3. Handling Imbalanced Data: SVCs are more sensitive to imbalanced data and require the input datasets to be standardized or normalized. On the contrary, RFCs do not require firmly distributed data inputs and are more robust to variations in the data.

4. Interpretability of Results: The results produced by SVCs are more interpretable because the decision boundaries are linear and can be easily visualized. RFCs produce more complex decision boundaries that can be more difficult to interpret.

5. Computational Efficiency: SVCs can be faster to train than RFCs because they involves solving a quadratic programming problem to find the optimal hyperplane. In contrast, RFCs involve training multiple decision trees, which can be computationally intensive.

5.7 Methodology

In this methodology, the EEG is analyzed subject-wise using the Welch's Method for spectral density estimation. Based on these features extracted across different frequency ranges, they are attempted to be classified into the different probability blocks listed in Table 3.1.

While designing this methodology, it was observed that a total of 5 participants (Subjects 08, 10, 11, 17, 20) were not showing the behavioural results which were shown by the rest of the subjects. Hence, these five subjects were excluded and Subject 06 was included, bringing the total number of subjects to 20. The following methodology was designed based on this total.

Another useful observation kept in mind was that when observed visually, differences in regional potentials of the brain were seen 1.30 seconds before the onset of movement. Instead of focusing on a broader window, the methodology includes only the -1.30 seconds to 0.00 seconds window.

The subjectwise averaging and PSD extraction steps are executed in MATLAB r2022b[1], and the subsequent steps are implemented using Python 3.9.16[2].

5.7.1 Subjectwise Averaging

For all the following steps, the probability blocks of all the participants are first averaged across all channels to create one single EEG file per patient with dimensions $1 \times 64 \times 1300$. For a given probability block, this step results in an array of size $20 \times 64 \times 1300$ for all subjects combined.

5.7.2 Extracting PSD and Time Domain Features

Welch's Method for Frequency Domain

Next, the PSD features are extracted from the subjectwise averaged data using Welch's method. This particular methodology has been adopted from the work by Hussain et al. from 2022 [40]. The following parameters are used for the extraction process:

- Sampling Frequency: 1000 Hz
- Timescale Resolution: 0.25 seconds

[1]The Mathworks Inc., Massachusetts, USA
[2]Python Software Foundation, Delaware, USA

- Window Size Averaging: 100
- Overlap = 0

After the extraction of the PSD features using these parameters, the averages for each EEG channel are calculated over the following ranges:

- Broadband: 4 Hz to 100 Hz, excluding line noise frequencies (58 Hz to 60 Hz).
- Theta: 4 Hz to 7 Hz
- Alpha: 8 Hz to 12 Hz
- Beta: 13 Hz to 35 Hz
- Low Gamma: 36 Hz to 58 Hz
- High Gamma: 62 Hz to 100 Hz

These ranges were identified by previous studies in the domain of EEG processing [41] [42] [43]. These gave a total of 384 features per subject per probability block to work with, yielding a feature array of dimensions $20 \times 64 \times 6$. This array is flattened to 20×384 for easier processing in subsequent steps.

Area under the Curve and Slope from Time Domain

Along with the frequency domain features, two time domain features are also computed for each EEG channel. To maintain information of the generic shape of the EEG amplitude, the overall area under the curve (AUC) and slope between -1.3 seconds and 0.00 seconds is computed, giving rise to a $20 \times 64 \times 2$ feature array, which upon flattening sizes to 20×128.

5.7.3 Approach 1: PSD Features with SVC, Validated with Data from Same Standard Distribution

In the first approach, a Support Vector Classifier for the probability blocks is built using only the spectral features as shown in Figure 5.2. The flattened subjectwise feature arrays are concatenated into an array of dimensions 40×384. This data is subsequently standardized, and a randomized split of 80%/20% is performed for model training and validation, giving 32 subjects for training/testing, and 8 for validation. Using Recursive Feature Elimination,

the top 20 features of the training data array are selected for training the SVC model. A LeaveOneOut implementation leads to a 32-fold cross validation training setup for the model. The 8 subjects from the same standard distribution are used to validate the model accuracy.

5.7.4 Approach 2: PSD Features with SVC, Validated with Data from Different Standard Distribution

In this approach, the methodology is similar to the one followed in the prior one, with one major difference. Here, we perform the 80%/20% split before the standardization step and standardize the train/test and validation sets separately as shown in Figure 5.3. This step is more resembling of real-life scenarios where the model testing will not be standardized with the training data.

5.7.5 Approach 3: Adding Time Domain Features to Approach 2

Building upon the methodology in the previous approach, the RFE feature selection algorithm is fed the standardized concatenated input of the AUC, slope, and PSD features as shown in Figure 5.3. As before, top 20 features are selected.

5.7.6 Approach 4: PSD Features with RFC

Standardization is an asset when data is abundant and comes from the same source. But to build a model which can take single subject inputs without any other data points to rely upon, a way around standardization must be sought. As discussed earlier in this chapter, SVCs are powerful but require standardized inputs. RFCs on the other hand, do not have this constraint.

For this approach, a Pearson's Correlation Coefficient-based feature selection approach is implemented after the 80%/20% split as shown in Figure 5.4. For RFE to work, the input data needs to be standardized or normalized to get the correct features according to importance for the model. Else, the features with smaller numerical values will have a higher coefficient associated and vice versa. Hence, a PCC approach is preferred in this case.

With the top 25 features extracted using PCC, a RFC is model is trained at the default hyperparameters.

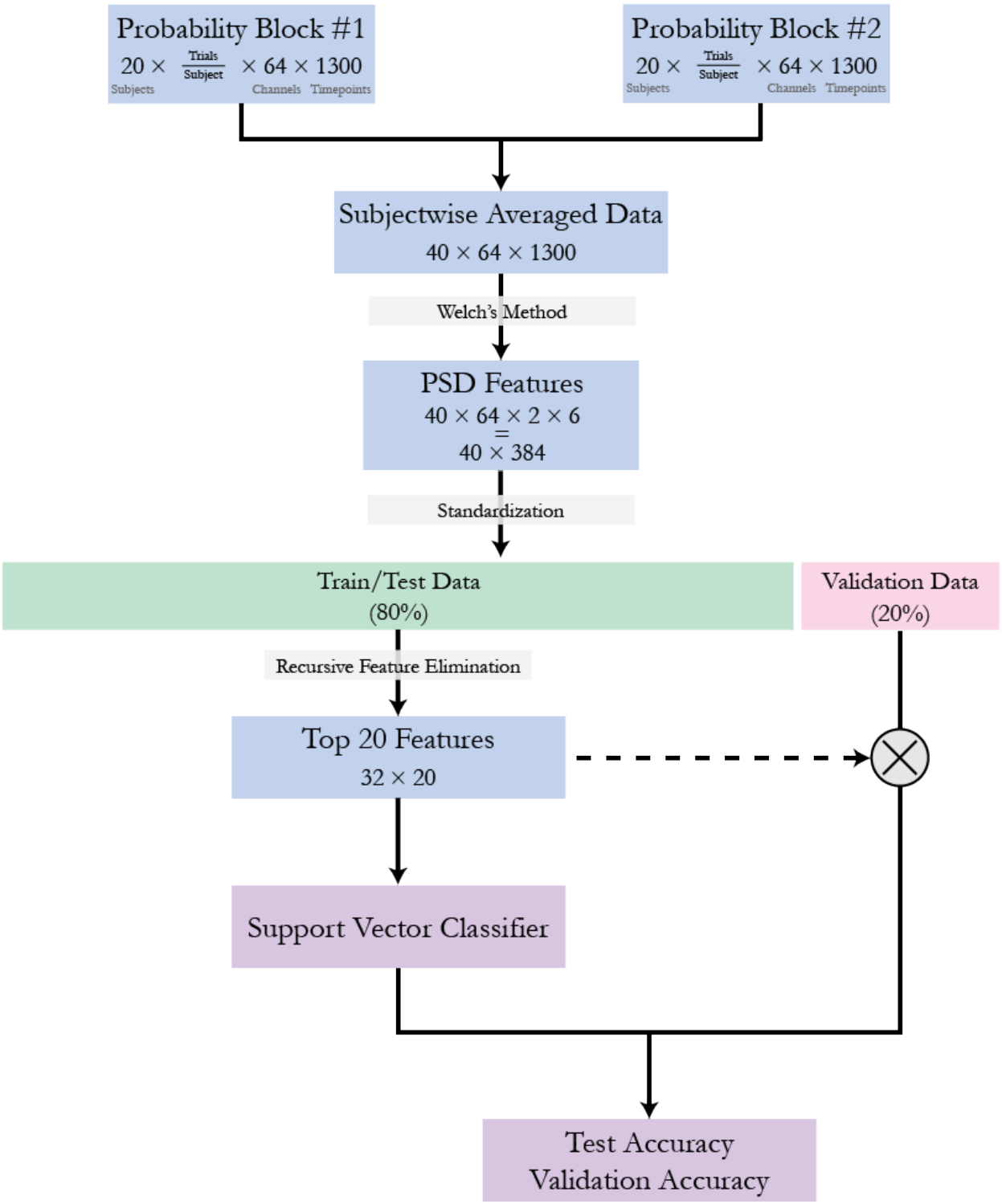


Figure 5.2: Methodology 2, Approach 1: PSD Features with SVC, Validated with Data from Different Standard Distribution.

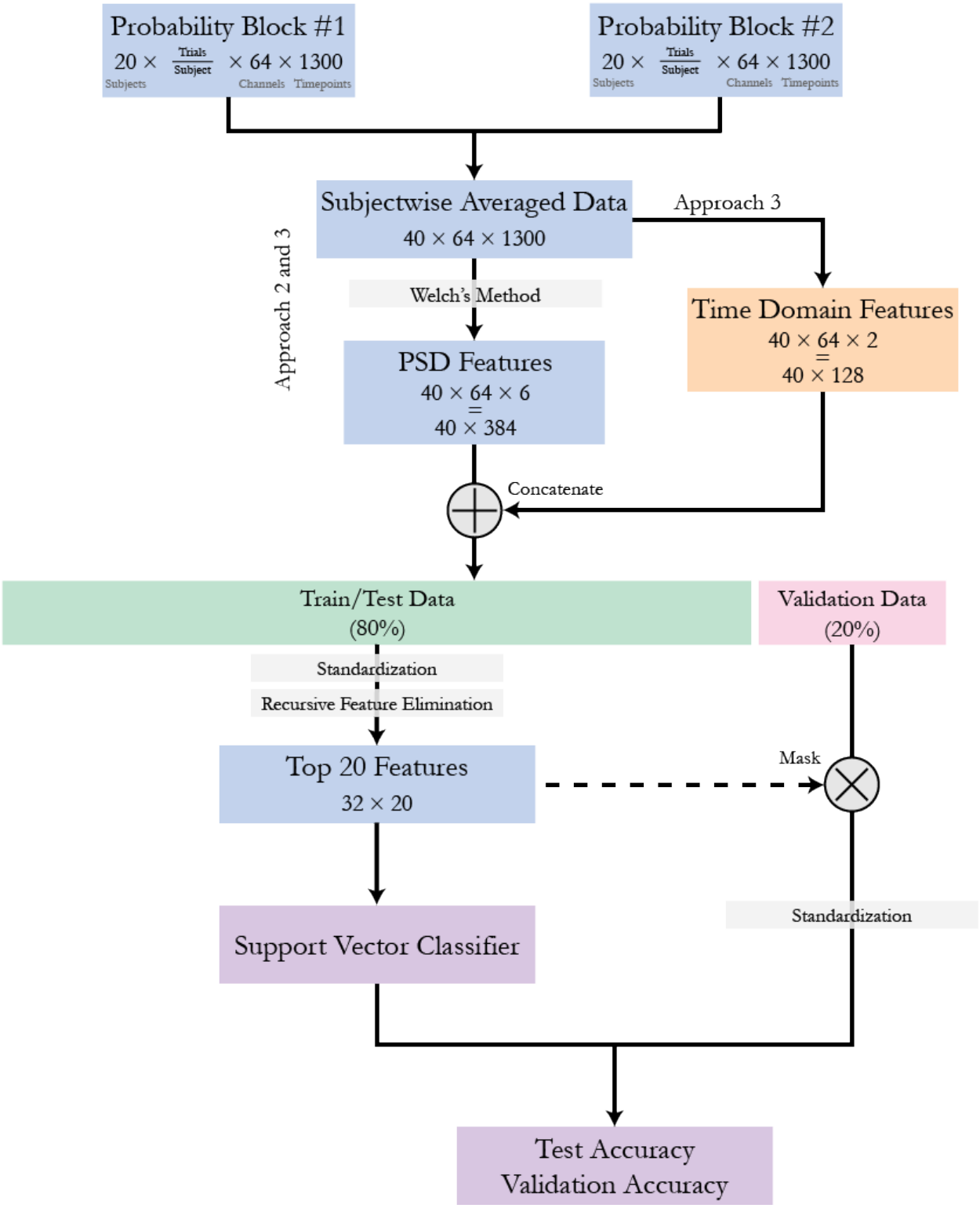


Figure 5.3: Methodology 2, Approach 2 and 3: PSD and Time Domain Features with SVC, Validated with Data from Different Standard Distribution.

5.7.7 Approach 5: PSD Features and Time Domain Features with RFC

Here, the time domain features of AUC and slope are concatenated to the feature array before the PCCs are calculated to select the top 25 features, similar to Approach 3 with respect to Approach 2 as shown in Figure 5.4.

5.7.8 Approach 6: K-Fold Nested LeaveOneOut SVC (Approach 3 Extension)

To verify the consistency of the validation results across all possible splits of train/test and validation, the LeaveOneOut algorithm is nested within a 5-Fold loop as shown in Figure 5.5. Choosing the split randomly allows all subjects to be tried as validation data atleast once, and will ensure the SVC model has picked up on the correct EEG features.

5.7.9 Approach 7: K-Fold Nested LeaveOneOut RFC (Approach 5 Extension)

Similar to Approach 6, a 5-Fold nested loop is added for the LeaveOneOut training methodology for RFCs (as mentioned in approach 5) as shown in Figure 5.5.

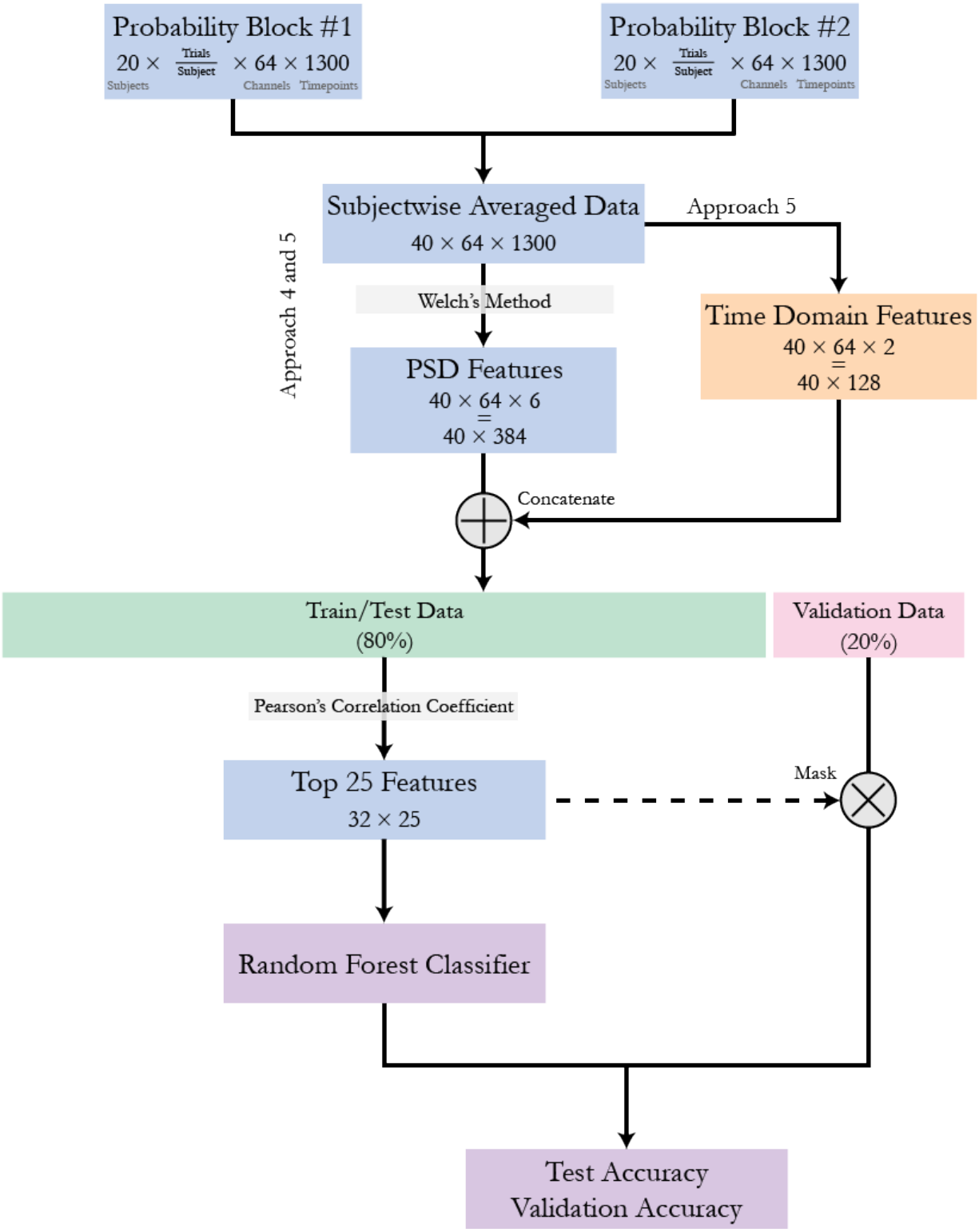


Figure 5.4: Methodology 2, Approach 4 and 5: PSD Features and Time Domain Features with RFC.

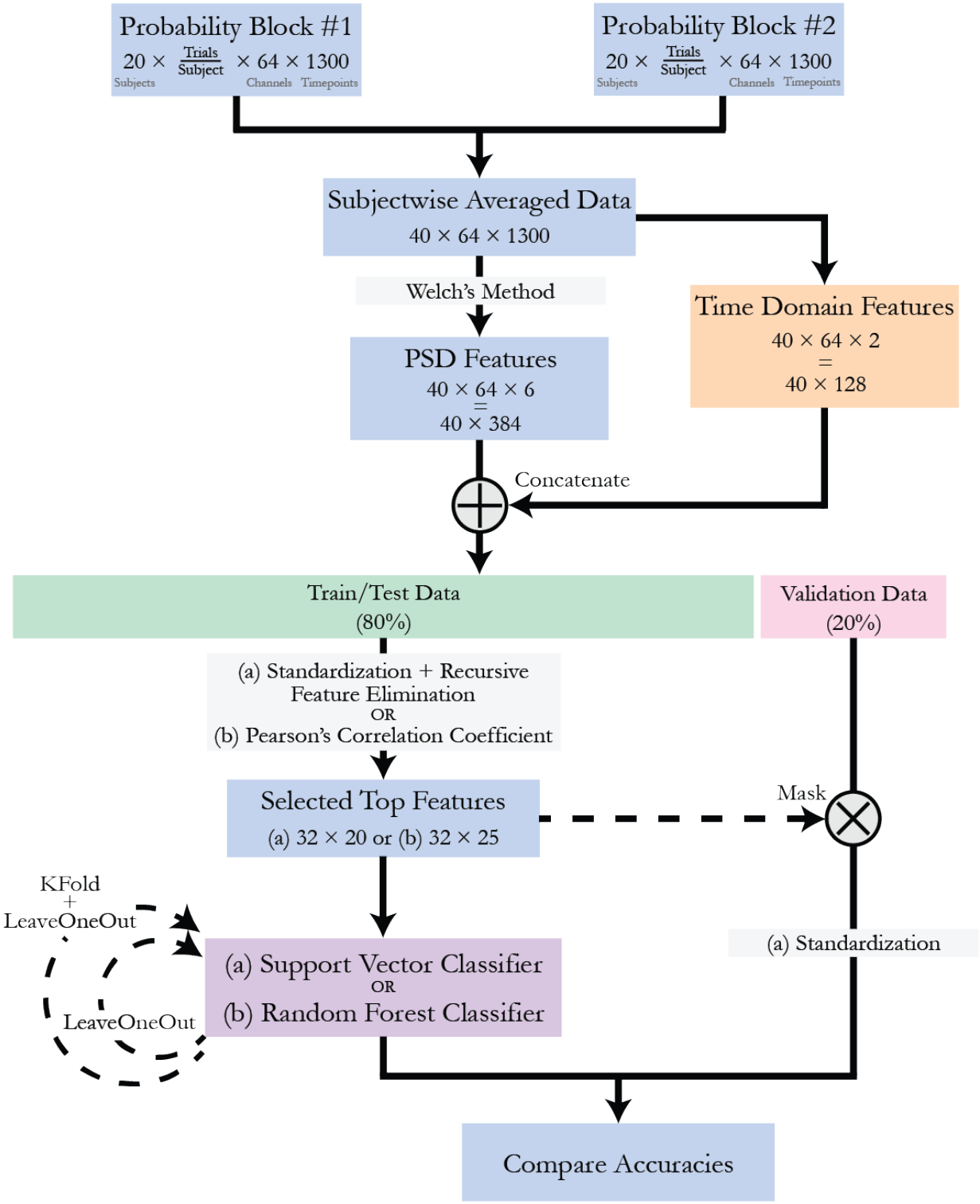


Figure 5.5: Methodology 2, Approach 6 and 7: K-Fold Nested LeaveOneOut Approach.

Chapter 6

Implementation, Results and Discussions

Both the methodologies followed produced notable results which have been discussed in detail in this chapter.

6.1 Methodology I: HCTSA

Under this methodology, the aim was to extract useful features through a Highly Comparative Time Series Analysis (HCTSA) approach. The Pearson's Correlation Coefficient was checked for all of these features against the judgement error. This method aimed on picking up characteristics important for regressor models.

6.1.1 Implementation

For all the trial-wise data processing steps and the HCTSA feature extraction, MATLAB is extensively used. The HCTSA MATLAB package by Ben Fulcher et al. [31] is utilized for the second part, and the further numerical analysis was performed in Python using `scikit-learn`.

6.1.2 Results

No notable results were obtained from this set of calculations. Across all 4 attempts of the 2 approaches, the absolute values of Pearson Correlation Coefficients were all less than 0.20, which are statistically insignificant.

A linear regression and SVM model were tried to be fit on these extracted features nonetheless, and returned a loss value of more than 10,000. This was the indicator that a more classical approach to the problem is required, as done in Methodology II.

6.2 Methodology II: Spectral Density Estimation

In this methodology, the spectral density estimates for the subjectwise EEG data was calculated over all channels and 6 frequency bands. This data was used to train classifier models to distinguish between two chosen probability blocks at a time. Unlike the results obtained from Methodology I, this approach yielded much more concrete conclusions statistically.

6.2.1 Implementation

Similar to the previous implementation, all the data processing steps are executed in MATLAB. Following this, the PSD and AUC/Slope features are also extracted in MATLAB using the `pwelch` and `trapz` functions. Once the features are extracted, they are imported to Python for the machine learning and analysis steps.

The binary classifiers are built across three groups:

1. 50 No Feedback vs 75 No Feedback
2. 50 Feedback vs 75 Feedback
3. Baseline No Feedback vs 75 Feedback

These groups are chosen as they can provide differentiating information across varying probability blocks, and can contribute to strengthening the support for prospective Sense of Agency (discussed in the subsection 'Discussions').

6.2.2 Results

All approaches listed in the Methodology II chapter yielded notable results. A table summarizing the outcomes is shown in Table 6.1.

Approach 1 uses a combined standardization and support vector classifier (SVC), excludes time domain features, performs RFE with the top 20 features, and achieves perfect median test and validation accuracy for both No Feedback conditions (50 and 75). It achieves 87.50%

Approach	Description	Time Domain Features	50% No Feedback vs 75% No Feedback	50% With Feedback vs 75% With Feedback	Baseline No Feedback vs 75% With Feedback
1	Combined Standardization + RFE (Top 20) + Support Vector	✗	MTA = 100% | VA = 87.5%	MTA = 100% | VA = 62.5%	MTA = 100% | VA = 37.5%
2	Independent Standardization + RFE (Top 20) + Support Vector	✗	MTA = 100% | VA = 75%	MTA = 100% | VA = 50%	MTA = 100% | VA = 87.5%
3	Independent Standardization + RFE (Top 20) + Support Vector	✓	MTA = 100% | VA = 100%	MTA = 100% | VA = 75%	MTA = 100% | VA = 50%
4	PCC (Top 25) + Random Forest	✗	MTA = 78.1% | VA = 37.5%	MTA = 57.8% | VA = 50%	MTA = 51.1% | VA = 50%
5	PCC (Top 25) + Random Forest	✓	MTA = 76.5% | VA = 87.5%	MTA = 60.9% | VA = 50%	MTA = 78.4% | VA = 25%
6	5-Fold Nested LeaveOneOut + Approach 3	✓	MTA = 100% | MVA = 100%	MTA = 100% | MVA = 75%	MTA = 100% | MVA = 50%
7	5-Fold Nested LeaveOneOut + Approach 5	✓	MTA = 73.3% | MVA = 75%	MTA = 64% | MVA = 50%	MTA = 72.26% | MVA = 50%

**Note: MTA = Median Test Accuracy; VA = Validation Accuracy; MVA = Median Validation Accuracy.*

Table 6.1: Summary of Results for Methodology II.

validation accuracy for the 50 No Feedback condition and 62.50% validation accuracy for the 75 No Feedback condition.

Approach 2 uses independent standardization and SVC, excludes time domain features, and performs RFE with the top 20 features. It achieves perfect median test accuracy for the 50 No Feedback versus 75 No Feedback condition, and yields an accuracy of 75% for validation.

Approach 3 is similar to approach 2, but includes time domain features. It also achieves perfect median test accuracy for all conditions and perfect validation accuracy for the 50 No Feedback versus 75 No Feedback condition. However, the validation accuracy drops to 75.00% and 50.00% for the 75 No Feedback and 75 Feedback conditions, respectively.

Approach 4 uses random forest, excludes time domain features, and selects the top 25 features based on Pearson correlation coefficient. It achieves median test accuracies ranging from 51.10% to 78.10% and validation accuracies ranging from 37.50% to 50.00%, shown in the table.

Approach 5 is similar to approach 4, but includes time domain features. It achieves median test accuracies ranging from 60.90% to 76.50% and validation accuracies ranging from 25.00% to 87.50%.

Approach 6 uses a 5-fold leave-one-out nested approach with SVC, includes time domain features, and performs RFE with the top 20 features. It achieves perfect median test for all conditions, with the validation accuracies ranging between 50.00% to 100.00%.

Approach 7 uses a 5-fold leave-one-out nested approach with random forest, includes time domain features, and selects the top 25 features based on Pearson correlation coefficient. It achieves median test accuracies ranging from 64.00% to 73.30% and validation accuracies ranging from 50.00% to 75.00%.

6.2.3 Feature Analysis

The features which were selected for the SVC and RFC approach were also analyzed, and patterns were observed. The summary of the features extracted from each channel, as well as the frequency band or AUC/slope are shown in Tables 6.2 and 6.3.

6.3 Discussions

Based on the results obtained, many relevant inferences can be drawn.

50 No Feedback vs 75 No Feedback (RFE for SVC)							
Frequency Domain (PSD)				Time Domain (AUC/Slope)			
Electrodes		Frequency Bands		Electrodes		Frequency Bands	
C4	1	Alpha	3	AF3	1	Area	6
F5	1	Beta	1	AF7	1	Slope	5
FT7	1	Broadband	2	AF8	1		
P1	1	Theta	3	F1	1		
P3	1			F2	1		
T7	1			F3	1		
T8	1			F6	1		
TP10	2			Fp1	1		
				Fz	1		
				Pz	2		

50 Feedback vs 75 Feedback (RFE for SVC)							
Frequency Domain (PSD)				Time Domain (AUC/Slope)			
Electrodes		Frequency Bands		Electrodes		Frequency Bands	
C1	2	Alpha	5	Afz	1	Area	3
C4	2	Broadband	2	F3	1	Slope	1
CP1	1	High Gamma	2	Fz	2		
Cz	2	Low Gamma	5				
F6	1	Theta	2				
F7	2						
FC2	1						
FC4	1						
FT10	1						
P2	1						
PO4	1						
POz	1						

Baseline vs 75 Feedback (RFE for SVC)							
Frequency Domain (PSD)				Time Domain (AUC/Slope)			
Electrodes		Frequency Bands		Electrodes		Frequency Bands	
C1	1	Alpha	1	AF8	1	Area	3
C4	1	Beta	1	F1	1	Slope	4
CP1	1	Broadband	4	F2	1		
F2	1	High Gamma	2	F4	1		
FC2	1	Low Gamma	2	F5	2		
FC3	1	Theta	3	Pz	1		
FC4	1						
Fz	1						
P1	1						
P2	1						
P4	1						
POz	1						
T7	1						

Table 6.2: Summary of Features Extracted for SVC Binary Classifiers.

50 No Feedback vs 75 No Feedback (PCC for Random Forest)							
Frequency Domain (PSD)				Time Domain (AUC/Slope)			
Electrodes		Frequency Bands		Electrodes		Frequency Bands	
F3	1	Alpha	1	AF7	1	Area	5
AF7	1	Beta	5	F2	1	Slope	5
Afz	2	Broadband	1	F6	6		
F4	1	High Gamma	2	F8	2		
F7	1	Low Gamma	2				
F8	2	Theta	4				
FC4	1						
FT8	1						
O1	1						
Oz	1						
P7	2						
PO4	1						

50 Feedback vs 75 Feedback (PCC for Random Forest)					
Frequency Domain (PSD)		Time Domain (AUC/Slope)			
Electrodes	Frequency Bands	Electrodes		Frequency Bands	
-	-	AF3	2	Area	13
		AF7	2	Slope	12
		AF8	2		
		CPz	3		
		Cz	2		
		F1	1		
		F2	1		
		F3	1		
		F4	2		
		F6	6		
		F8	2		
		Fp1	1		

Baseline vs 75 Feedback (PCC for Random Forest)							
Frequency Domain (PSD)				Time Domain (AUC/Slope)			
Electrodes		Frequency Bands		Electrodes		Frequency Bands	
AF3	2	Alpha	4	AF3	1	Area	4
CP3	1	Beta	1	AF7	1	Slope	5
CP5	1	Broadband	4	F6	1		
F4	1	High Gamma	2	F8	1		
F6	1	Low Gamma	3				
F8	1	Theta	2				
FT10	2						
FT8	1						
Fp2	1						
O1	1						
O2	1						
P5	1						
P7	1						

Table 6.3: Summary of Features Extracted for RFC Binary Classifiers.

6.3.1 HCTSA vs Welch's Method

Comparing the success of the two methodologies, it can be seen that a Spectral Density Approach is more relevant than HCTSA to the processing of EEG data in this case. This is a preliminary observation and to confirm it's outcome, the first methodology will have to be modified so as to include subjectwise data similar to the approach in the second methodology. It is highly possible that HCTSA may perform at par or better than Welch's Method under similar circumstances due to the extensiveness of the package.

6.3.2 Probability Block Classification using PSD

The comparisons of the second methodology are key to achieving our aim of supporting the observations of prospective Sense of Agency.

For the comparison between 50 Feedback and 75 Feedback conditions, it is expected that the 75 Feedback block should show a higher intentional binding effect as compared to the other. This is because here, the outcome is more likely to occur, and hence it is more likely to be associated with the action performed by the subject. It is reasonable to expect visible differences between the two conditions. The SVC classifier was able to pick up features well when the AUC and Slope features were added. It is seen that majority of the frequency features arose from the centroparietal region in the alpha and low gamma bands. The RFC did not perform as well on this block, and latched onto only the time domain features during classification.

In the 50 No Feedback versus 75 No Feedback, we obtained the results with highest accuracies. Both the SVC and RFC classifiers were able to pick up features in the EEG that distinguished them excellently, without overfitting as proven through the K-Fold nesting test. This demonstrates that even when the conclusion does not materialise, a potential element of Sense of Agency is still present. This is compelling evidence that the neural activity before an action differs significantly when a particular outcome is more likely to be connected to it. Upon examining the binding effects, it is also seen that the judgement error is lower (albeit marginally) for the subjects in the 75 No Feedback block than the 50 No Feedback block. This supports the view that the experience of one's own actions are fundamentally linked to the predictions of the outcomes and not just the retrospective inferences made after the outcome [44]. This goes hand-in-hand with the prospective Sense of Agency observations by Moore et al. [15].

It is observed through the results in Figure 6.1 that the SVC classifier has outperformed

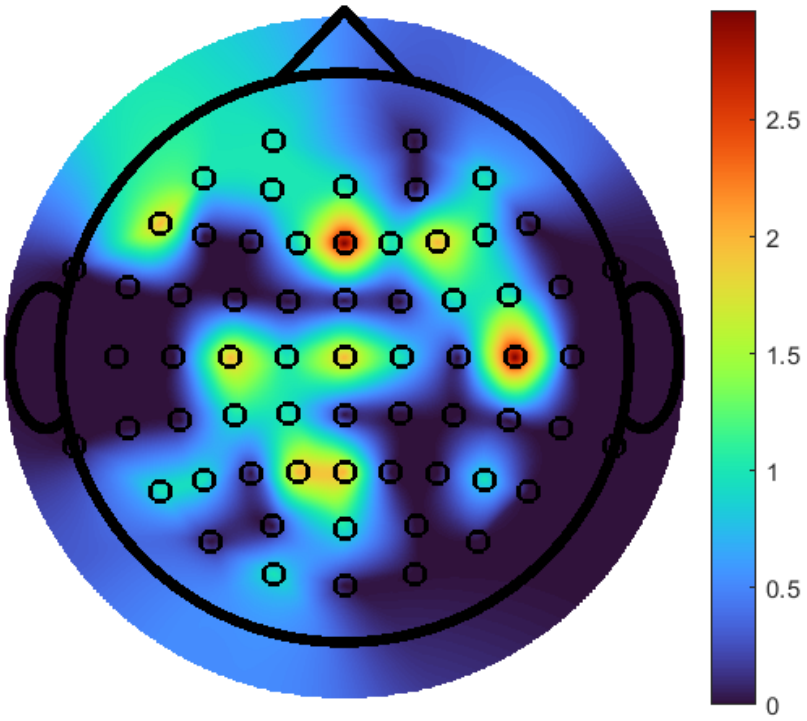


Figure 6.1: Visualization of the Electrodes Used to Train the SVC Model (50NoFBvs75NoFB + 50FBvs75FB).

the RFC classifier for the 50 No Feedback vs 75 No Feedback classification block. Upon examination of the features picked up by the RFE selector, it is seen that they originate mainly from the centroparietal region of the brain as shown in Figure 6.1. These may be features predominantly correlated to the Bereitschaftspotential sourced in the pre-SMA region (Figure 1.1). This observation supports the experimental results of Moore et al. from a 2010 study [3]. Upon introducing a disturbance to the pre-SMA region using transcranial magnetic stimulation, a reduction in the intentional binding was observed. The results observed in this project strengthen the hypothesis emphasizing the importance of the importance of the pre-SMA region for generating a Sense of Agency.

6.4 Future Directions

Building upon the research conducted in this project, several future directions are possible to be explored.

- The features extracted show that various frequency bands have been extracted from different regions of the brain for training the machine learning models. These results show that there must be cross-frequency coupling between these signals, and that the brain behaves much more complexly than through just defined frequency bands. This is an upcoming field of research in neuroscience and can be studied further in this project.

- Predicting the judgement error using subjectwise data, i.e. building a regressor for the data.

- Using HCTSA to extract features subjectwise, and compare the performance to the Welch's Method results.

- Another approach is to work with the data in the frequency domain, to generate spectrograms and use Convolutional Neural Networks (CNNs) as regressors or classifiers. This approach will help recognize any patterns present in the data, and may have better outputs than machine learning based approaches.

Chapter 7

Summary and Conclusions

Sense of Agency is the experience of feeling control over one's own actions. This cognitive phenomenon is highly subjective, and the core causes of it are a central question in neuroscience.

The source of this experience can be attributed to two components - retrospective and prospective Sense of Agency. Feedback signals which arise due to the perception of a causal outcome are the source of the retrospective component. This component banks on the inference that the brain creates a Sense of Agency based on inferences of it's own actions after the completion of it. The prospective component, on the other hand, suggests that the brain's Sense of Agency is also associated with the prediction of outcomes, and not only the retrospective inferences. This has been observed in only a handful of studies so far, and is the crux of this project.

This project aimed to strengthen the literature supporting the prospective component of Sense of Agency, and use a machine learning approach to show differences in neural activity before the onset of movement.

In the first methodology, trialwise EEG data is processed, and 7000+ features are extracted from it using a HCTSA approach. The Pearson's Correlation Coefficient for the features were calculated against the judgement error, but unfortunately none of the features showed statistical significance. Subsequently, in the second methodology, PSD features of the subjectwise averaged EEG data were calculated across six frequency bands using Welch's Method. Along with these frequency domain features, two time domain features (area under the curve and slope) were calculated for all the channels of the subjectwise data. Together, these were used to train two types of machine learning classifiers - Support Vector Classifiers and Random Forest Classifiers.

These machine learning models were used to differentiate between three pairs of proba-

bility blocks which had the potential to showcase differences in neural activity across these blocks. These were Baseline No Feedback vs 75 Feedback, 50 Feedback vs 75 No Feedback, and 50 No Feedback vs 75 No Feedback.

Of the three mentioned binary classification problems, the highest differences were noticed in the 50 No Feedback vs 75 No Feedback blocks. Upon examining the intentional binding effects, it was also observed that the judgement error is lower for the subjects in the 75 No Feedback block than the 50 No Feedback block. This bolsters the idea that one's own actions and experiences are essentially related to predictions of outcomes rather than merely judgements drawn in hindsight after the fact.

We can conclude from this study that a prospective component of Sense of Agency plays a key role in generating a Sense of Agency. Even on trials where the action had no effect, temporal binding took place when predictability of the consequence of the action was high. The results of the followed methodologies align with the outcomes of previous studies in this domain, and further analysis will lead to a solid conclusion for the same.

Bibliography

[1] P. Haggard, "Sense of agency in the human brain," *Nature Reviews Neuroscience*, vol. 18, pp. 196–207, Mar. 2017.

[2] R. Vastano, E. Ambrosini, J. L. Ulloa, and M. Brass, "Action selection conflict and intentional binding: An ERP study," *Cortex*, vol. 126, pp. 182–199, May 2020.

[3] J. W. Moore, D. Ruge, D. Wenke, J. Rothwell, and P. Haggard, "Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the sense of agency," *Proceedings of the Royal Society B: Biological Sciences*, vol. 277, pp. 2503–2509, Apr. 2010.

[4] M. Brass and P. Haggard, "The hidden side of intentional action: the role of the anterior insular cortex," *Brain Structure and Function*, vol. 214, pp. 603–610, May 2010.

[5] Q. Welniarz, Y. Worbe, and C. Gallea, "The forward model: A unifying theory for the role of the cerebellum in motor control and sense of agency," *Frontiers in Systems Neuroscience*, vol. 15, Apr. 2021.

[6] N. Sidarus and P. Haggard, "Difficult action decisions reduce the sense of agency: A study using the eriksen flanker task," *Acta Psychologica*, vol. 166, pp. 1–11, May 2016.

[7] P. Haggard, S. Clark, and J. Kalogeras, "Voluntary action and conscious awareness," *Nature Neuroscience*, vol. 5, pp. 382–385, Mar. 2002.

[8] M. Ruess, R. Thomaschke, and A. Kiesel, "Intentional binding of visual effects," *Attention, Perception, &amp Psychophysics*, vol. 80, pp. 713–722, Jan. 2018.

[9] H.-G. Jo, M. Wittmann, T. Hinterberger, and S. Schmidt, "The readiness potential reflects intentional binding," *Frontiers in Human Neuroscience*, vol. 8, June 2014.

[10] H. Shibasaki and M. Hallett, "What is the bereitschaftspotential?," *Clinical Neurophysiology*, vol. 117, pp. 2341–2356, Nov. 2006.

[11] H.-G. Jo, M. Wittmann, T. L. Borghardt, T. Hinterberger, and S. Schmidt, "First-person approaches in neuroscience of consciousness: Brain dynamics correlate with the intention to act," *Consciousness and Cognition*, vol. 26, pp. 105–116, May 2014.

[12] D. M. Wolpert and Z. Ghahramani, "Computational principles of movement neuroscience," *Nature Neuroscience*, vol. 3, pp. 1212–1217, Nov. 2000.

[13] I. Fried, A. Katz, G. McCarthy, K. Sass, P. Williamson, S. Spencer, and D. Spencer, "Functional organization of human supplementary motor cortex studied by electrical stimulation," *The Journal of Neuroscience*, vol. 11, pp. 3656–3666, Nov. 1991.

[14] P. Haggard, "Conscious intention and motor cognition," *Trends in Cognitive Sciences*, vol. 9, pp. 290–295, June 2005.

[15] J. Moore and P. Haggard, "Awareness of action: Inference and prediction," *Consciousness and Cognition*, vol. 17, pp. 136–144, Mar. 2008.

[16] J. W. Moore, "What is the sense of agency and why does it matter?," *Frontiers in Psychology*, vol. 7, Aug. 2016.

[17] E. Daprati, N. Franck, N. Georgieff, J. Proust, E. Pacherie, J. Dalery, and M. Jeannerod, "Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients," *Cognition*, vol. 65, pp. 71–86, Dec. 1997.

[18] N. Franck, C. Farrer, N. Georgieff, M. Marie-Cardine, J. Daléry, T. d'Amato, and M. Jeannerod, "Defective recognition of one's own actions in patients with schizophrenia," *American Journal of Psychiatry*, vol. 158, pp. 454–459, Mar. 2001.

[19] A. Gentsch, S. Schütz-Bosbach, T. Endrass, and N. Kathmann, "Dysfunctional forward model mechanisms and aberrant sense of agency in obsessive-compulsive disorder," *Biological Psychiatry*, vol. 71, pp. 652–659, Apr. 2012.

[20] E. Oren, N. Friedmann, and R. Dar, "Things happen: Individuals with high obsessive–compulsive tendencies omit agency in their spoken language," *Consciousness and Cognition*, vol. 42, pp. 125–134, May 2016.

[21] O. G. Brim, C. D. Ryff, and R. C. Kessler, *The adaptive value of feeling in control during midlife,*, p. 320–349. University of Chicago Press, 2004.

[22] J. Metcalfe, T. S. Eich, and A. D. Castel, "Metacognition of agency across the lifespan," *Cognition*, vol. 116, pp. 267–282, Aug. 2010.

[23] B. Shneiderman, *Designing the user interface: Strategies for effective human-computer interaction*. Addison-Wesley, 1998.

[24] D. Coyle, J. Moore, P. O. Kristensson, P. Fletcher, and A. Blackwell, "I did that! measuring users' experience of agency in their own actions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, May 2012.

[25] H. Limerick, D. Coyle, and J. W. Moore, "The experience of agency in human-computer interactions: a review," *Frontiers in Human Neuroscience*, vol. 8, Aug. 2014.

[26] P. Haggard and M. Tsakiris, "The experience of agency," *Current Directions in Psychological Science*, vol. 18, pp. 242–246, Aug. 2009.

[27] C. D. Frith, "Action, agency and responsibility," *Neuropsychologia*, vol. 55, pp. 137–142, Mar. 2014.

[28] S. Nichols, "Experimental philosophy and the problem of free will," *Science*, vol. 331, pp. 1401–1403, Mar. 2011.

[29] B. Libet, C. A. Gleason, E. W. Wright, and D. K. Pearl, "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential)," *Brain*, vol. 106, no. 3, pp. 623–642, 1983.

[30] B. D. Fulcher, M. A. Little, and N. S. Jones, "Highly comparative time-series analysis: the empirical structure of time series and their methods," *Journal of The Royal Society Interface*, vol. 10, p. 20130048, June 2013.

[31] B. D. Fulcher and N. S. Jones, "hctsa : A computational framework for automated time-series phenotyping using massive feature extraction," *Cell Systems*, vol. 5, pp. 527–531.e3, Nov. 2017.

[32] Pooja, S. Pahuja, and K. Veer, "Recent approaches on classification and feature extraction of EEG signal: A review," *Robotica*, vol. 40, pp. 77–101, May 2021.

[33] S. Bode, E. Schubert, H. Hogendoorn, and D. Feuerriegel, "Decoding continuous variables from event-related potential (ERP) data with linear support vector regression

using the decision decoding toolbox (DDTBOX)," *Frontiers in Neuroscience*, vol. 16, Nov. 2022.

[34] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, pp. 37–40, Springer, 2009.

[35] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, pp. 70–73, June 1967.

[36] S. Cho, J. P. Lynch, J.-J. Lee, and C.-B. Yun, "Development of an automated wireless tension force estimation system for cable-stayed bridges," *Journal of Intelligent Material Systems and Structures*, vol. 21, pp. 361–376, Oct. 2009.

[37] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik *Machine Learning*, vol. 46, no. 1/3, pp. 389–422, 2002.

[38] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, Sept. 1995.

[39] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1, pp. 278–282, IEEE, 1995.

[40] S. J. Hussain and R. Quentin, "Decoding personalized motor cortical excitability states from human electroencephalography," *Scientific Reports*, vol. 12, Apr. 2022.

[41] C. Zrenner, D. Desideri, P. Belardinelli, and U. Ziemann, "Real-time EEG-defined excitability states determine efficacy of TMS-induced plasticity in human motor cortex," *Brain Stimulation*, vol. 11, pp. 374–389, Mar. 2018.

[42] S. J. Hussain, M. K. Vollmer, J. Stimely, G. Norato, C. Zrenner, U. Ziemann, E. R. Buch, and L. G. Cohen, "Phase-dependent offline enhancement of human motor memory," *Brain Stimulation*, vol. 14, pp. 873–883, July 2021.

[43] S. J. Hussain, L. Claudino, M. Bönstrup, G. Norato, G. Cruciani, R. Thompson, C. Zrenner, U. Ziemann, E. Buch, and L. G. Cohen, "Sensorimotor oscillatory phase–power interaction gates resting human corticospinal output," *Cerebral Cortex*, vol. 29, pp. 3766–3777, Oct. 2018.

[44] M. Voss, J. Moore, M. Hauser, J. Gallinat, A. Heinz, and P. Haggard, "Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences," *Brain*, vol. 133, pp. 3104–3112, Aug. 2010.