Exploring Classification Methods for Motor Imagery and Execution EEG Signal Fluctuations

Pragati Dode

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Erika Parsons

Afra Mashhadi

Pierre Mourad

Program Authorized to Offer Degree:

Computer Science & Software Engineering

University of Washington

Abstract

Exploring Classification Methods for Motor Imagery
and Execution EEG Signal Fluctuations

Pragati Dode

Chair of the Supervisory Committee:
Erika Parsons
School of Science, Technology, Engineering & Mathematics

Brain-Computer Interfaces (BCIs) offer promising applications in neurological rehabilitation through motor imagery (MI)-based training, which is the "intent" of performing an action. This research addresses the challenge of accurately classifying MI and motor execution (ME) based on Electroencephalography (EEG) signals. This kind of data is often limited by subject variability, non-stationarity, environmental noise during data collection, EEG device quality, and small dataset sizes. For our study, we propose to make use of an external large dataset, including data from 103 subjects (compared to the 9–12 subject datasets used in prior work). One of the main goals of this research is to integrate multiple feature extraction techniques spanning time, frequency, and spatial domains. Effective EEG channel selection was guided by fMRI studies identifying MI- and ME-relevant Brodmann areas, combined with EEG-based statistical analysis, resulting in a refined set of 12 informative electrodes. Several machine learning models (SVM, RF, KNN, XGBoost, MLP) are evaluated, achieving up to 80% accuracy with improved robustness across subjects. These findings demonstrate enhanced generalizability and support the development of more reliable BCI applications for real-world rehabilitation scenarios.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to Professor Parsons for her unwavering support and guidance throughout this research. Her twice-weekly meetings were instrumental in keeping me focused and progressing in the right direction, and her thoughtful feedback greatly enriched the writing and documentation phases. I am also sincerely thankful to Professor Mourad, whose expertise in neurological research was particularly valuable during the early stages of investigation and data preprocessing. I extend my appreciation to Professor Mashhadi for serving on my committee and for her support throughout the process. Finally, I am truly honored to have had the opportunity to be part of the Smart NeuroRehab Ecosystem research group under Professor Parsons' mentorship.

DEDICATION

To my husband, Amol, and our children, Arus and Prisha —
your love, patience, and encouragement made this journey possible.

Chapter 1

INTRODUCTION

The brain is the most vital organ in the human body, controlling essential functions such as breathing, movement, cognition, and memory. Its immense complexity makes it difficult to fully comprehend, as it consists of billions of neurons and glial cells [92]. To study brain activity, researchers typically rely on two primary methods: analyzing the electrical signals generated by neurons or using brain imaging techniques.

Among the most promising advancements in neural research is the development of Brain-Computer Interfaces (BCIs) [79], which enable direct communication between the brain and external devices. BCIs can be categorized into three types: invasive, where electrodes are implanted within the cerebral cortex for high-precision neural recording; semi-invasive, where signals are measured directly from the cortical surface; and non-invasive, where brain activity is recorded externally from the scalp, typically through Electroencephalography (EEG). EEG-based BCIs have become especially popular due to their portability, safety, affordability, and ease of use. Recent advancements in machine learning [5] have further enhanced the reliability of EEG-based BCIs, broadening their applications in neuroscience, rehabilitation, and assistive technologies.

A critical area of application of BCIs is in some aspects of neurological rehabilitation, which aims to restore lost skills or develop new ones through structured practice and adaptation. While other types of rehabilitation therapies rely on active physical exercises to generate sensory feedback for motor recovery, research has shown that physical movement is not always necessary. **Motor Imagery** (MI) is the mental rehearsal of a movement without actual execution—activates brain regions involved in actual movement, particularly the primary motor cortex, supplementary motor area, and pre-motor cortex. This has been

demonstrated in both EEG-based studies from the computer science domain [54, 43, 21] and in neuro-imaging research using fMRI techniques [84, 16, 43, 21, 83, 61]. These findings provide robust evidence that MI activates the motor system in a manner similar to ME, inducing neuro-plastic changes that support its role as a potent non-invasive tool in rehabilitation.

The research work presented in this thesis is part of the larger project "Smart NeuroRehab Ecosystem" (SNeuroRehab) [78][12] at the University of Washington Bothell Computing and Software Systems, lead by Prof. Erika Parsons. This project proposes the application of modern technology trends such as Edge Computing (EC), Electroencephalopgraphy and Electromyography, Computer Vision, Virtual and Augmented Reality (AR/VR), Machine Learning and Artificial Intelligence (ML/AI) and Data Science to develop strategies that can aid with the rehabilitation of some certain patients[23][62].

Figure 1.1 presents an architectural overview of the Smart NeuroRehab Ecosystem, including the integration of the methodology developed in this study. Initially, the project focused on designing rehabilitation solutions for stroke patients suffering from Spatial Neglect – a condition characterized by impaired perception and response to stimuli on one side of space [7] – as well as upper extremity motor weakness [4].

Although this thesis utilizes an external dataset, it is well aligned with the objectives of the Smart NeuroRehab project. The dataset[31] features upper-limb motor tasks, which are particularly relevant to the group's long-term research goals. It was collected by the developers of the BCI2000 system—an open-source, general-purpose software platform widely used in brain-computer interface (BCI) research. Contributed to PhysioNet, the dataset comprises over 1,500 high-quality EEG recordings capturing both motor imagery (MI) and motor execution (ME) tasks. These recordings were acquired using a medical-grade 64-channel EEG system, ensuring high signal fidelity and strong clinical relevance.

Analyzing this dataset in the current study has provided valuable insights that have guided the development of our own enhanced data acquisition protocols. These protocols form the foundation for future experimental efforts within the Smart NeuroRehab initiative.

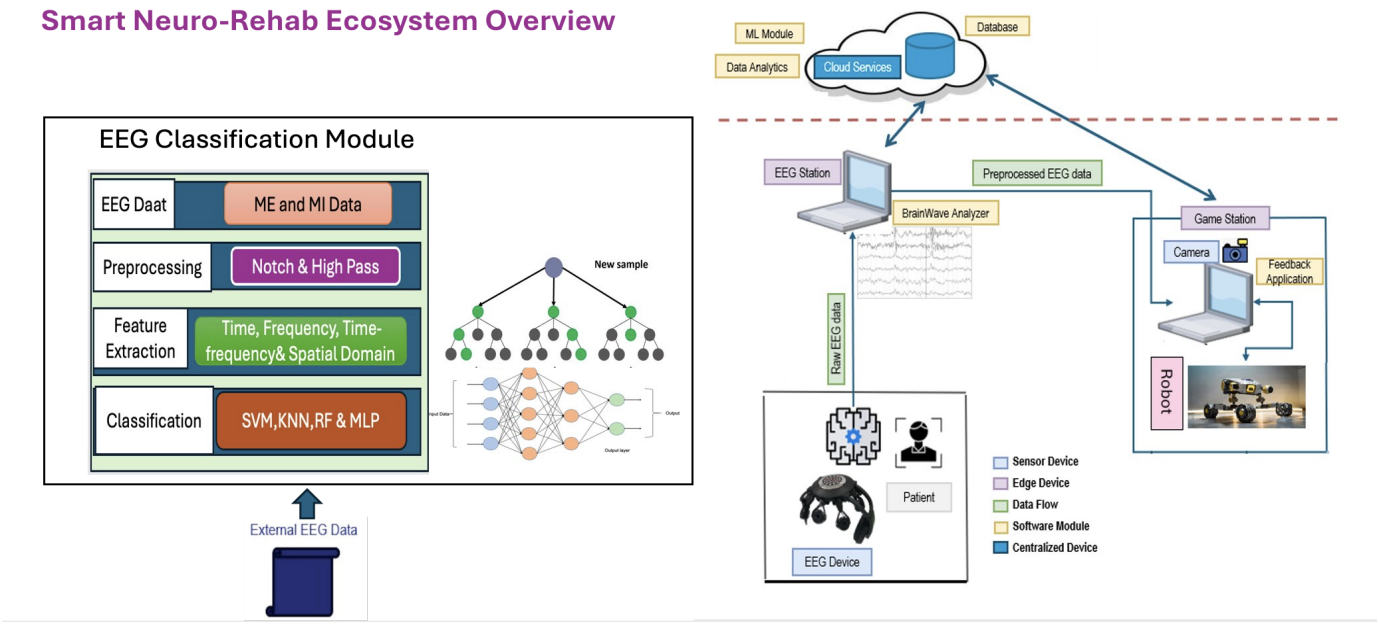


Figure 1.1: Integration of the Proposed Study within the Smart Neuro-Rehabilitation Ecosystem. The components inside the EEG Classification Module constitute the work presented in this thesis.

1.1 Problem Statement

Despite the progress in EEG-based BCIs, accurately distinguishing between Motor Imagery (MI) and Motor Execution (ME) remains a significant challenge. EEG signals are highly non-stationary, prone to noise, and vary considerably across individuals, making classification difficult. Most prior studies within the SNeuroRehab project have relied on datasets with a limited number of subjects (typically 9–12), which limits ***generalizability***. Furthermore, previous studies predominantly used a single feature extraction technique, often overlooking key information embedded in EEG signals. Moreover, the processing of EEG data and training ML models can be computationally intensive, limiting analysis for those researchers who lack access to supercomputing resources. These limitations reduce classification accuracy or hinder the real-world application of EEG-based BCIs.

1.2 Goals

This study aims to assist with EEG-based classification of Motor Imagery (MI) and motor execution (ME) tasks using ML techniques. We aim to achieve this by:

1. Devising a strategy for effective channel (feature) selection to optimize computational efficiency while preserving classification accuracy for differentiating ME versus MI EEG signals.

2. Using a significantly larger dataset containing EEG recordings from 103 subjects[1], obtained with an EEG device with 64 channels resulting in approximately 24,324,480 observations. This can provide deeper insights into EEG signal variability compared to previous studies.

3. Applying multiple feature extraction techniques, including time-domain, frequency-domain, time-frequency-domain, and spatial-domain methods, as well as advanced approaches such as Common Spatial Pattern (CSP) and Power Spectral Entropy.

[1]Six subjects were excluded due to data inconsistencies)

4. Evaluating multiple machine learning models in combination with the feature extraction techniques to comparatively evaluate their accuracy performance and determine viability for future use. Models considered include Support Vector Machine (SVM), Random Forest(RF), K-Nearest Neighbors(KNN), Extreme Gradient Boosting(XGBoost) and Multilayer Perceptron(MLP).

By addressing these goals, this research aims to advance EEG-based BCI applications in rehabilitation and assistive technology by identifying an efficient and accessible way to study and differentiate brain signals. Furthermore, in the context of the SNeuroRehab project, it will provide valuable ground work to:

- Understand how to filter, select and preprocess EEG signals as dataset features.
- Understand which approaches may have a higher chance of better classifying between MI and ME.
- Suggest information on how to collect and analyze EEG features, including an understanding of which EEG nodes may provide more valuable information.
- Provide a comparative baseline of classification strategies for MI vs. ME brain signals using feature extraction methods.

1.3 Research Gap and Contributions

Previous research on EEG-based motor imagery (MI) and motor execution (ME) classification has achieved notable accuracy, often exceeding 80%. However, these studies typically rely on datasets with a limited number of subjects (9–12), which restricts the generalizability of their findings. For example, the study by Song J. and Zhai Q.[81] used the BCI Competition IV-2a and IV-2b datasets, each containing data from nine subjects. Similarly, G. Dornhege and co-authors [19] collected EEG recordings from 10 subjects, achieving comparable classification performance. Another study, conducted by multidisciplinary students

from the USA, Bangladesh, and Japan [73], analyzed data from the BCI Competition III (IVA) and IV, which included recordings from five and nine subjects, respectively. The Grasp-and-Lift EEG dataset, used by Yahya N. and co-authors [88], contained data from only 12 subjects.

In addition to dataset limitations, many studies[58][53][42] [37] have primarily relied on a single feature extraction technique, which may overlook critical information embedded in EEG signals. While methods such as Common Spatial Pattern (CSP), time-frequency decomposition, and power spectral analysis have been explored, their effectiveness varies across different datasets. Moreover, although multiple classifiers—including Support Vector Machine (SVM), Multilayer Perceptron (MLP), k-Nearest Neighbors (KNN), Quadratic Discriminant Analysis (QDA), and Deep Learning models—have been tested, no consensus exists on the most effective approach for robust classification.

Notably, much of the existing work has focused on MI versus Rest or multiclass MI classification, with relatively high accuracy reported in some cases. However, significantly fewer studies have addressed the classification between ME and MI—despite neuroimaging evidence (e.g., fMRI)[50] [15]showing both tasks activate overlapping motor-related brain areas. Existing ME vs. MI classification efforts using CNNs and SVMs [87][35]have reported low accuracy, often below 60%, highlighting the difficulty and underexplored nature of this problem.

This study aims to address these gaps by leveraging a significantly larger dataset, incorporating data from 109 subjects. By applying multiple feature extraction techniques and implementing an effective channel selection strategy, this research enhances EEG-based classification of MI and ME tasks. The findings contribute to advancing BCI applications in neurological rehabilitation by improving classification accuracy and computational efficiency.

To address these gaps, this study leverages a large-scale EEG dataset involving 109 subjects—significantly more than in prior studies—to improve the generalizability of results. It integrates diverse feature extraction methods, employs an effective EEG channel selection strategy, and evaluates multiple classification algorithms for MI vs. ME tasks. These con-

tributions enhance classification performance while reducing computational complexity and offer insights to guide future EEG-based experiment design within the Smart NeuroRehab initiative.

The proposed work is novel as there is no related work that tackles the problem of MI versus ME discrimination parting from 64 channels, and using feature extraction methods, nor offers a comparison of different combinations of these with various classifiers.

1.3.1 Hypotheses:

Based on our problem statement as well as the research gap discussed above, we outline these hypotheses:

Null Hypothesis (H_0): The use of Independent Component Analysis (ICA) in combination with a feature extraction methods (time-domain, frequency-domain, time-frequency-domain, or spatial-domain) cannot yield a strategy to efficiently differentiate ME from MI EEG signals.

Alternative Hypothesis (H_1): The use of Independent Component Analysis (ICA) in combination with at least one of the extraction methods (time-domain, frequency-domain, time-frequency-domain, and spatial-domain) can yield a strategy to efficiently differentiate ME from MI EEG signals.

The term *efficiently* refers to the ability to classify these signals (by with an accuracy percentage close the highest one reported in related works that use more than one channel ($79\pm1\%$). However, it also refers to being able to execute the classification pipeline using ubiquitous, accessible hardware (e.g., a laptop[2]).

1.3.2 Scope

For this research, we focus solely on EEG signals attributable with *upper-limb* (hand) movements only. One of the reasons being that the neurological issues presently being targeted

[2]For this study, the hardware used had an M1 chip with an 8-core CPU (4 performance cores, 4 efficiency cores) with 8GB of RAM

by the SNeuroRehab initiative are associated with rehabilitation therapies involving hand movements (e.g., certain type of stroke affections and muscle weakness), and their data collected in-house has focused on hand movements. In terms of the data analyzed, it consist of EEG recordings from 103 *healthy* subjects, acquired using a 64-channel system based on the international 10-10 electrode placement scheme. The work emphasizes *offline* processing and classification using traditional ML approaches, aiming to identify effective **features** and models for differentiating between MI and ME tasks.

1.4 Document Outline

The remainder of this paper is organized as follows: Section 2 reviews related work, summarizing existing research on motor imagery (MI), its applications in rehabilitation, and Brain-Computer Interfaces (BCIs). Section 3 provides essential background on the brain, EEG, BCI systems, and explains the concepts of motor imagery (MI) and motor execution (ME). Section 4 details the methodology, including dataset selection, preprocessing, feature extraction, and machine learning models. Section 5 presents the experiments and results, including comparative analysis and a summary of key findings. Finally, Section 6 concludes the study by discussing contributions, outlining limitations, and proposing future research directions.

Chapter 2

RELATED WORK

Over the last decade, research on Motor Imagery (MI) has expanded significantly, particularly in its application to rehabilitation and Brain-Computer Interfaces (BCIs). Numerous studies have explored a variety of feature extraction and classification techniques to improve the performance of MI-based systems. However, many are limited by small sample sizes—often involving fewer than a dozen subjects—limiting the generalizability of their findings. While MI has been shown to be effective both independently and in combination with Motor Execution (ME), the highest classification accuracies typically arise in subject-specific models using small cohorts.

A key challenge in this area is identifying reliable feature extraction methods for distinguishing between MI and ME. Studies vary in their classification goals—some distinguish MI from rest, ME from rest, different classes within MI or ME, or directly compare MI and ME. Each scenario may require distinct feature extraction strategies. Preprocessing choices such as filtering and artifact removal also play a crucial role in classification performance. This review aims to critically evaluate the literature to assess whether existing evidence is sufficient to identify optimal preprocessing and feature extraction methods for robust MI and ME classification.

2.1 Dataset

Compared to two decade ago, the landscape of EEG research has significantly evolved. Previously, EEG devices were less accessible, and publicly available EEG datasets were scarce. In the last ten years, however, there have been substantial advancements in both hardware and data availability. Portable, user-friendly EEG devices are now widely accessible, allowing

researchers and students to collect data more easily.

One major contribution to the availability of EEG datasets comes from the Brain-Computer Interface (BCI) Competitions, such as BCI Competition III and IV [71, 9]. These competitions were established to provide standardized datasets and benchmarking platforms for evaluating signal processing and classification techniques in brain-computer interface (BCI) research. The datasets from these competitions are publicly available and have been extensively used in the BCI community.

BCI Competition IV, for instance, includes several datasets recorded using various EEG configurations—such as 22-channel, 64-channel, and 3 bipolar-channel EEG setups. These datasets involve recordings from 9, 7, and 9 subjects respectively. The primary focus is on motor imagery (MI) tasks, making them particularly relevant for studies involving MI classification.

One of the most frequently used datasets from BCI Competition IV is Dataset 2a, which includes recordings of four different motor imagery tasks: left hand, right hand, foot, and tongue movements. This dataset has been widely applied in research, including studies by Yang et al. [89] and Barachant et al. [8], to test and compare feature extraction and classification methods.

Despite their popularity, the BCI-III and BCI-IV datasets have notable limitations. They include data from fewer than ten subjects and focus exclusively on imagined motor tasks, with no data related to actual motor execution. This restricts their generalizability and limits their applicability for studies aiming to compare motor imagery with motor execution or to develop subject-independent models.

While the number of MI-related studies has grown substantially, most of the existing literature—and the datasets used—are limited in two critical ways: they either focus solely on motor imagery without including motor execution data, or they involve a small number of subjects (typically fewer than ten). These limitations hinder the development of robust, generalizable models that can perform well across individuals and task types. There is a clear need for datasets that include both MI and ME tasks collected from a larger and more

diverse group of subjects.

2.2 MI for Rehabilitation

Recent research in both computer science and neurology has demonstrated the potential of using motor imagery (MI) for rehabilitation. Numerous studies from both disciplines support this possibility. Neurological rehabilitation aims to help patients who have experienced nervous system-related incidents—such as strokes, surgeries, trauma, major accidents, or amputations—relearn essential skills required for daily life [38].

Mulder's review study [54] investigated whether MI activates the same brain areas as motor execution (ME). Drawing on findings from both the computer science and neurology domains, this interdisciplinary review provides strong evidence supporting the overlap between MI and ME in brain activation. The study discusses a variety of experiments from different fields:

1. **Playin Piano in Mind[50]:** This study explored how both music performance and music imagery engage overlapping brain regions. Twelve music academy students participated by mentally and physically performing a right-hand piano piece. fMRI results showed that both conditions activated a bilateral frontoparietal network, including the premotor areas, the precuneus, and Brodmann Area 40. However, the posterior parietal cortex was only activated during actual performance, likely due to the higher visuomotor demands of playing an instrument.

2. **Mentally simulated leg exercise[15] :** This study demonstrated that imagining a leg exercise can increase heart and breathing rates proportionally to the imagined workload—similar to real exercise. However, unlike actual physical activity, there were no changes in muscle metabolism. Participants still reported greater fatigue when imagining heavier loads. These findings suggest that MI can trigger autonomic responses, indicating mental preparation for action without physical execution.

3. **Spinal Reflex Activation During Imagined Finger Movements**[47]**:** This experiment showed that imagining finger movements increases spinal reflex excitability. When muscle stretches were applied during imagined finger flexion, short-latency EMG responses—similar to those observed during real movement—were recorded. No such responses were observed during rest. The study suggests that MI can activate spinal motor circuits through subthreshold activation of motoneurons, highlighting the strong physiological link between mental simulation and motor readiness.

These studies collectively demonstrate that motor imagery can activate neural and physiological systems in ways that resemble actual movement, highlighting its potential for rehabilitation. However, from a machine learning perspective, a critical gap remains. While neuroscience has shown the viability of MI, the development of reliable MI-based rehabilitation systems using BCI requires models that can robustly distinguish between motor imagery (MI) and motor execution (ME) signals. This demands not only accurate feature extraction but also effective classification frameworks capable of generalizing across individuals. Bridging this gap is essential to translate these findings into real-world, ML-driven BCI applications for rehabilitation.

2.3 Feature Extraction and Classification

Feature extraction plays a critical role in enabling the use of motor imagery (MI) for rehabilitation through brain-computer interface (BCI) systems. One of the main challenges in this domain is the difficulty of achieving sufficient variance between motor imagery (MI) and motor execution (ME) signals. Since both tasks activate overlapping regions of the brain during similar actions, distinguishing between them using EEG data can be particularly challenging.

The feature extraction process involves transforming raw EEG signals into meaningful representations that capture discriminative patterns between MI and ME. A variety of studies have explored different techniques for this purpose, applying signal processing and machine

learning approaches to enhance classification performance. This section reviews relevant studies, highlighting the feature extraction methods used and the classification accuracies reported.

A study by Geethanjali and Ray [30] utilized EEG data collected from seven subjects at Colorado State University to investigate time-domain feature extraction. The features extracted included Mean Absolute Value (MAV), Zero Crossing (ZC), Slope Sign Changes (SSC), and Waveform Length (WL). The dataset comprised five distinct mental tasks:

1. **Baseline (B):** Relax the mind without engaging in any thought or task.
2. **Multiplication (M):** Mentally solve a multiplication problem (e.g., 12×13).
3. **Letter Composition (L):** Imagine writing a letter to a relative.
4. **Rotation (R):** Visualize a previously shown object rotating in space.
5. **Counting (C):** Mentally visualize and count numbers being written and erased on a blackboard.

Binary classification was performed using all pairwise combinations of the five tasks, including B–M, B–L, B–R, B–C, M–L, M–R, M–C, L–R, L–C, and R–C. Among all subjects, Subject 4 achieved the highest average classification accuracy of 98.9% across all task pairs.

Although the mental tasks used in this study are not directly related to rehabilitation activities, the results demonstrate that time-domain features, when combined with a Linear Discriminant Analysis (LDA) classifier, can effectively enhance the separability between rest and MI conditions, as well as among different MI tasks.

In a separate study, Ramos et al. [63] used the publicly available epilepsy EEG dataset from Bonn University, which includes recordings from healthy individuals and epileptic patients. The study focused on binary classification of healthy versus epileptic seizure EEG signals. Three different feature extraction techniques were applied to extract time–frequency

domain features from EEG spectrograms. The classification performance was evaluated using multilayer perceptron (MLP), support vector machines (SVM), and k-nearest neighbors (KNN). The authors reported achieving up to 100% accuracy in distinguishing between the two classes.

Another study [87] utilized the BCI Competition IV-2a and 2b datasets, which include EEG data from nine subjects. The study focused on binary classification of left-hand versus right-hand motor imagery (MI) tasks. Band power features were extracted by filtering EEG signals into five frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz). Both Support Vector Machine (SVM) and Convolutional Neural Network (CNN) classifiers were employed for classification. The highest accuracy reported was 94.1% for Subject 5. While subject-specific accuracies were relatively high, the overall average accuracy across all subjects using both classifiers reached a maximum of 61.3%.

A recent study [75] explored the use of Riemannian geometry decoding algorithms for classifying motor imagery and motor execution EEG signals. The authors applied several Riemannian decoding strategies on data recorded with both 64- and 29-electrode configurations. The best average classification accuracy achieved was 78.95%. Since this study used a comparatively large dataset, the results highlight the need to extract additional types of features and combine them with different classifiers to further improve performance.

In summary, while substantial progress has been made in classifying motor imagery (MI) using diverse feature extraction and machine learning techniques, the development of generalizable models remains constrained by the limited availability of large-scale datasets that include both MI and motor execution (ME) data. This review underscores the pressing need for integrated datasets, standardized and robust preprocessing protocols, and the exploration of hybrid or complementary feature extraction strategies. Addressing these gaps is essential for translating neuroscientific insights into effective, real-world brain-computer interface (BCI) systems for rehabilitation.

Chapter 3

BACKGROUND

The human brain is an organ considered by many as the "cornerstone of our existence", governing essential functions that define our everyday experiences—breathing, movement, cognition, and memory, among others. Its intricate complexity surpasses that of any known structure, reflecting a network of billions of cells, predominantly glia and neurons[93]. Neurons, with their electrochemical signaling, and glia cells, providing essential support, collectively sustain the brain's vitality and functionality. We can analyze either electrical signals or brain images to comprehend the cognitive behavior of the brain.

Brain-Computer Interfaces (BCIs) [68] are innovative tools for studying brain signals. EEG-based BCIs have gained momentum due to their convenience, portability, safety, and cost-effectiveness. Advancements in machine learning algorithms further enhance EEG's reliability in detecting various brain functions[79].

In practical scenarios, EEG still faces issues due to its limited reliability in interpreting and decoding brain signals. There are various reason behind this like noisy data,high non-stationarity of data,insufficient calibration data, subject specific nature, and poor spatial resolution[75]. These challenges require attention across multiple fronts.For example at signal processing level we can develop features and classifiers that are adaptable to varying contexts and require minimal data for calibration[90]. Whereas in neuroscience, we can pinpoint changes in brain wave frequency or amplitude and the activation of specific brain areas. These could serve as more reliable neurophysiological markers.

Research demonstrates residual brain activity during motor imagery in paralyzed individuals, indicating its promise for rehabilitation and prosthesis control. Consequently, endeavors to develop brain-computer interfaces for paralyzed patients frequently incorporate motor

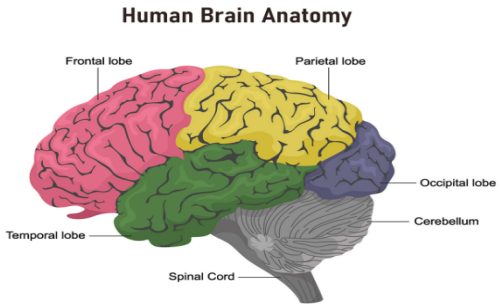
imagery for controlling external devices [51]. Accurate classification of motor imagery and motor execution is crucial for BCI applications, such as prosthetics, enabling patients to perform daily tasks with a prosthetic arm.

3.1 The Human Brain

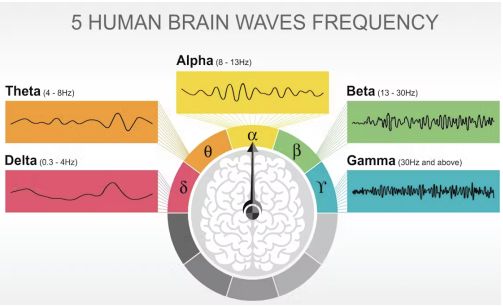
The brain comprises two hemispheres, the Left and Right hemispheres, each of which is further subdivided into four lobes: frontal, temporal, parietal, and occipital. The diagram Figure 3.1a illustrates the positioning of these lobes within the human brain. The frontal lobe, situated behind the forehead, is the largest lobe and governs functions related to emotions, language, and memory. Adjacent to the frontal lobe is the parietal lobe, responsible for integrating sensory information from various parts of the body. Positioned on the side of the brain, the temporal lobe is associated with hearing, language, and sensory processing. Finally, located at the lower back of the head, the occipital lobe handles the perception and processing of visual information.

3.2 Brain Computer Interface

Brain copmputer interfaces also known as brain machine interfaces(BMI's) are used to control devices using brain signals.BCIs find applications in various sectors such as medical for disease prevention, detection, and diagnosis, as well as rehabilitation. They are also used in neuroergonomics and smart environments to achieve smart homes and workplaces. Additionally, BCIs have applications in neuromarketing and advertising, education and self-regulation, games and entertainment, as well as security and authentication [5]. There are three types of BCIs used to analyze brain signals: invasive (microelectrode array), partially invasive (electrocorticography), and non-invasive (electroencephalograph). Invasive BCIs require neurosurgery for implantation into the grey matter of the brain, partially invasive BCIs are implanted inside the skull but outside the brain, and non-invasive BCIs are attached to the scalp.[1]



(a) Structure of Human Brain and its Lobes (Image Source:https://www.hopkinsmedicine.org/)



(b) Brain Wave Frequencies (Image Source: https://theconversation.com/)

Figure 3.1: Human Brain and its Wave Frequencies

3.3 Electroencephalography

In 1875, scientist Richard Caton achieved the first successful recording of brain activity through electrical signals. Using a galvanometer, he positioned two electrodes on the scalp of a human subject. This milestone marked the introduction of the term Electroencephalogram (EEG), with "electro-" referring to the recording of brain electrical activity, "encephalo-" denoting the signals emitted from the head, and "-gram" indicating graphy.[69]

When neurons are stimulated, synaptic currents occur within their dendrites, leading to the creation of an electrical field close to the scalp. Pyramidal neurons, mainly found in layers three and five of the cerebral cortex, generate the electrical signals detected by EEG electrodes[56]. The recorded scalp electrical signals arise from the accumulation of inhibitory or excitatory postsynaptic potentials(IPSPs or EPSPs) from numerous pyramidal cells surrounding each electrode.

The electrical activity of the brain, as captured by EEG, can be categorized based on frequency. The most common classification includes alpha, beta, theta, delta, and gamma waves, as depicted in Figure 3.1b. Each frequency band has distinct associations[44]. Alpha

waves are associated with relaxation, beta waves with an active state, theta waves with meditation or inward focus, delta waves with deep sleep, and gamma waves with the integration of senses and memory, often linked with states or tasks requiring heightened concentration.

3.3.1 Motor Execution and Motor Imagery

The difference between the intent moving and actually doing it, can be formalized with these two concepts, which are focal to the proposed research:

- **Motor Execution:** Refers to the actual physical act of performing a movement of body.

- **Motor Imagery:** Refers to mentally visualizing or imagining movement without performing [18].

Recent studies have shown that the mental visualization of an action activates the identical neural networks within the motor system as those utilized during the actual execution of the movement [54] [21].

A study [54] involving 16 stroke patients aimed to investigate the impact of combining motor imagery with occupational therapy compared to using occupational therapy alone. The findings revealed that eight patients who received the combination therapy exhibited significant improvement compared to those who underwent occupational therapy alone.

Chapter 4

METHODOLOGY

This chapter describes in detail the data, concepts and methods used in this research. Data selection, filtering and processing is one of the dominant aspects in here, as this work is being used as a building block for other aspects of the NeuroRehab Ecosystem project.

The core objective of this study is to investigate whether effective preprocessing combined with various domain-specific feature extraction methods—spanning time, frequency, time-frequency, and spatial domains—can significantly improve the classification performance for Motor Execution (ME) and Motor Imagery (MI) EEG signals. The methodological steps presented in this chapter are designed to test this hypothesis through rigorous preprocessing, targeted channel selection, diverse feature computation, and consistent classification techniques.

4.1 Dataset

We begin with the introduction and description of the dataset used, as it is the foundation of this research. Collecting EEG data is challenging for various reasons, device characteristics, subject availability, experiment design, environment noise conditions, etc. In the case of the NeuroRehab Ecosystem project team, we faced challenges with a device nearing end of life, which in many occasions resulted in noisy or unreliable data collection. In their thesis work, alumni Raj[62] and Gahatraj[23] conducted experiments using the aforementioned device, which was the Wireless DSI-7 from Wearable Sensing[86] with 8 electrodes. However, the amount of data collected in their work was insufficient and not adequate for the goals of this research, which require a large sample size and ideally, data from more electrodes that can provide more features. In addition, one of the goals of this research requires data that

includes **motor imagery** (intent to do an action or imagine an action).

We define an action in terms of those can be measured using the BCI devices used in the NeuroRehab project and it's larger goal. The main reason is to maintain consistency in and create potential benchmarks for current and future work. We also assume that an action entails going from a rest position to a specific position, for instance, Figure 4.1 depicts the action "make a fist".

Definition 1. Let $\mathbb{A}$ represent the set of hand gestures that are detectable by our EEG and EMG devices and has some therapeutic relevance (for upper extremity purposes).

Definition 2. In the context of this research, an ***action*** $\mathcal{A} \in \mathbb{A}$, is a single instance of a hand gesture.

Definition 3. Let $\mathcal{T}$ represent a ***task***, which is a time interval time during which an action $\mathbf{A}_i$ is repeated several times.

Figure 4.1: Hand Clench, i.e., from a relaxed position to making a fist.

[3]

Taking these requirements into consideration, research was done to find existing viable datasets from external sources. The dataset collected by the developers of the BCI2000 system[22], is an open-source, general-purpose software platform for brain-computer interface (BCI) research was found to meet the requirements for the scope of this thesis. This dataset, which contributed to PhysioNet, includes over 1,500 EEG recordings capturing both **motor imagery (MI)** and **motor execution (ME)** (actual actions). Next, we will describe the different aspects of the BCI2000 dataset.

The BCI2000 Experiment Outline

The BCI2000 system (http://www.bci2000.org) used 64 channels to collect EEG data from various subjects. Each subject completed 14 experiment sessions consisting of:

- Two 1-minute baseline sessions: one with eyes open and one with eyes closed
- Three 2-minute sessions for each of four predefined tasks (listed below)
- A relaxation interval between each 2-minute session

During these sessions, the subject sits in front of a monitor and is prompted to perform a task. The tasks performed during the 2-minute sessions are as follows:

Task 1 First Motor Execution: A target appears on the left or right side of a screen. The subject opens and closes the fist on the same side as the target until it disappears, then relaxes.

Task 2 First Motor Imagery: A target appears on the left or right side of the screen. The subject imagines opening and closing the fist on the same side as the target until it disappears and then relaxes.

Task 3 Motor Execution Hand / Foot: A target appears at the top or bottom of the screen. If the target is at the top, the subject opens and closes both fists. If the target is at the bottom, the subject opens and closes the toes on both feet. The action is continued until the target disappears, then the subject relaxes.

Task 4 Motor Imagery Hand / Foot: A target appears at the top or bottom of the screen. If the target is at the top, the subject imagines opening and closing both fists. If the target is at the bottom, the subject imagines opening and closing both feet. The imagining continues until the target disappears, then the subject relaxes.

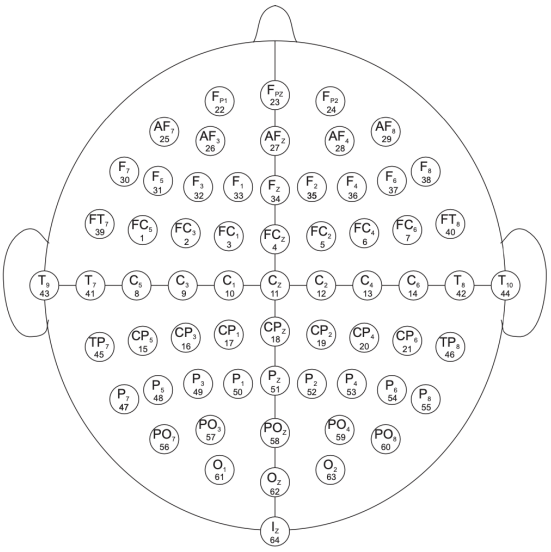


Figure 4.2: EEG Montage Utilized During Data Collection [31].

EEG Node Montage

Electrode placement for EEG data collection followed the international 10-10 standard montage, which provides a higher resolution of spatial coverage compared to the 10-20 system[57]. This montage is particularly beneficial for capturing motor-related brain activity due to its dense arrangement over sensorimotor areas. Figure 4.2 illustrates the electrode configuration used during the BCI2000 experiment sessions.

Action Labels in the BCI2000

One of the ultimate goals of this work is to determine whether it is possible to differentiate between EEG signals generated by different actions – and eventually make automated decisions based on these.

The BCO2000 dataset contains **labels** that identify a task $\mathcal{T}_i$ within each experiment

timeline and subject. For this research, we focused on the upper limb by selecting tasks specifically related to upper limb movements, i.e., opening and closing a fist.

This decision was made to match – to some extent – previous and current work being done by the NeuroRehab Ecosystem team, so it can serve as a comparative study and perhaps even to provide a rough baseline for future work. Table 4.1 gives a description of the tasks we will considered and their corresponding labels, the colors indicate how we group these tasks to create the three different classes. Specifically, we focus on tasks associated with the action of "making a fist", $\mathcal{A}_{Fist}$. Hence, we grouped existing data samples from "similar" tasks into categories we can then use to evaluate analysis and discrimination strategies. These classes are:

1) Relaxed state: Labels 1 and 4.
2) Motor Execution (ME): Labels 2 and 3.
3) Motor Imagery (MI): Labels 5 and 6.

Table 4.1: Task Descriptions by Label for Data Used in This Study: each color is a class.

Label	Task $\mathcal{T}$
1	Relax
2	Open and Close Left Fist (OCLF)
3	Open and Close Right Fist (OCRF)
4	Relax
5	Imagine Open and Close Left Fist (IOCLF)
6	Imagine Open and Close Right Fist (IOCRF)

In this study, we consider only "Fist" actions for both ME and MI and without distinguishing between left and right hand. To simplify our notation, we will use $\mathcal{T}_E$ to represent

executing the action of making a fist, and $\mathcal{T}_I$ to represent imagining making a fist. Similarly, we will use $\mathcal{T}_R$ to represent a relaxation state.

Establishing an Experimental Protocol

The BCI2000 dataset includes data from 103 subjects organized across several files. Each file has a run length of 2 minutes that includes 3 types of tasks performed during each *session*: relaxation followed by one of the two motor tasks (execution or imagination). Table 4.2 summarizes relevant dataset characteristics.

Table 4.2: Dataset Parameters

Parameter	**Value**
Subjects	103
Run files per subject	6
Run length	2 min
Tasks	3 (ME, MI or Relaxation)
Runs per tasks	Each task is repeated 3 times across separate runs
Trials per run	Each run(1 file for one subject) consists of 30 task occurrences
Electrodes	64
Sampling rate(Hz.)	160

Table 4.3 provides an example of an annotation file for a single subject. Such a file consists of 30 records and 5 columns. The first column represents the label of the action, the second column indicates the duration of the action in seconds, the third column specifies the number of rows corresponding to that action, and the fourth and fifth columns show the starting and ending rows for the respective label (action).

Definition 4. Let $\mathbf{X}_L$ represent a ***trial***, which is a time interval of approximately 4 seconds during which a task $\mathcal{T}_L$ is performed ($L \in \{1, 2, ..., 6\}$ is a label).

Definition 5. Let $\mathcal{R}$ represent a ***run*** or *Session*, which is a 2-minute interval during which there is a sequence of 30 trials.

A file contains a *Session*, where a sequence of trials $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ is carried out. For instance, in a scenario where the tasks considered are $\mathcal{T}_1$: Relaxation, $\mathcal{T}_2$: Open and close left fist, and $\mathcal{T}_3$: Open and close right fist. The trial sequence within the file alternates in a pattern such as: $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_1$, $\mathbf{X}_3$, $\mathbf{X}_2$, $\mathbf{X}_1$, $\mathbf{X}_3$, ..., continuing until there are a total of 30 trials in the file.

Table 4.3: Annotation data example table

Serial No.	**Trial Label**	**Trial Length (sec)**	**Trial Length (samples)**	**Trial Onset**	**Trial End**
1	1	4.2	672	1	672
2	3	4.1	656	673	1328
3	1	4.2	672	1329	2000
... Table contains 30 rows in total ...					

Our experiments focus on distinguishing between $\mathcal{A}_{Fist}$ (motor execution) 4.1 with either hand as one class, $\mathcal{A}_{Fist}$ **intent** (motor imagery), and having the hand(s) in a state of relaxation ($\mathcal{A}_{Relax}$), either before or after a task has been performed.

Since we consider only "Fist" actions for both ME and MI and without distinguishing between left and right hand, to simplify our notation, we can use $\mathcal{T}_E$ to represent executing the action of making a fist, and $\mathcal{T}_I$ to represent imagining making a fist. Similarly, we will use $\mathcal{T}_R$ to represent a relaxation state.

Data Availability

The dataset used in this study is publicly available and can be accessed through the links provided below. A curated version of the dataset, available in CSV and MATLAB formats, was used for this research. The original dataset, available in EDF format, is also included for reference.

1. EEG Motor Movement/Imagery Dataset (Original Dataset)[31]
2. Increasing Accessibility to a Large Brain–Computer Interface Dataset: Curation of PhysioNet EEG Motor Movement/Imagery Dataset for Decoding and Classification (Curated Dataset)[77]

4.2 Data Pre-processing

4.2.1 Filters Used[10]

The data in the BCI2000 files consists of raw EEG signals, which contain noise as well as other signals not relevant for our study, we use the following filters:

1. **Notch Filter:** A notch filter was applied at 60 Hz to eliminate power line interference, which is commonly observed as a sharp peak in EEG recordings. This frequency corresponds to the standard electrical power frequency in the United States, while in Europe, the standard is 50 Hz. In this study, a noticeable peak at 60 Hz was observed in the power spectral density (PSD) plot during exploratory data analysis, indicating the presence of electrical noise. Applying a notch filter effectively removed this narrowband interference without affecting the underlying brain wave activity.

2. **High-Pass Filter (Low Frequency Filter):**To eliminate slow, low-frequency artifacts such as baseline drifts and sweat-related noise, a high-pass filter with a cutoff around 0.5 Hz was used. These low-frequency components do not contribute meaningfully to the classification of Motor Execution (ME) and Motor Imagery (MI) tasks. By applying this filter, the signal was cleaned of irrelevant low-frequency noise while preserving the frequency bands important for brain activity analysis.

4.2.2 Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is a technique used to decompose mixed EEG signals into statistically independent components, effectively isolating the underlying brain sources

responsible for the observed scalp activity. ICA reconstructs the original source signals by multiplying the EEG data matrix with an weight matrix:

$$\mathbf{U} = \mathbf{W} \cdot \mathbf{X}$$

Where:

X: EEG data (Channels × Time)

U: ICA Activity (Components × Time)

W: Weight matrix (Components × channels)

The unmixing matrix **W** can be computed using various algorithms. EEGLAB, for instance, employs the Infomax algorithm[17] to extract statistically independent components from EEG data.

The matrix **U** can be used to visualize ICA components and identify noisy or artifact-related components. These unwanted components can be suppressed by setting their values to zero. The cleaned EEG data can then be reconstructed using the inverse of the unmixing matrix:

$$\mathbf{X} = \mathbf{W}^{-1} \cdot \mathbf{U}$$

In this study, EEGLAB was used to perform ICA, visualize the components, remove noise sources, and reconstruct the cleaned EEG signals. Initially, ICA was performed separately for each class (Motor Execution and Motor Imagery). Noisy components were removed from each class, and the cleaned data were then combined. However, this approach is not ideal for EEG classification tasks, as ICA components derived from classwise decomposition may not generalize across the entire dataset.

To address this limitation, ICA was instead applied to the entire dataset. This approach yielded components that differed significantly from those obtained classwise. Table 4.4 presents a comparison of ICA component counts by type for the full dataset versus the MI and ME subsets.

Table 4.4: ICA component count for Entire and Classwise subsets of Data

Type	Entire Data	MI Data	ME Data
Brain	17	19	18
Muscle	16	14	15
Eye	4	3	3
Other	4	8	7
Channel Noise	5	1	3

Although applying ICA separately to each class may seem intuitive, it goes against the fundamental purpose of ICA—to isolate independent sources from the entire dataset. Class-wise decomposition can lead to inconsistent components that do not reliably represent the true underlying sources. However, this approach still provided a valuable insight: both Motor Execution (ME) and Motor Imagery (MI) conditions showed the presence of muscle-related components. This indicates that even without actual physical movement in MI trials, muscle artifacts are still present and detectable in the EEG data.

Figures 4.3 and 4.4 show the ICA components extracted from the entire dataset. EEGLAB also supports the visualization of each component's properties, as shown in Figures 4.5, 4.6, 4.7 and 4.8,which aids in verifying the classification of components. For example, the first ICA component was automatically labeled as an eye artifact by EEGLAB; this was confirmed by examining the corresponding property plots. The labeling guidance provided by EEGLAB's documentation (EEG Independent Component Labeling) was used to distinguish between brain, muscle, eye, channel noise, and other types of components.

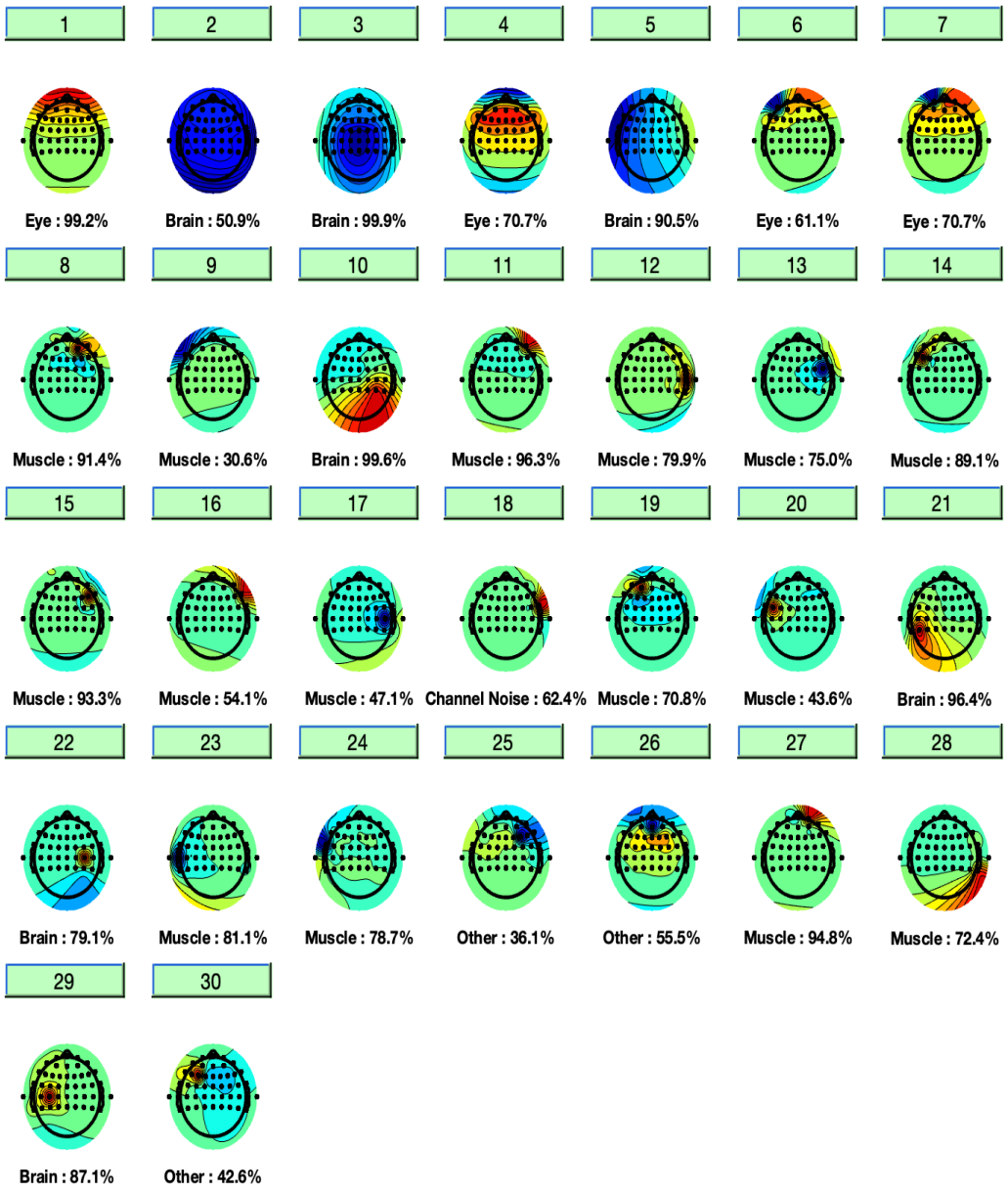


Figure 4.3: ICA components 1–30 for the entire dataset.

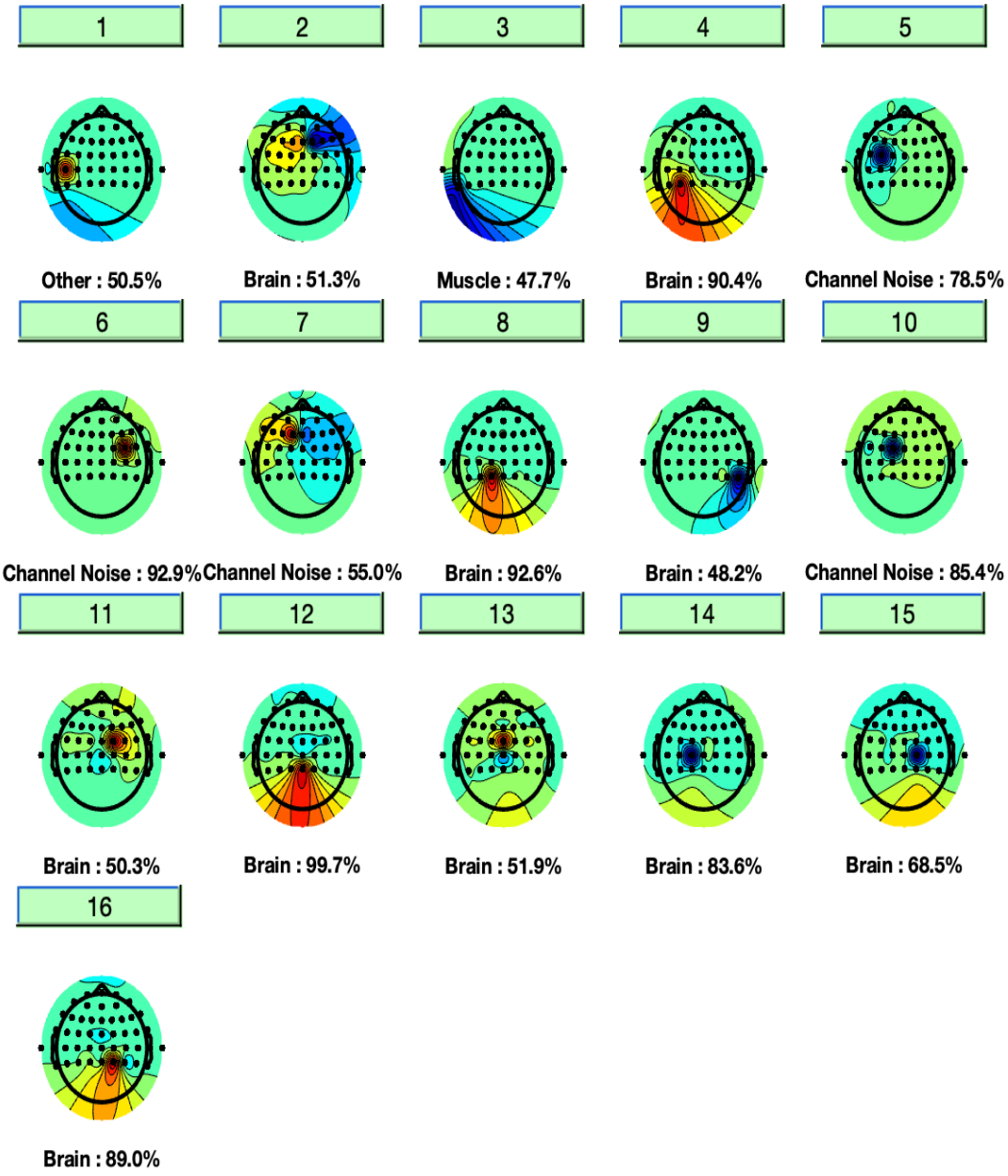


Figure 4.4: ICA components 31–46 for the entire dataset.

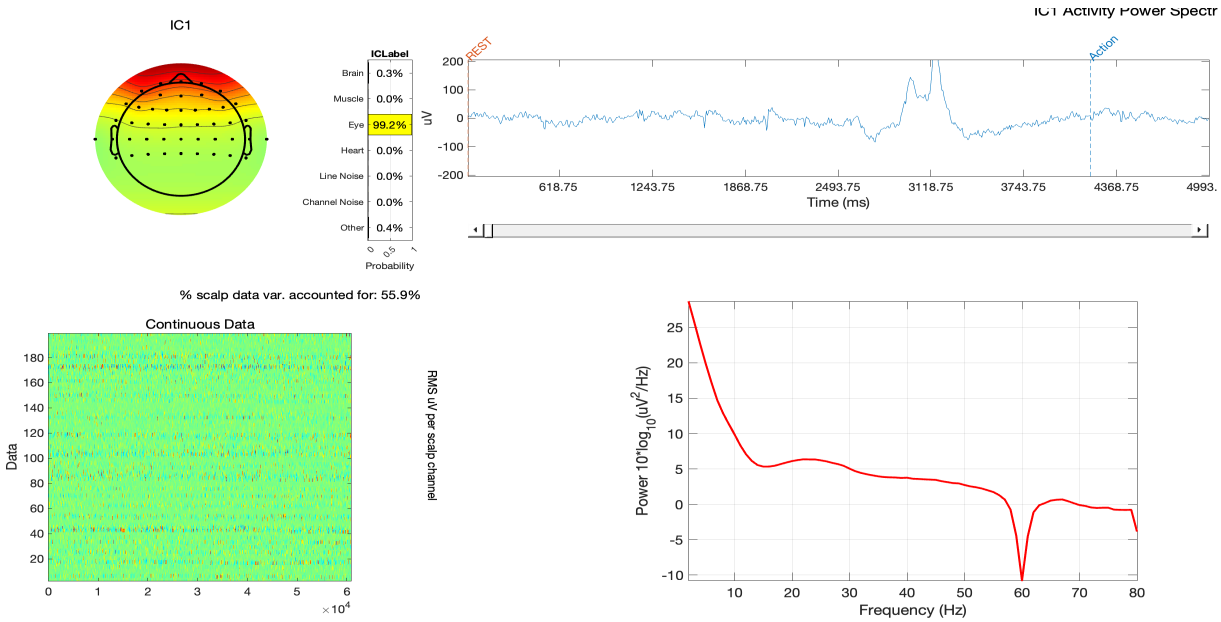


Figure 4.5: ICA 1

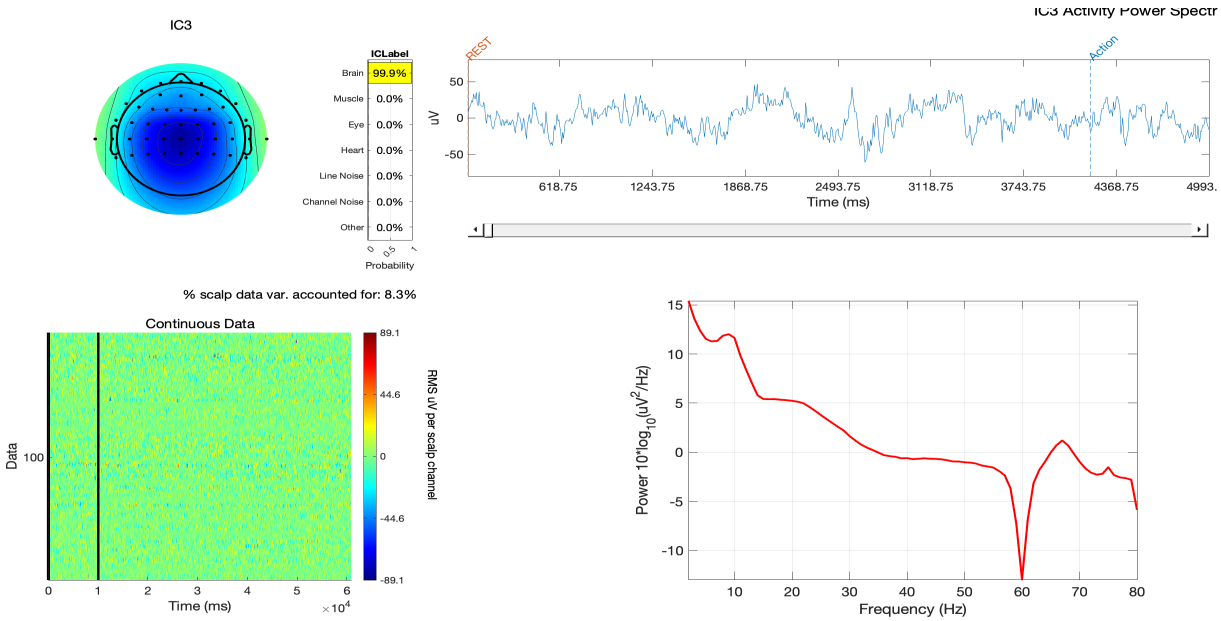


Figure 4.6: ICA 3

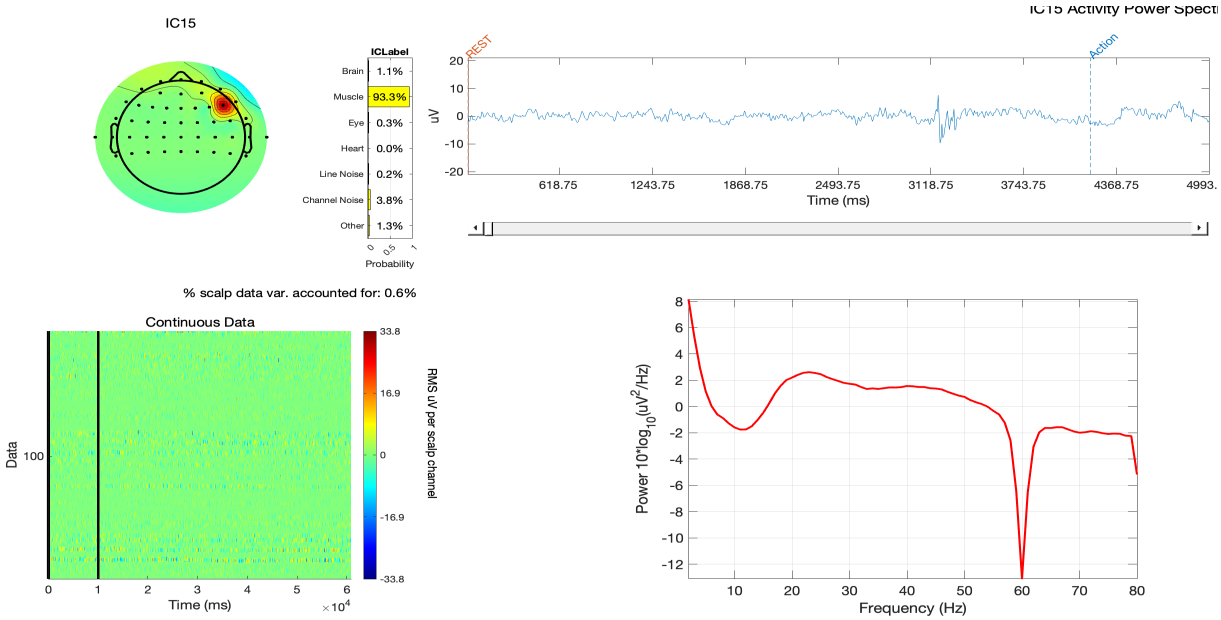


Figure 4.7: ICA 15

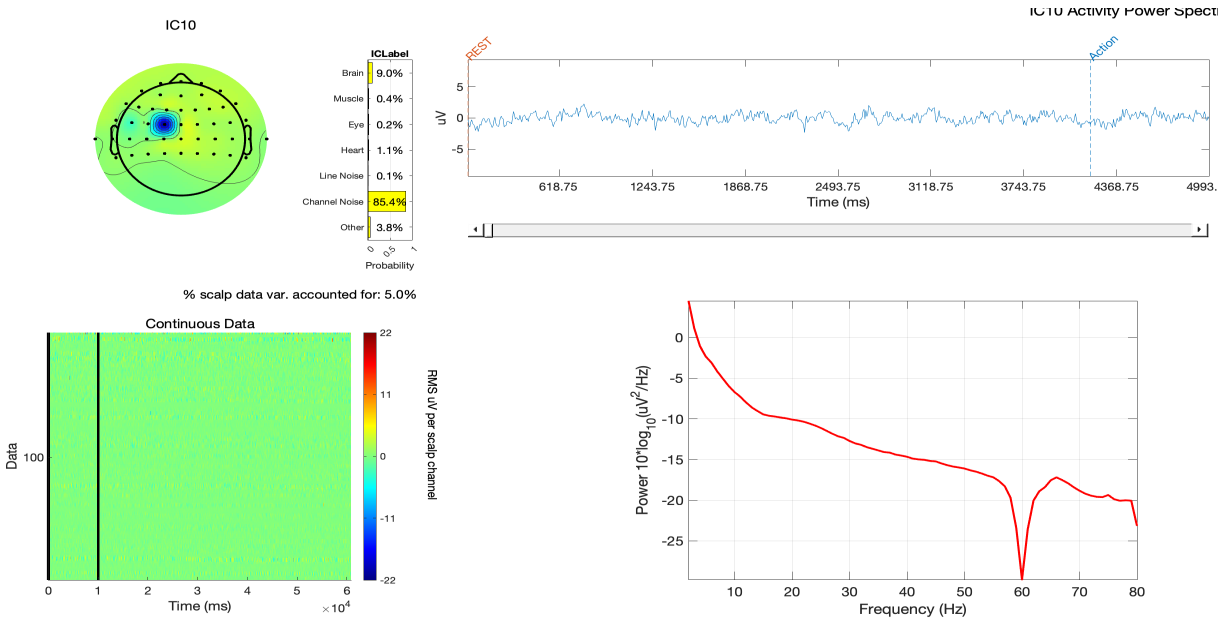


Figure 4.8: ICA 40

Based on this process, components 1, 4, 36, and 40 were identified as noise-related and removed. Figure 4.9 illustrates the EEG signals before (blue) and after (red) ICA removal. Notably, the red trace shows a clear reduction in prominent peaks at around 3, 7, and 11 seconds, indicating successful noise removal while preserving the underlying brain signal. After ICA-based artifact removal, the resulting cleaned EEG data were used for the feature extraction methods implemented in the subsequent stages of this study.

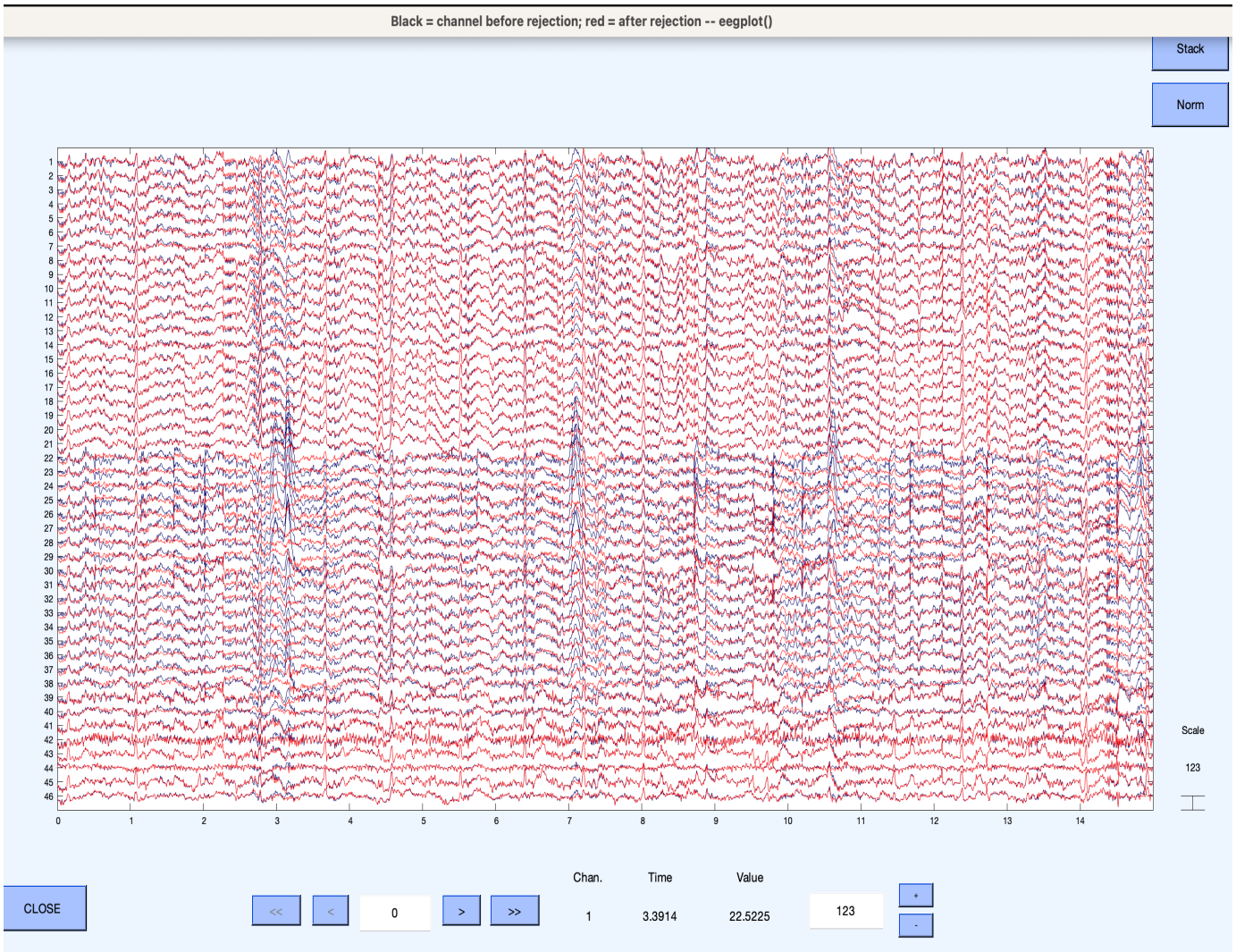


Figure 4.9: Signal Plot After Dropping Noisy ICA

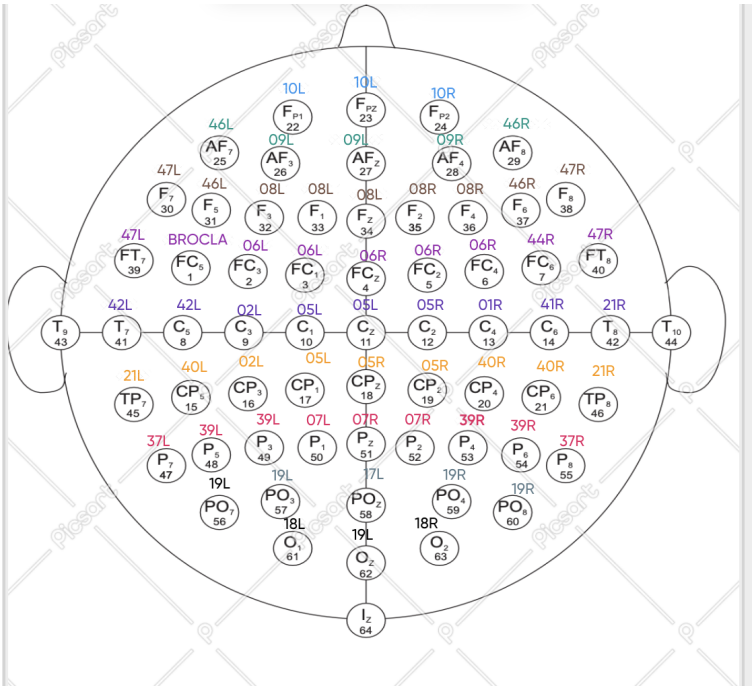


Figure 4.10: Closest 10-10 Electrode Position to each Broadmann area [31][11][52].

4.3 Effective Channel Selection

Effective EEG channel selection is crucial for accurately classifying Motor Execution (ME) and Motor Imagery (MI) tasks. It reduces setup time, computational load, and data redundancy by focusing on the most informative channels, thereby enhancing the practicality and performance of BCI systems in real-world applications[73].

While numerous EEG-based studies on the classification of Motor Execution (ME) and Motor Imagery (MI) emphasize feature extraction and classification techniques, few address the spatial localization of brain activity underlying these tasks. In contrast, neuroscience research[84][61][21][66] often maps ME- and MI-related brain activity using neuroimaging methods such as fMRI, typically reporting findings in terms of Brodmann areas—functionally distinct regions of the cortex.

For EEG-based studies, a practical approach involves aligning the 10–10 electrode mon-

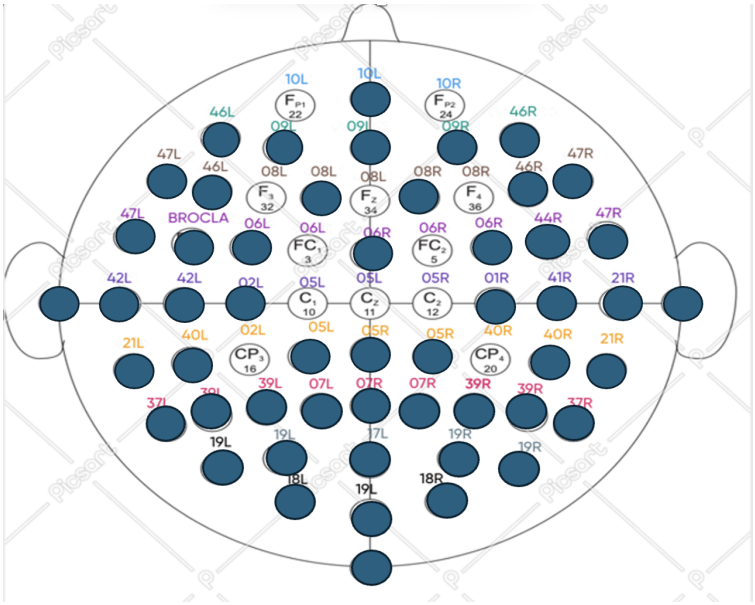


Figure 4.11: EEG 10–10 Montage Highlighting Selected Electrodes for MI and ME Classification

tage with Brodmann areas to guide electrode selection. A review of relevant literature was conducted to derive an association between EEG 10–10 electrode positions and Brodmann areas[11][52].The electrode-to-Brodmann area mapping is presented in Figure 4.10. Additionally, an EEG-based study utilizing Shannon entropy was reviewed to identify electrodes most relevant to MI classification. By combining insights from both anatomical mapping and signal-based feature selection, an initial set of 46 candidate electrodes was identified. Based on further analysis of classification studies, this list was refined to a more concise set of 12 electrodes,, as shown in Figure 4.11. These electrodes were used for all subsequent processing and feature extraction in this study.

4.4 Feature Extraction

The spatio-temporal nature of EEG data presents significant challenges for handling and analysis. Data contain complex patterns in both space (electrode locations) and time (signal dynamics).As a result, feature extraction becomes a crucial step to simplify the data and identify meaningful patterns that can enhance the performance of subsequent analysis or classification tasks.

4.4.1 Shannon entropy:

Shannon entropy is a statistical measure used to quantify the degree of uncertainty or randomness within a dataset[70]. Introduced by Claude Shannon in 1948, it evaluates the unpredictability associated with a set of possible outcomes. There is a direct relationship between entropy and the amount of information present in a signal: the more information or variability in the signal, the higher the entropy; conversely, less information leads to lower entropy due to reduced variability. For example[2], if all data points belong to a single class, predicting the class of a new data point is straightforward, and the entropy in this case is zero. However, if the data is distributed across multiple classes, predicting the class becomes more uncertain, resulting in higher entropy. In the context of EEG signals, a high entropy value typically indicates greater randomness or noise in the signal, making it harder to extract meaningful patterns. On the other hand, lower entropy suggests the presence of more consistent and structured information, which is beneficial for classification tasks. Shannon entropy H is calculated using the following formula:

$$H = -\sum_{i} p_i \log_2(p_i) \tag{4.1}$$

Where

H = Shannon Entropy,

Pi = represents the probability of occurrence of the i^{th} class or event

4.4.2 Common Spatial Patterns (CSP):

EEG signals inherently suffer from low spatial resolution due to volume conduction, which causes relevant information to be distributed across multiple channels. To mitigate this and improve signal quality, spatial filtering techniques are used to extract more localized and discriminative neural activity. CSP is one such technique that computes optimal spatial filters by projecting raw EEG signals into a space where the variance of one class is maximized while the variance of the other class is minimized. This projection enhances class discriminability, making CSP particularly useful for classification and other analytical tasks [53]. The following steps outline the CSP implementation used in this work:

1. **Covariance Matrix Calculation[49]:**
   For each trial i, the EEG data $\mathbf{X}_i \in \mathbb{R}^{C\times T}$ (with C channels and T time samples) is used to compute the spatial covariance matrix:
   $$\mathbf{C}_i = \mathbf{X}_i\mathbf{X}_i^\top$$

2. **Normalization:**
   Each covariance matrix is normalized by its trace to obtain the normalized covariance:
   $$\tilde{\mathbf{C}}_i = \frac{\mathbf{C}_i}{\mathrm{tr}(\mathbf{C}_i)}$$
   Trials where $\mathrm{tr}(\mathbf{C}_i) < 10^{-10}$ are discarded, as they indicate flat or corrupted signals.

3. **Mean Covariance Matrix:**
   The normalized covariance matrices are averaged across all valid trials for each class A and B:
   $$\mathbf{R}_A = \frac{1}{N_A}\sum_{i=1}^{N_A}\tilde{\mathbf{C}}_i^A, \quad \mathbf{R}_B = \frac{1}{N_B}\sum_{i=1}^{N_B}\tilde{\mathbf{C}}_i^B$$

4. **Generalized Eigenvalue Decomposition[72]:**
   The spatial filters are obtained by solving the generalized eigenvalue problem using the

eigh function from SciPy.

$$\mathbf{V}^H\mathbf{R}_A\mathbf{V} = \mathbf{W}, \quad \mathbf{V}^H(\mathbf{R}_A + \mathbf{R}_B)\mathbf{V} = \mathbf{I}$$

- $\mathbf{V}$: matrix of eigenvectors
- $\mathbf{W}$: diagonal matrix of eigenvalues
- $\mathbf{R}_A$: normalized covariance matrix for class A
- $\mathbf{R}_B$: normalized covariance matrix for class B
- $\mathbf{I}$: identity matrix

5. **Filter Selection:**

   The eigenvalues are sorted in descending order. The spatial filters are formed by selecting the eigenvectors corresponding to the m largest and m smallest eigenvalues:

$$\mathbf{W} = [\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{C-m+1}, \ldots, \mathbf{v}_C]$$

6. **Feature Extraction:**

   For each trial, the original EEG signals are projected using the CSP filters:

$$\mathbf{Z} = \mathbf{W}^\top\mathbf{X}$$

   where $\mathbf{X} \in \mathbb{R}^{C\times T}$ is a trial, and $\mathbf{Z} \in \mathbb{R}^{2m\times T}$ contains the spatially filtered signals. Then, the variance of each projected component is computed to form the feature vector:

$$f_j = \text{var}(\mathbf{z}_j), \quad j = 1, \ldots, 2m$$

   where $\mathbf{z}_j$ is the j-th row of $\mathbf{Z}$.

4.4.3 Time Domain Features

1. **Hjorth Parameters**[58]: They are a collection of statistical measures used to characterize the attributes of a signal in the time domain. These parameters provide valuable

insights into the signal's amplitude, rate of variation, and complexity, making them useful for signal analysis, such as EEG.There are three types of parameters: activity, mobility, and complexity[33].

(a) **Activity:** Is a measure of the mean power of a signal over time, represented by the variance of its amplitude. This parameter reflects the overall energy or intensity of the signal and possesses an additive property, allowing multiple observations to be combined into a single representative value. In the frequency domain, it corresponds to the total power across the spectrum.

(b) **Mobility:** It reflects the signal's average rate of change and corresponds to the standard deviation of the frequency distribution in the spectral domain.

(c) **Complexity:** Complexity quantifies how closely the shape of a signal resembles a pure sine wave by comparing the mobility of its first derivative to that of the original signal, with values approaching 1 indicating greater similarity.

Table 4.5: The Hjorth Parameter [58]

Parameter	Notation
Activity	$\text{var}(y(t))$
Mobility	$\frac{\text{var}(y'(t))}{\text{var}(y(t))}$
Complexity	$\frac{\text{mobility}(y'(t))}{\text{mobility}(y(t))}$

2. **Root Mean Square (RMS)**[60]: is a statistical measure that represents the effective magnitude of a varying signal. It is calculated as the square root of the average of the squares of the signal values and is particularly useful for analyzing signals that contain both positive and negative components, such as sinusoidal waveforms.

3. **Zero Crossing Rate (ZCR)**[82]: is a time-domain feature that measures how often a signal changes sign, i.e., crosses the zero amplitude axis.

4. **Waveform Length (WL)**[48]: This feature reflects the degree of variation or activity in the signal over time. It quantifies signal complexity by computing the total accumulated length of the waveform within a given analysis window.

4.4.4 Band Power[49]:

EEG signals are commonly decomposed into distinct frequency bands—namely delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz)—each associated with different cognitive and physiological states. These band-specific power values serve as discriminative features in EEG signal classification tasks. In this study, band power features were extracted following the application of Common Spatial Patterns (CSP). The incorporation of band power significantly improved classification accuracy compared to using CSP features alone.

4.4.5 Time-Frequency Domain

Haar wavelet energ(HWE)[67]:

Haar Wavelet Energy (HWE) is a time–frequency domain feature used to characterize the distribution of signal energy across various frequency bands. It is derived from the Discrete Wavelet Transform (DWT) using the Haar wavelet, which is particularly well-suited for analyzing non-stationary signals like EEG due to its simplicity and localization capability. In wavelet decomposition, the EEG signal is split into multiple levels of approximation and detail coefficients. These coefficients represent the signal's behavior at different frequency ranges and resolutions. The energy of these coefficients is calculated by summing the square of each value, providing a measure of how much signal power is concentrated in each frequency band. By computing the wavelet energy for each decomposition level and EEG channel, HWE features capture the signal's energy distribution in both time and frequency. This makes them

useful for distinguishing different cognitive or motor states, such as Motor Execution (ME) and Motor Imagery (MI), based on how energy is distributed in the brain's electrical activity.

4.4.6 Classification Techniques

In this study, a diverse set of classification algorithms—including SVM, Random Forest, KNN, XGBoost, and MLP—was selected based on their demonstrated effectiveness in previous EEG-based Motor Execution (ME) and Motor Imagery (MI) classification tasks. These classifiers have consistently shown competitive performance when combined with various feature extraction techniques, making them well-suited for EEG signal classification.

1. **Support Vector Machine (SVM)[29]:**SVM is a supervised learning algorithm widely used for classification tasks due to its effectiveness in both linear and non-linear settings. The core idea is to find the optimal hyperplane that separates different classes by maximizing the margin—the distance between the hyperplane and the closest data points from each class, known as support vectors. When data is not linearly separable in the original feature space, SVM applies kernel functions to map the data into a higher-dimensional space where linear separation is possible. Commonly used kernels include the linear kernel for simple data, polynomial kernels for capturing interactions, and the Radial Basis Function (RBF) kernel for more complex, non-linear patterns. Based on the nature of the data, SVMs can thus be categorized as linear or non-linear.

2. **Random Forest (RF)[28]:** Random Forest is a powerful machine learning classifier based on ensemble learning that combines the predictions of multiple decision trees to improve accuracy and reduce overfitting. Each tree is trained on a different random subset of the dataset, and at each split, only a random subset of features is considered, which introduces diversity among the trees. For classification tasks, the final prediction is determined by majority voting; for regression, it is the average of all tree predictions. This ensemble approach enhances model stability and performance, particularly on

large and complex datasets. Additionally, Random Forest provides feature importance rankings, offering valuable insights for feature selection and interpretability.

3. **K-Nearest Neighbors (KNN)[25]:** KNN is a simple machine learning algorithm that predicts the label of a new data point by looking at the 'K' closest points in the training set. It uses distance metrics like Euclidean or Manhattan to find similarity. For classification, it uses majority voting among the K neighbors; for regression, it averages their values. Choosing the right K is important—too small can overfit, too large can underfit. KNN is easy to implement and works well for pattern recognition tasks, including EEG signal classification.

4. **Extreme Gradient Boosting (XGBoost)[26]:** XGBoost is a powerful machine learning algorithm that builds an ensemble of decision trees sequentially, where each new tree corrects the errors of the previous ones to minimize the overall loss. It uses advanced optimization techniques and regularization (L1 and L2) to reduce overfitting and enhance performance. Key parameters like learning rate, max depth, gamma, and subsample control how trees are built and how much they contribute to the final prediction. XGBoost is highly flexible, efficient, and often delivers top performance on structured data tasks, including EEG signal classification.

5. **Multilayer Perceptron (MLP)[27]:**MLP is a type of feedforward artificial neural network composed of fully connected dense layers that transform input data across dimensions. It includes an input layer (with nodes equal to input features), one or more hidden layers (which process information using weighted sums and activation functions), and an output layer (with nodes equal to the number of target classes or regression outputs). During training, data flows forward through the network (forward propagation), and errors are minimized using a loss function such as cross-entropy or mean squared error. The model learns by adjusting weights and biases through back-propagation and optimization algorithms like Stochastic Gradient Descent or Adam.

Activation functions like ReLU, Sigmoid, or Tanh introduce non-linearity, enabling MLPs to model complex patterns in data.

4.5 Dimensionality Reduction [40]

Dimensionality refers to the total count of attributes or input variables within a dataset. When this count becomes very large—typically reaching hundreds or more—the dataset is classified as high-dimensional[41]. Dimensionality reduction is the process of mapping data from this high-dimensional space into a lower-dimensional one, while striving to retain the most significant patterns and essential information. This technique is particularly valuable for easing data analysis and enhancing the visualization of complex datasets.

As the number of dimensions grows, classifier accuracy may initially increase—up to an optimal point. However, adding more features beyond this threshold without increasing the training sample size can cause performance to decline. This phenomenon, known as the curse of dimensionality, arises because data become sparse or noisy in high-dimensional spaces, reducing model effectiveness.

In this study, both linear techniques—Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)—and a non-linear method, t-Distributed Stochastic Neighbor Embedding (t-SNE), are applied for dimensionality reduction.

1. **Principal Component Analysis (PCA)[24]:** PCA reduces the number of variables (features) in a dataset by transforming the feature space into a different one where variables are uncorrelated. These new variables are called principal components. In addition, in the new feature space, a smaller subset of these features is usually enough to capture the variance in the data. With these components, it may be possible to represent important patterns of variation within the original data, facilitating easier analysis and visualization of complex, high-dimensional information.

2. **Linear Discriminant Analysis (LDA):** LDA seeks a linear combination of features that best separates different classes within a dataset. It finds the directions (or

hyperplanes in higher dimensions) that maximize the separation between class means while minimizing variance within each class, thereby enhancing the distinction between categories.

3. **t-Distributed Stochastic Neighbor Embedding (t-SNE):** t-SNE is a non-linear dimensionality reduction technique mainly used for visualizing high-dimensional data in two or three dimensions. It generates a simplified low-dimensional representation that preserves local relationships, ensuring that points close in the original high-dimensional space remain close in the reduced space.

4.6 Evaluation Metrics

We employ the following metrics to evaluate the performance of the machine learning classifiers:

- **Accuracy:** reflects the proportion of all instances that the classifier correctly:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

- **Precision:** is the proportion of positive classifications that are actually positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.3}$$

- **Recall:** is the proportion of true positive predictions out of all actual positive instances in the dataset:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.4}$$

- **F1 Score:** is the harmonic average of precision and recall, offering a single metric that balances the importance between these:

$$\text{F1 Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{4.5}$$

Chapter 5

EXPERIMENTS AND RESULTS

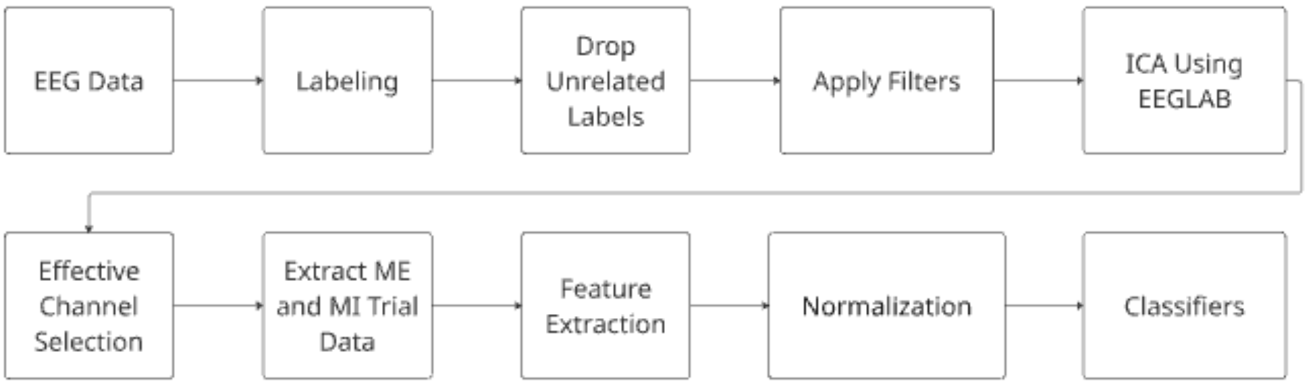


Figure 5.1: Overall EEG Data Processing Work Flow

This study employs four distinct feature extraction methods, as it was outlined in the Feature Extraction section 4.4. The overall data processing and classification pipeline is presented in Figure 5.1. Following effective channel selection, data corresponding to Motor Execution (ME) and Motor Imagery (MI) trials is extracted and subsequently used for feature extraction and classification.

In the following sections,each feature extraction approach is described in detail along with the classification performance achieved using different machine learning models.

ME and MI Trial Data: Throughout the experiments, the term "ME and MI trial data" refers to EEG recordings collected during individual trials in which participants executed either executed or imagined motor tasks. Task was defined in Section 4.1 definition 3. Each

subject completed multiple runs, with each run consisting of a sequence of trials corresponding to one of the two conditions. For all experiments, trial data from 103 subjects—extracted from a selected subset of 12 electrodes out of the original 64—were aggregated and used as input to the feature extraction and classification pipelines.

5.1 Experiment 1: Shannon Entropy-Based Features

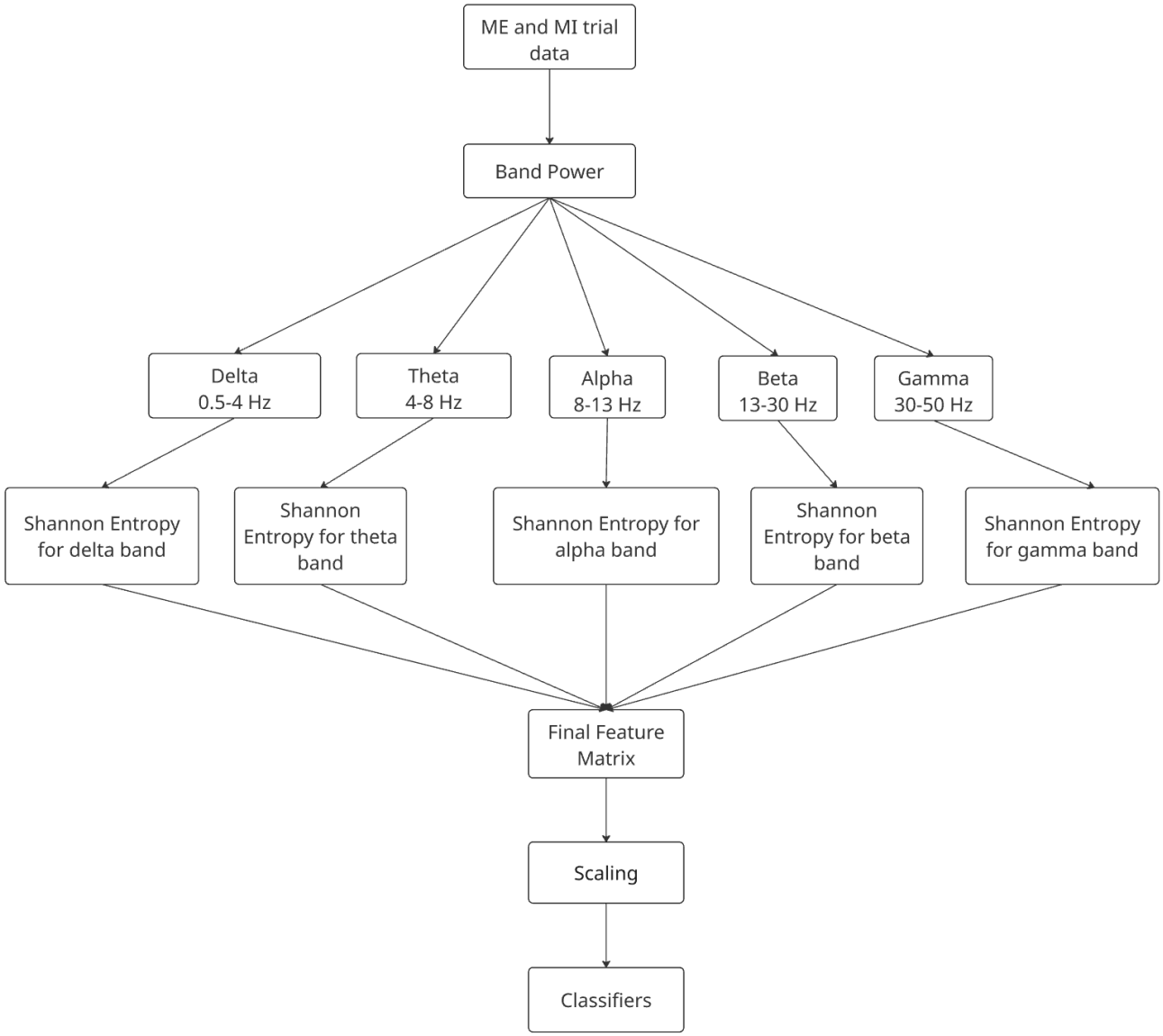


Figure 5.2: Shannon Entropy Feature Extraction Flow Diagram

The first feature extraction method implemented is based on Shannon entropy. The extracted Motor Execution (ME) and Motor Imagery (MI) trial data are used to compute band powers across the following EEG frequency ranges: Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–50 Hz). Shannon entropy is calculated for each of these band power components, and the final feature matrix is formed by aggregating the entropy features from all frequency bands along the column axis. These features are then standardized using scikit-learn's `StandardScaler` before being passed to the classifiers. The entire process is illustrated in Figure 5.2, which outlines the transformation from raw ME/MI EEG data to entropy-based feature vectors.

To understand the contribution of each frequency band, several configurations of the entropy features were evaluated:

1. Excluding the Delta band while retaining all other frequency bands
2. Excluding both Delta and Theta bands
3. Excluding Delta, Theta, and Gamma bands

Table 5.1: Performance Metrics for Different Classifiers using Shannon Entropy Based Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	51.03	51.03	51.03	51.03
RF	50.11	50.10	50.11	50.01
KNN	49.89	49.89	49.89	49.89
XGBoost	51.52	51.52	51.52	51.52
MLP	51.84	51.96	51.84	51.32

The best classification performance was achieved when only entropy features from the

Alpha, Beta, and Gamma bands were retained. Table 5.1 presents the accuracy, precision, recall, and F1 score for each classifier using this configuration.

To investigate whether reducing feature dimensionality could enhance class separability, three dimensionality reduction techniques were applied to the entropy feature set: PCA (results shown in Table 5.2), LDA (Table 5.3), and t-SNE (Table 5.4). The original feature set had 36 features, resulting in a matrix with dimensions (9220×36). Since this is a relatively low-dimensional dataset (typical high-dimensional datasets usually have over 100 features), the benefit of dimensionality reduction may be limited.

Table 5.2: Performance Metrics for Different Classifiers using PCA-Reduced Shannon Entropy Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	49.57	49.56	49.57	49.53
RF	47.89	47.86	47.89	47.79
KNN	51.63	51.63	51.63	51.63
XGBoost	48.64	48.65	48.64	48.64
MLP	50.16	50.15	50.16	50.13

Applying PCA reduced the dimensionality to (9220×17), retaining 95% of the variance. However, classification performance remained comparable to that of the original features, suggesting that PCA was ineffective in improving class separation. The main reason behind this may be that PCA performs best when the original features present *linear* correlations between them, allowing the transformation to "create" new features that capture these as they can be seen as "redundant".

LDA works well for binary classification problems and it aims to maximize the distance between the means of the classes while minimizing the variance in a class. It projects the features onto a single discriminant axis, reducing the features to one dimension (9220×1).

While LDA showed a marginal improvement over PCA, with accuracies approaching 52.57%, it still failed to yield meaningful class separation across classifiers (see Table 5.3). The reason for these results may be similar to that of PCA in the sense that LDA also makes certain assumptions about the data, particularly that 1) it is normally distributed and 2) the classes are *linearly* separable [34].

Table 5.3: Performance Metrics for Different Classifiers using LDA-Reduced Shannon Entropy Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	52.57	52.93	52.77	52.24
RF	50.00	50.00	50.00	50.00
KNN	48.48	48.48	48.48	48.48
XGBoost	49.73	49.74	49.73	49.64
MLP	51.25	51.25	51.25	51.00

Table 5.4: Performance Metrics for Different Classifiers using t-SNE-Reduced Shannon Entropy Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	49.78	49.79	49.78	49.76
RF	49.89	49.88	49.89	49.87
KNN	49.62	49.62	49.62	49.61
XGBoost	49.51	49.52	49.51	49.50
MLP	49.73	49.75	49.73	49.23

Since we cannot assume that our data satisfies the linearity assumptions required by

PCA and LDA for effective classification, we instead apply a non-linear dimensionality reduction technique: distributed stochastic neighbor embedding or t-SNE. This method tries to represent data in lower dimensionality by measuring the similarity between data points and representing it as a probability. However, this technique, did not enhance classification results either, with all classifiers achieving approximately 50% accuracy as seen in Table 5.4. This suggests that feature transformations alone – even non-linear ones – were insufficient to distinguish the two classes effectively in this dataset.

The classification accuracies obtained suggest that Shannon entropy, when used as the sole feature extraction method, does not produce a strongly discriminative feature space for differentiating ME and MI tasks. Several observations support this conclusion.

1. **Limited Class Separability:** The accuracy is close to the chance level (around 50%) across all classifiers, indicating poor separability of the two classes and feature may not be informative enough.

2. **Consistency Across Models:** All classifiers produce nearly identical metrics, implying that the limitation lies in the quality of the features rather than in the learning algorithms themselves.

3. **No Indication of Underfitting or Overfitting:** The consistently low accuracy near chance level, along with similar metrics across classifiers, indicates that the Shannon entropy features alone do not provide sufficient discriminative information for ME vs. MI classification. This suggests the limitation lies primarily in the feature representation rather than model capacity. Additionally, the similarity between accuracy, precision, recall, and F1 scores across models reflects stable performance with no evidence of high variance (overfitting).

4. **Potential Feature Overlap:** The classifiers appear to struggle with drawing clear decision boundaries, likely due to substantial overlap between the entropy values of the ME and MI tasks in the feature space.

5.2 Experiment 2:Common Spatial Patterns (CSP)

The second type of features extracted from the EEG data are spatial features derived using the Common Spatial Pattern (CSP) method. CSP aims to identify spatial filters that maximize the variance for one class while minimizing it for the other, thereby enhancing the separability between the two classes. This is achieved by performing a generalized eigenvalue decomposition on the covariance matrices of both classes and selecting an equal number of eigenvectors corresponding to the largest and smallest eigenvalues. These eigenvectors are then used as spatial filters. Figure 5.3 illustrates the flow diagram for the CSP-based feature extraction process.

Once the spatial filters were obtained, they were applied to both the training and testing datasets to generate spatially filtered signals. Subsequently, band power features were extracted from these filtered signals across the standard EEG frequency bands: Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–50 Hz). The band power features from all frequency bands were then concatenated to form the final feature matrix, which was used for classification.

Initially, classification was attempted using only the spatially filtered signals without applying band power extraction, resulting in an accuracy of approximately 65% using the XGBoost classifier. Subsequently, band power features were extracted from the spatially filtered signals, which significantly improved classification performance, yielding an accuracy of 76%. Among all configurations tested, the best results were achieved when spatial features obtained through CSP were combined with band power features extracted from the CSP-filtered signals. This highlights the effectiveness of leveraging both spatial and frequency-domain information to enhance the discriminability of motor task-related EEG patterns. In addition, various combinations of spatial and frequency-domain features were explored to assess their impact on classification performance. Specifically, the following approaches were evaluated: (1) using only spatial features, (2) extracting band power features first and then applying CSP, and (3) applying CSP first and then extracting band power features. Among

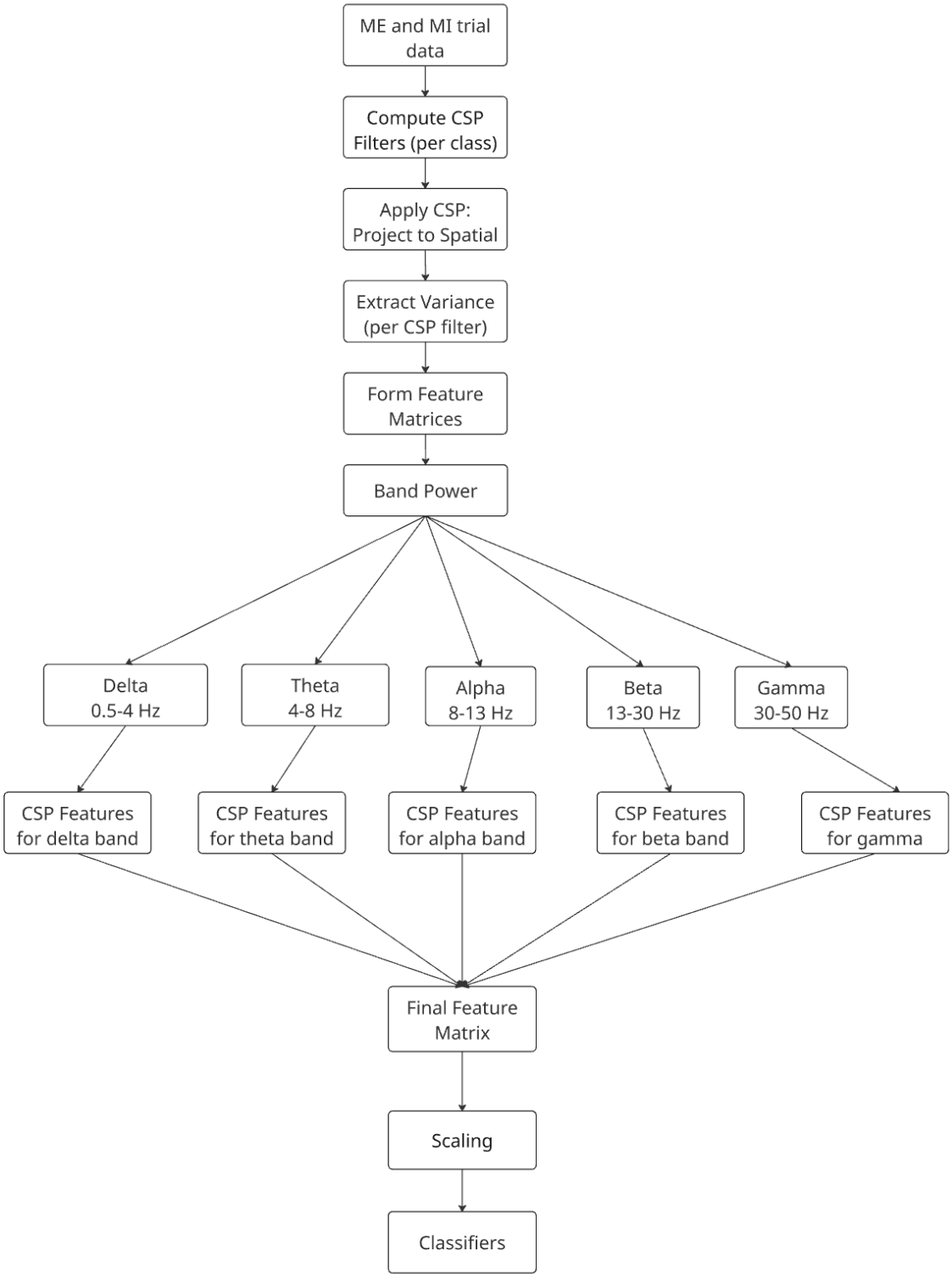


Figure 5.3: CSP Feature Extraction Flow Diagram

these, the third approach—applying CSP followed by band power extraction—yielded the highest classification accuracy. The detailed performance results for this combination are presented in Table 5.5.

Table 5.5: Performance Metrics for Different Classifiers using Spatial Features (No. of CSP filters m = 8)

	Accuracy%	Precision%	Recall%	F1 Score%
SVM	61.61	62.53	61.61	60.89
RF	77.39	77.39	77.39	77.38
KNN	77.33	77.34	77.33	77.33
XGBoost	79.34	79.34	79.34	79.34
MLP	69.20	69.20	69.20	69.20

The other combinations, particularly using only spatial features or extracting band power prior to CSP, resulted in comparatively lower performance. These results suggest that the sequence of feature extraction steps plays a critical role, with the CSP-then-band-power strategy providing the most discriminative representation for the classification task.

These results indicate that combining CSP spatial features with band power features extracted from CSP-filtered signals creates a highly informative and discriminative feature space for distinguishing between Motor Execution (ME) and Motor Imagery (MI) tasks. Several observations support the effectiveness of this combined approach:

1. **Complementary Feature Integration:**
   By integrating CSP spatial patterns (which enhance class discriminability in the spatial domain) with band power (which captures task-relevant frequency information), the combined feature set captures both spatial and spectral dynamics of motor-related brain activity.

2. **Improved Classification Performance:**
   The accuracy increases significantly across classifiers, with XGBoost achieving the highest at 79.34%. This suggests that the combined features provide a richer representation of task-specific neural signatures compared to using either CSP or band power alone.

3. **Strong Performance of Ensemble and Distance-Based Models:**
   Classifiers such as XGBoost, KNN, and Random Forest yield high and consistent performance across all metrics, indicating their ability to leverage the non-linear relationships in the combined feature space.

4. **Balanced and Robust Metrics:**
   Precision, recall, and F1 scores are closely aligned with accuracy, reflecting stable performance and no significant bias toward any class. This consistency suggests that the combined features enable effective generalization.

5. **Notable Improvement Over Entropy-Based Features:**
   Compared to the Shannon entropy-based results (Table 5.1), where performance hovered around chance level, this method offers a significant improvement, highlighting the superiority of task-specific spatial and spectral features for EEG-based motor classification.

To evaluate the impact of dimensionality reduction on CSP-based spatial features, PCA, LDA, and t-SNE were applied to the CSP feature set (generated using 8 spatial filters).The original feature matrix of shape (9220×144) was reduced to (9220×25) dimensions using PCA while retaining 95% of the variance. The classification performances obtained using the reduced feature sets are summarized in Tables 5.6, 5.7, and 5.8, respectively.

PCA reduced the dimensionality while retaining 95% of the variance. This approach maintained strong classification performance, with accuracy reaching up to 73.37% (Table 5.6), although slightly lower than that of the full CSP feature set (Table 5.5).

Table 5.6: Performance Metrics for Different Classifiers using PCA-Reduced Spatial Features (No. of CSP filters m = 8)

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	60.52	61.23	60.52	59.89
RF	70.23	70.23	70.23	70.23
KNN	73.37	73.37	73.37	73.37
XGBoost	71.69	71.69	71.69	71.69
MLP	69.14	69.14	69.14	69.14

In contrast, LDA, which projects data onto a single dimension in binary classification tasks, led to a notable decline in performance. Across all classifiers, accuracy dropped to near chance levels (52.11–56.89%) as shown in Table 5.7.

Table 5.7: Performance Metrics for Different Classifiers using LDA-Reduced Spatial Features (No. of CSP filters m = 8)

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	56.78	57.40	56.78	55.85
RF	53.09	53.09	53.09	53.08
KNN	52.11	52.12	52.11	52.11
XGBoost	54.50	54.52	54.50	54.44
MLP	56.89	56.93	56.89	56.82

The application of t-SNE, a non-linear dimensionality reduction method, produced mixed results (Table 5.8): while classifiers such as Random Forest, KNN and XGBoost retained moderate performance (70.23% , 70.34% & 69.47% accuracies), others like SVM and MLP performed closer to chance. This variability indicates that although t-SNE may uncover

underlying manifold structures, it does not consistently preserve class-separable representations suitable for all classifiers. These findings underscore that CSP features encapsulate rich spatial information critical for ME vs. MI classification, and that excessive dimensionality reduction—particularly using aggressive techniques like LDA—can obscure relevant patterns. PCA strikes a better balance by reducing redundancy while retaining useful variance.

Table 5.8: Performance Metrics for Different Classifiers using t-SNE-Reduced Spatial Features (No. of CSP filters m = 8)

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	51.79	51.83	51.79	51.54
RF	70.23	70.23	70.23	70.23
KNN	70.34	70.34	70.34	70.34
XGBoost	69.47	69.47	69.47	69.47
MLP	58.30	58.30	58.30	58.30

In conclusion, the integration of CSP and band power features effectively captures the complex neural signatures of ME and MI tasks. This multi-domain feature strategy enables robust classification performance and provides a promising direction for EEG-based BCI applications.

5.3 Experiment 3:Time Domain Features

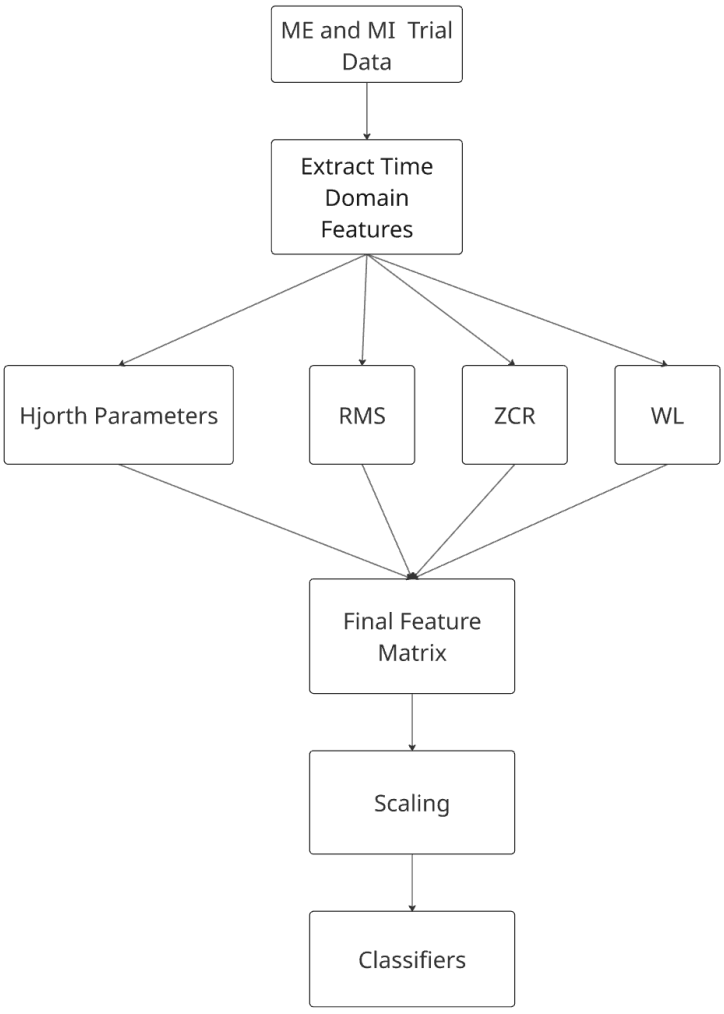


Figure 5.4: Time Domain Feature Extraction Flow Diagram

The third type of features extracted are time-domain features. These include Hjorth parameters—Activity, Mobility, and Complexity—as well as Root Mean Square (RMS), Zero Crossing Rate (ZCR), and Waveform Length (WL). The flow diagram illustrating the extraction of these time-domain features is shown in Figure 5.4. All extracted features were

combined to form the final feature matrix, which was then used for classification.

Several combinations of time-domain features were explored to evaluate their individual and collective impact on classification performance. The following configurations were tested:

1. Activity features only
2. Mobility features only
3. Complexity features only
4. Zero Crossing Rate (ZCR) only
5. Waveform Length (WL) only
6. Root Mean Square (RMS) only
7. A combination of Complexity, WL, and RMS
8. All time-domain features combined

The best classification performance was achieved using all time-domain features, as summarized in Table 5.9.

Table 5.9: Performance Metrics for Different Classifiers using Time Domain Features

	Accuracy%	Precision%	Recall%	F1 Score%
SVM	61.50	61.51	61.50	61.36
RF	70.82	70.90	70.82	70.83
KNN	78.90	78.91	78.90	78.90
XGBoost	72.56	72.57	72.56	72.56
MLP	70.99	71.00	70.99	70.99

The classification performance using time-domain features reveals several important insights regarding their effectiveness for distinguishing Motor Execution (ME) and Motor Imagery (MI) tasks:

1. **Moderate Discriminative Power of Time-Domain Features:** Classifiers demonstrate moderate accuracy ranging from 61.50% (SVM) to 78.90% (KNN), indicating that time-domain features capture meaningful patterns related to motor tasks, though not as strongly as combined spatial-spectral features.

2. **Superior Performance of Distance-Based Classifier (KNN):** KNN outperforms all other classifiers with an accuracy of 78.90% and consistently high precision, recall, and F1 scores (78.90%). This suggests that KNN effectively leverages the local structure in the time-domain feature space to distinguish between ME and MI.

3. **Consistent Performance of Ensemble and Neural Models:** Random Forest, XGBoost, and Multilayer Perceptron classifiers deliver similar performance (70.82–72.56%), demonstrating their ability to model non-linear relationships within the time-domain features but not reaching the highest accuracy achieved by KNN.

4. **Limited Performance of SVM:** Support Vector Machine yields the lowest accuracy (61.50%), suggesting that its linear or kernel-based decision boundaries may not effectively capture the complexity present in time-domain features.

5. **Balanced Classification Metrics Across Models:** Precision, recall, and F1 scores closely mirror accuracy for all classifiers, indicating no strong bias toward false positives or false negatives and reflecting stable classification performance.

6. **Robustness Across Cross-Validation Folds:** Classification accuracies obtained from all five cross-validation folds fall within a similar range, showing minimal variance. This consistency suggests that the model's performance is stable and generalizes well across different data subsets, reducing the likelihood of overfitting or data-specific bias.

Table 5.10: Performance Metrics for Different Classifiers using PCA-Reduced Time Domain Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	58.51	58.55	58.51	58.52
RF	65.08	65.21	65.08	65.07
KNN	64.97	64.97	64.97	64.97
XGBoost	64.48	64.51	64.48	64.49
MLP	69.52	69.52	69.52	69.50

Dimensionality reduction applied to time-domain features yielded results consistent with trends observed in CSP features. The original feature matrix had a shape of (9220×72), which PCA reduced to (9220×10) while preserving most of the variance in the data. As shown in Table 5.10, PCA retained reasonable performance, with MLP achieving the highest accuracy (69.52%) and other classifiers maintaining moderate results (58.51–65.08%).

In contrast, LDA (Table 5.11) led to a substantial decline in performance across all models, with accuracies ranging from 51.52% to 57.38%, suggesting that projecting data to a single dimension likely suppressed discriminative variability.

Table 5.11: Performance Metrics for Different Classifiers using LDA-Reduced Time Domain Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	57.32	57.49	57.32	57.28
RF	51.52	51.52	51.52	51.52
KNN	52.60	52.57	52.60	52.57
XGBoost	55.64	55.74	55.64	55.63
MLP	57.38	57.52	57.38	57.35

Table 5.12: Performance Metrics for Different Classifiers using t-SNE-Reduced Time Domain Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	52.01	52.12	52.01	51.98
RF	66.00	66.04	66.00	66.00
KNN	66.16	66.17	66.16	66.16
XGBoost	65.40	65.42	65.40	65.41
MLP	53.96	53.87	53.96	53.66

t-SNE results (Table 5.12) were mixed: classifiers like KNN and RF retained decent accuracy (66.16%), while others such as SVM and MLP dropped closer to chance levels. These findings suggest that certain dimensionality reduction techniques, particularly those involving strong compression or complex non-linear mappings, may obscure critical temporal features necessary for distinguishing between ME and MI tasks.

5.4 Experiment 4: Time-Frequency Domain

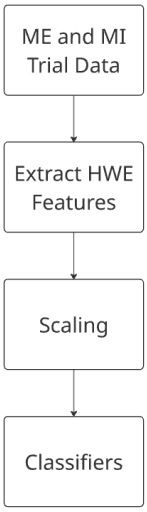


Figure 5.5: Haar Wavelet Energy Feature Extraction Flow Diagram

The fourth feature extraction method employed was based on the time-frequency domain using Haar Wavelet Energy (HWE). This approach leverages the ability of wavelet transforms to capture both temporal and spectral characteristics of non-stationary signals such as EEG. For each trial of Motor Execution (ME) and Motor Imagery (MI) data, a Discrete Wavelet Transform (DWT) was applied using the Haar wavelet as the mother wavelet. A three-level decomposition was performed, resulting in one set of approximation coefficients (A3) and three sets of detail coefficients (D1, D2, D3) for each EEG channel.The energy of each coefficient set was computed and used as the feature representation. Specifically, for each decomposition level and coefficient type, the energy was calculated as the sum of squared values of the coefficients. These energy values effectively represent how the signal's power is distributed across different frequency bands and time intervals.

The resulting feature vectors from all channels were then concatenated to form the final HWE feature matrix. Prior to classification, the feature matrix was scaled using standard

normalization techniques to ensure uniformity in feature ranges. The classification per-

Table 5.13: Performance Metrics for Different Classifiers using Time-Frequency Domain Features

	Accuracy%	Precision%	Recall%	F1 Score%
SVM	58.13	58.34	58.13	57.87
RF	72.13	72.16	72.13	72.12
KNN	68.17	68.17	68.17	68.17
XGBoost	73.16	73.16	73.16	73.15
MLP	67.62	68.29	67.62	67.32

formance using time-frequency domain features, specifically Haar Wavelet Energy (HWE), provides insights into the effectiveness of capturing both temporal and spectral characteristics of EEG signals. The resulting feature vectors were standardized and used to train multiple classifiers. Table 5.13 presents the evaluation metrics obtained for each model. Key observations from these results are as follows:

1. **Moderate-to-High Discriminative Capacity of HWE Features:** Classification performance using Haar Wavelet Energy (HWE) features shows moderate to high accuracy, ranging from 58.13% (SVM) to 73.16% (XGBoost), suggesting that HWE captures relevant time-frequency information for distinguishing ME and MI tasks.

2. **Best Performance with Tree-Based Classifiers:** XGBoost and Random Forest outperform other classifiers, achieving accuracies of 73.15% and 72.13% respectively. This indicates that ensemble-based, non-linear models are particularly effective in leveraging the multi-scale decomposition captured by wavelet features.

3. **Limited Performance of SVM:** Support Vector Machine shows the lowest classification performance (58.13%), implying that the decision boundaries required to separate

ME and MI classes in the HWE feature space may be highly non-linear and not well-suited for SVM kernels used in this study.

4. **KNN and MLP Provide Competitive Results:** K-Nearest Neighbors and Multilayer Perceptron classifiers perform comparably (68.17% and 67.62% accuracy, respectively), reflecting the effectiveness of local similarity-based and shallow neural approaches in handling the time-frequency feature structure.

5. **Balanced Metrics Across Classifiers:** Precision, recall, and F1 scores closely match the accuracy values for all models, indicating balanced classification performance with no significant skew toward false positives or false negatives.

Table 5.14: Performance Metrics for Different Classifiers using PCA-Reduced Time-Frequency Domain Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	56.67	56.92	56.67	56.26
RF	63.88	63.88	63.88	63.88
KNN	62.36	62.37	62.36	62.36
XGBoost	63.29	63.30	63.29	63.28
MLP	64.86	64.87	64.86	64.86

The HWE feature matrix, originally shaped (9220,48), was reduced to 10 dimensions using PCA while preserving 95% variance. As shown in Table 5.14, PCA moderately impacted performance, with MLP achieving the highest accuracy (64.86%), and other models such as RF, KNN, and XGBoost maintaining acceptable performance (62.36–63.88%).

In contrast, LDA reduced the feature space to a single dimension, leading to a notable drop in classification accuracy across all models (Table 5.15), with accuracies ranging from 51.30% to 57.00%. This suggests that projecting multi-band wavelet energy features onto a

Table 5.15: Performance Metrics for Different Classifiers using LDA-Reduced Time-Frequency Domain Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	56.07	56.44	56.07	55.41
RF	51.95	51.96	51.95	51.93
KNN	51.30	51.31	51.30	51.27
XGBoost	53.42	53.42	53.42	53.41
MLP	57.00	57.31	57.00	56.55

single axis suppresses critical discriminative information.

Table 5.16: Performance Metrics for Different Classifiers using t-SNE-Reduced Time-Frequency Domain Features

	Accuracy %	Precision %	Recall%	F1 Score %
SVM	52.60	52.63	52.60	52.46
RF	65.08	65.08	65.08	65.08
KNN	63.50	63.50	63.50	63.50
XGBoost	62.58	62.58	62.58	62.58
MLP	52.77	5.87	52.77	52.30

Results from t-SNE (Table 5.16) were mixed: RF and KNN retained relatively better accuracy (65.08% and 63.50%, respectively), while SVM and MLP dropped to near-chance levels (52.60% & 52.77%). These findings indicate that dimensionality reduction—especially aggressive linear projections or non-linear embeddings—can distort or eliminate important time-frequency patterns in HWE features, ultimately degrading ME vs. MI classification performance.

5.5 Comparative Analysis

In this section, we aggregate and discuss observed results comparatively across the four different feature extraction methods and classifiers used. It also offers interpretations for the outcomes and exploring the underlying reasons behind them.

5.5.1 Feature Extraction Methods and Classification Strategies

Classifiers will perform differently depending on the input dataset used. Datasets are not just about the observations, but equally – or perhaps more – importantly is the set of features used, both in terms of quantity and type. For this reason, a large amount of our work has focused feature selection, and previous sections in this chapter have outlined experiments focusing on each of these strategies. In these experiments we presented and discussed the methodology and classification results for four distinct feature extraction techniques with the goal of evaluating their effectiveness with different classifiers for differentiating EEG signals associated with Motor Execution (ME), Motor Imagery (MI) as detailed in Chapter 4. We now turn to discuss the differences from a higher-level perspective, when results assembled side by side as in Figure 5.6. These figures summarize accuracies and F1 scores obtained with each classifier respectively, combined with each feature extraction method.

As we can observe, the best classification performance was achieved using **Common Spatial Pattern (CSP) features**, particularly when combined with band power features. The reason may be due to CSP being a methodology designed to separate signals, such as EEG ones, into subcomponents, essentially working as a filter that enhances the signal. When combined with XGBoost (79.34%) classifier, the highest accuracy was achieved. **XGBoost** is an ensemble classifier (using decision trees), greedy algorithm that tends to perform well with large datasets. Very close comes Random Forest (77.39%), another tree-based ensemble algorithm, while **KNN** (77.33%) – a distance based method – that works well with spatial data. When it comes to other feature extraction strategies, we observe that those three classifiers tend to perform well, except with Shannon Entropy.

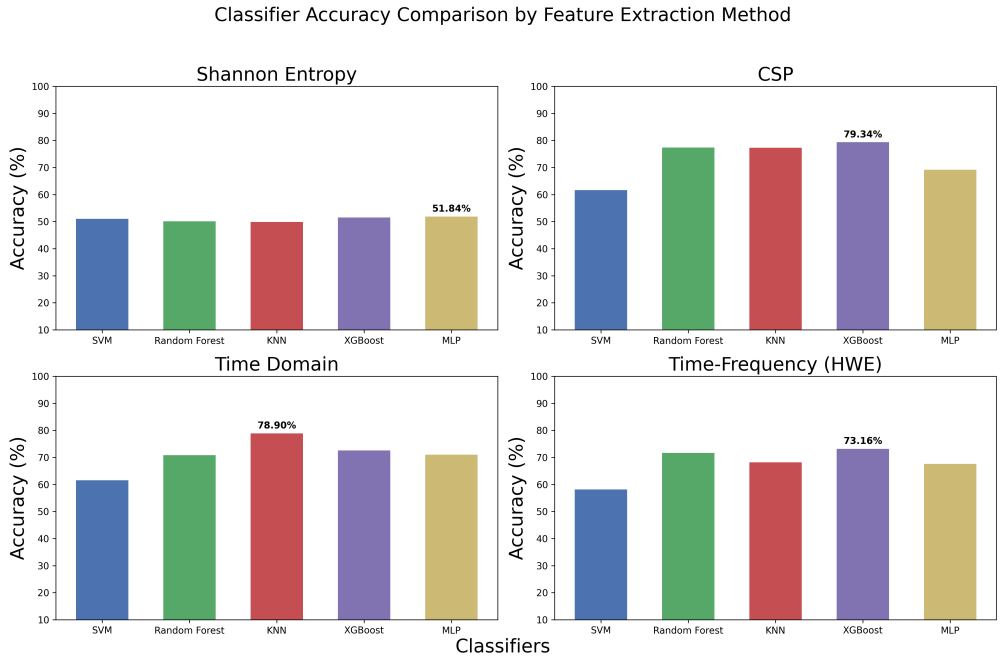


Figure 5.6: Comparison of Classifier Accuracy Using Different Feature Extraction Techniques

On the other hand, **MLP** (a non-linear Neural-network-based classifier) does not perform well with this data in general, which suggests that our data may be insufficient to capture complex patterns, particularly since MLP tends to work better with high dimensionality data and our dataset had originally 36 features only. When it comes to **SVM**, it is not surprising that it has lowest performance across all feature extraction approaches, particularly since SVM works best with small datasets where finding a hyperplane that separates two classes and our dataset is large.

Analyzing **Shannon Entropy**, we need to consider that it was employed to extract features that quantify the amount of information present in the EEG signal. In general, higher entropy values in EEG data are indicative of noisy data, whereas lower values suggest signal is more informative.For the dataset used in this study, the entropy values observed were in lower range, indicating that the EEG signals were relatively clean and less noisy. However, a key observation was that the range of entropy values for both classes(ME and MI) was nearly identical. This substantial overlap in entropy distributions between two

classes significantly reduced their separability during classification. Classifiers were unable to effectively distinguish between the two classes based on Shannon entropy alone.

Next, we assess the generalization of the best performing classifier for each of the feature extraction method. We use cross-validation scores, generally are used to evaluate model performance on "unseen" data during the training phase with positive outcomes showing that the model can generalize well beyond "seen" data used during training [36]. For training models, K-fold validation was used (with 10 folds). The feature-extraction-method/classifier of interest are presented in Figure 5.7, showing that scores are quite consistent on each fold and very closely match the corresponding accuracy results (Figure 5.6), which indicates that the models generalize well and it would be expected that their behavior on new data would be very similar.

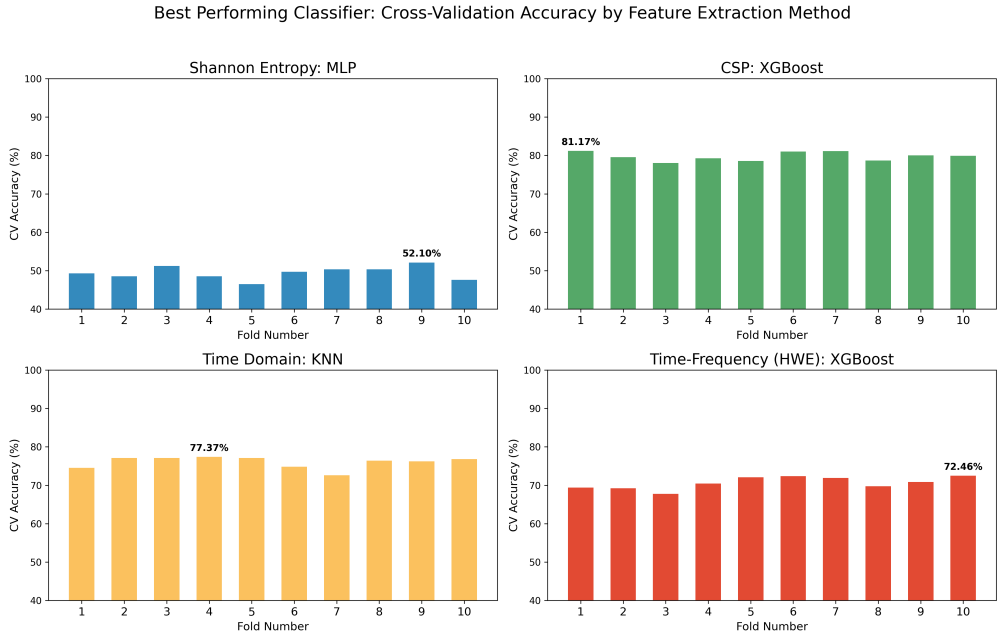


Figure 5.7: Best Performing Classifier: Cross-Validation Accuracy by Feature Extraction Method

While overall accuracy provides a general performance overview, class-wise evaluation reveals nuanced differences. We use summary Figure 5.8 to evaluate classification performance

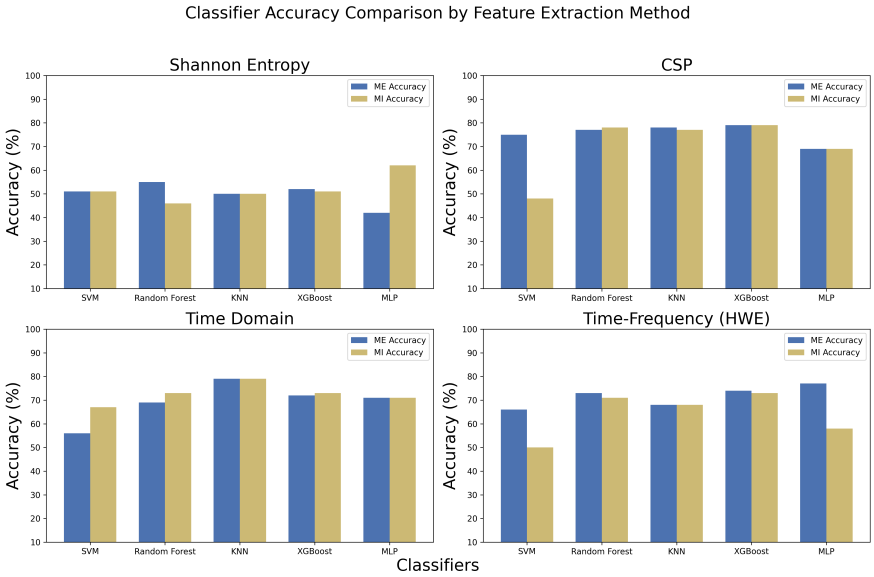


Figure 5.8: Class-wise Accuracy Comparison for ME and MI Tasks Across Classifiers and Feature Extraction Methods

specifically for the two classes of interest, ME and MI, across different feature extraction methods. In most cases, classifiers performed similarly over both classes. The instances showing noticeable divergence are SVM with CSP, suggesting that the hyperplane created by the SVM cannot well-round MI data, as it likely intersects ME in higher dimensions. MLP did also has noticeable difference when used with Shannon Entropy or Time-Frequency features. Overall, we observe once more that CSP effectively separates these classes, given the underlying coordinate space of the features.

Furthermore, precision, recall, and F1 scores presented in each of the experiment sections, were consistent with the accuracy values, indicating balanced classification performance across both classes. Despite these results, the best accuracies still fall below 80%, highlighting room for improvement. The uniformly close metric values suggest that while the classifiers are balanced, the current features may lack sufficient discriminative power. Since ME and MI involve similar neural substrates for related motor functions, distinguishing between them remains inherently challenging. This physiological overlap, while beneficial for

rehabilitation, complicates the classification process by limiting the feature space variance available for separation.

It is also worth noting that in both Figures 5.6 and 5.8 we can see the Effectiveness of **Time Domain** feature extraction, where four classifiers exceeded 70% accuracy: RF (70.80%), KNN (78.90%), XGBoost (72.56%), and MLP (70.99%). The consistent advantageous performance across classifiers indicates that time-domain features effectively capture discriminative variance between ME and MI. Notably, the better performance of KNN suggests that distance-based models, when combined with temporal features, are well-suited for this classification task.

5.5.2 Model Assessment

In here, we evaluate classification performance using confusion matrices and Area Under the Curve of Receiver Operating Characteristic (ROC) to help with the interpretation of the findings covered in the previous subsection. The ROC plots the true positive rate (TPR) against the false positive rate (FPR) at every possible interval, and it is commonly used to visualize a classifier's overall performance by measuring the area under the curve [32]. We focus on those Feature-Extraction-Method/Classifier combinations that yield highest and lowest accuracy (as seen in Figure 5.6) as well as those combinations with close or disparate class accuracy depicted in Figure 5.8.

CSP with XGBoost

The combination of CSP feature extraction with the XGBoost classifier resulted in one of the highest classification accuracies in this study. Figure 5.9 shows the confusion matrix and corresponding ROC curve plot. The confusion matrix indicates that, out of a total of 1844 observations, the count of true positives (732 or $\sim$ 39.7%) and true negatives (731 or $\sim$ 39.6%) are nearly the same, suggesting that the model's performance is well balanced across both classes: the total number of observations correctly identified is 1463 out of 1844 (or $\sim$80% of them).

The ROC curve further supports the model's effectiveness, with an area under the curve (AUC) of 0.87 (where an AUC of 1 being the ideal). This high AUC demonstrates strong discriminative ability between the two motor tasks. The curve rises sharply toward the top-left corner, indicating high *sensitivity* (the model correctly identifies the positive class with high accuracy) and a low false positive rate (or high specificity, meaning most negative instances are correctly classified).

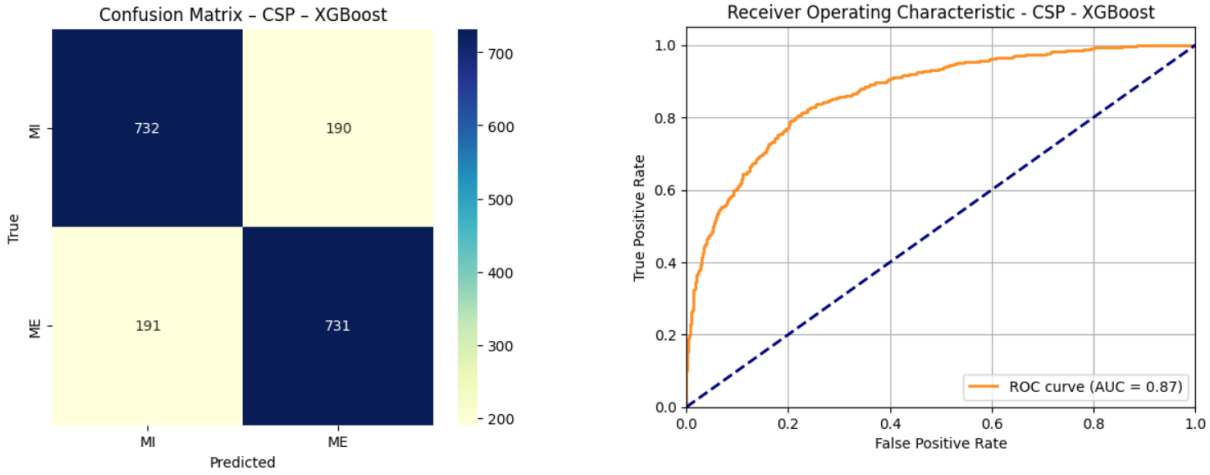


Figure 5.9: Confusion Matrix and ROC curve for CSP with XGBoost

These results support corroborate the effectiveness of CSP feature extraction, combined with the XGBoost classifier in distinguishing between MI and ME in EEG signals, which was observed in Figures 5.6 and 5.8.

CSP with KNN

This combination also yielded high accuracy, Figure 5.10 shows the confusion matrix and ROC curve. The confusion matrix indicates a slightly higher number of false positives (217 or 11.76%) and false negatives (201 or 10.9%) compared to the XGBoost classifier.

The classifier shows a slight tendency to predict ME tasks (negative class) more accurately than MI tasks (positive class), with $\sim$ 39.9% of true negatives (a count of 721) marginally

exceeding true positives with $\sim 38.23\%$ (a count of 705). This difference is minimal, with an overall accuracy differs by only about 1%, indicating balanced and reliable performance. We note that the False Positives and Negatives are both $\sim 10\%$.

Furthermore, the corresponding AUROC of 0.85, demonstrates the model's ability to differentiate between MI and ME. Although slightly lower than the XGBoost model, these results support the suitability of KNN when combined with CSP features.

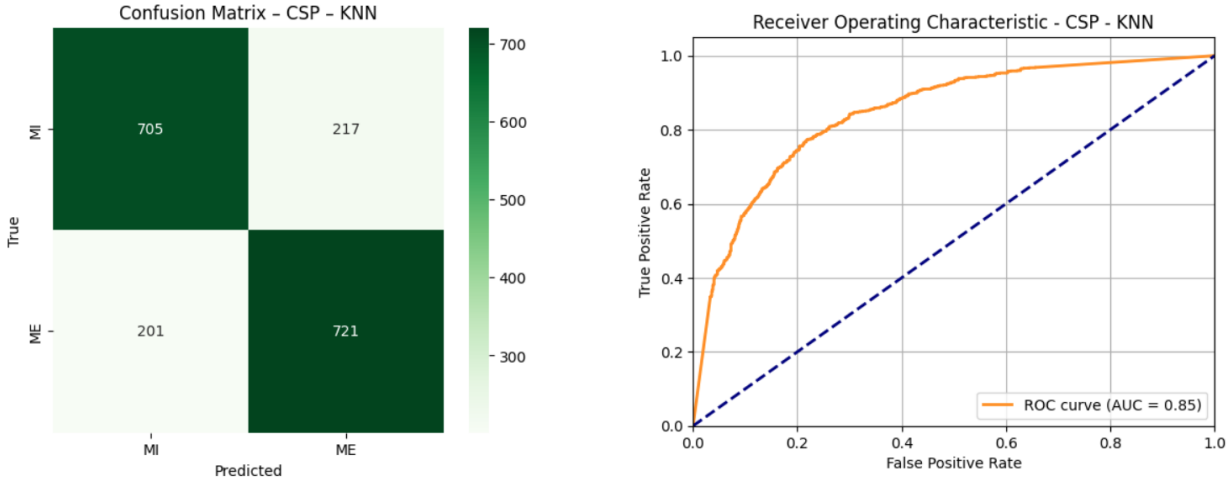


Figure 5.10: Confusion Matrix for CSP with KNN

CSP with SVM

Is an interesting combinations that yielded high accuracy for ME classification as seen in Figure 5.8). Its confusion matrix and corresponding ROC curve are depicted in Figure 5.11. Proportionally speaking, the False Negative count is high, with a proportion of $\sim 26\%$ (compared to CSP with XGBoost with $\sim 10\%$). More importantly, the True Negative count is low, with a proportion of 12.41%, supporting the idea that the two classes overlap in this space making it difficult for a hyperplane to separate them, resulting in MI points to be confused with ME, drastically decreasing the accuracy of MI in the class-wise comparison. Not surprisingly, The corresponding AUC under the ROC curve is only 0.67, further confirming

this model's difficulty discriminating between the two classes.

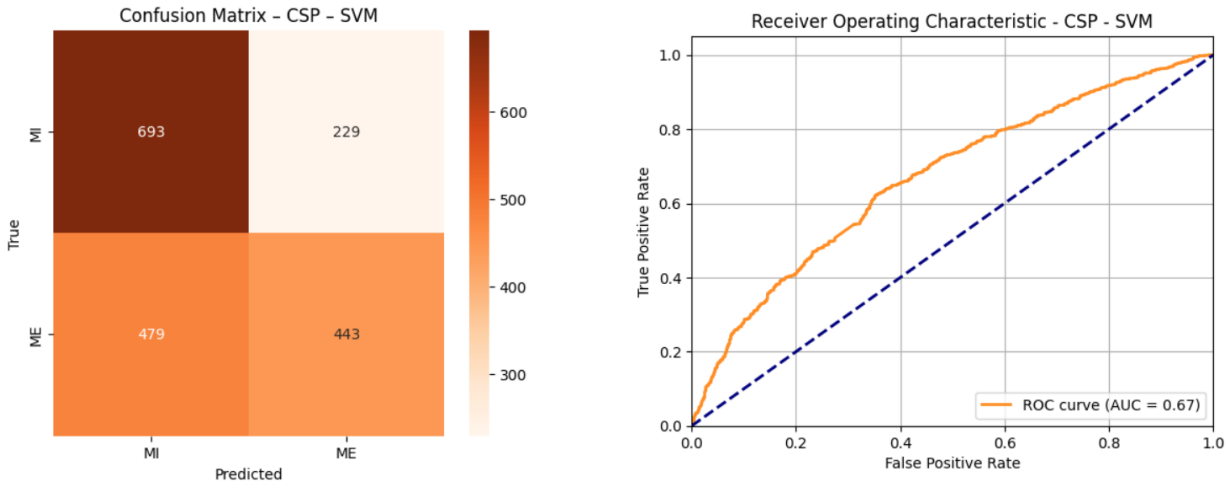


Figure 5.11: Confusion Matrix for CSP with SVM

Time Domain with with KNN

The KNN classifier achieved high accuracy, as shown in Figure 5.12, which presents the confusion matrix and ROC curve. The false positive count (190) matches that of CSP with XGBoost, while the false negative count (199) is slightly higher, suggesting some confusion between ME and MI due to their similarity. Nonetheless, the relatively low FP and FN values indicate that time-domain features are effective in distinguishing the two classes.

The ROC curve rises steeply up to 0.5 and continues to increase gradually to 0.9, reflecting a strong true positive rate relative to false negatives. With an AUC of 0.86, the classifier shows solid performance using this feature set, as further supported by Figures 5.6 and 5.8.

Shannon Entropy with MLP

Like CSP with SVM, has disparate accuracy between classes, but in this case, it is ME points that get confused with MI as seen in the confusion matrix in Figure 5.13 shows the confusion matrix and the corresponding ROC curve. As pointed out previously, Shannon entropy assumes that lower entropy channels are less noisy and carry relevant information,

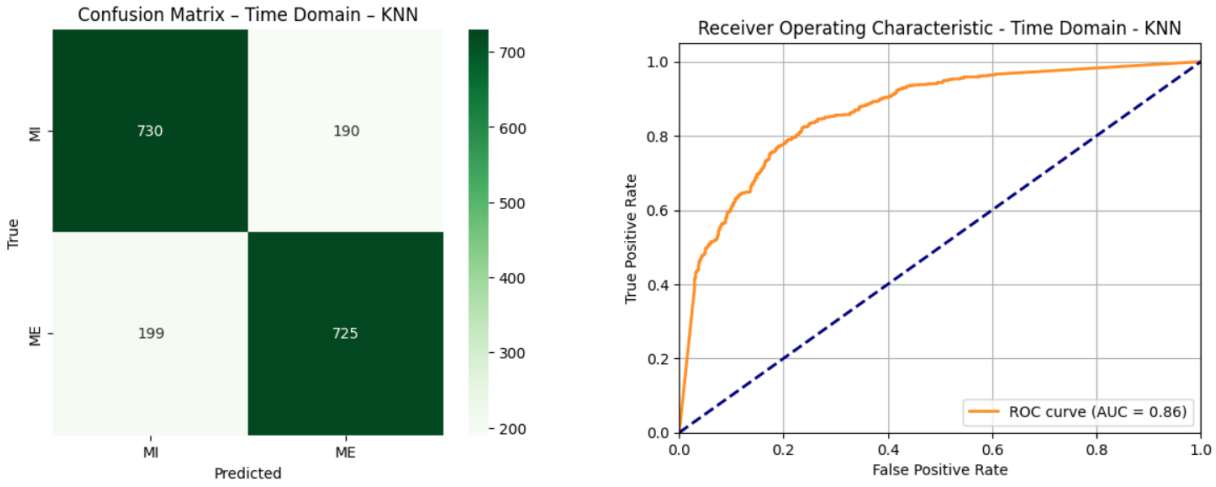


Figure 5.12: Confusion Matrix and AUROC curve for Time Domain with KNN

however, in the case of MI and ME, such channels have nearly the same entropy values. Nearly all the ROC curves for Shannon entropy look like the one in this figure, and the confusion matrices indicate large proportion of False (Negatives or Positives) – these images are available in Appendix A. This supports our findings that using Shannon entropy is not a good approach for this problem.

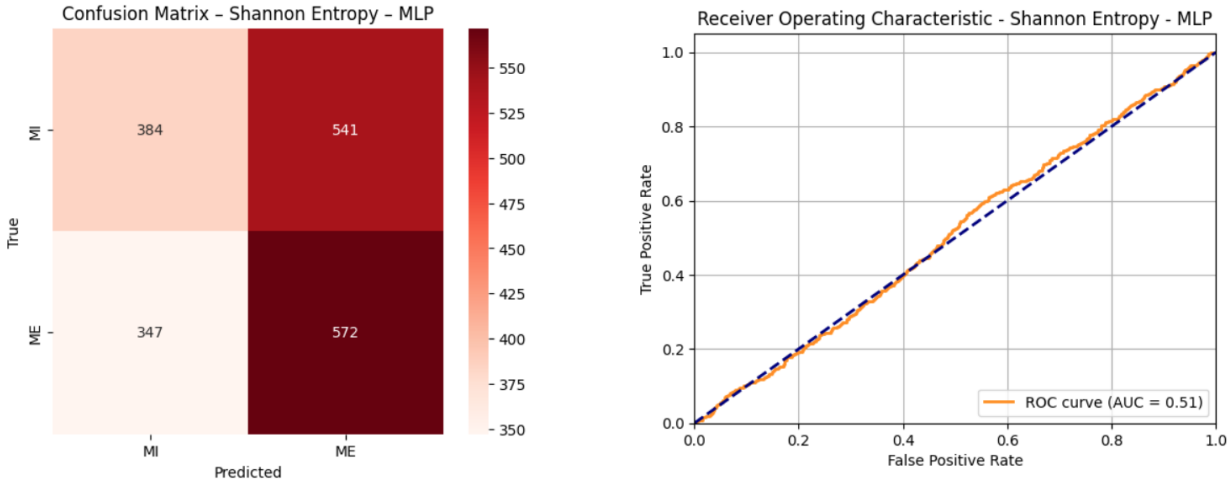


Figure 5.13: Confusion Matrix and AUROC curve Shannon Entropy with MLP

5.6 Influence of Subject-Specific Variability

Several studies in the literature support the subject-specific nature of classification performance in this domain. For instance, Yang et al. [89] reported a significant variation in AUROC curve values among subjects using their proposed Average Score (AS) algorithm, with subject-specific AUC values ranging from 0.598 (for one subject) to 0.998 (for a different one). Similarly, Barachant et al. [8] investigated four different feature extraction approaches across nine subjects and achieved their highest accuracy of 86.1% for Subject 9 using TSLDA, but the overall best average accuracy across all approaches was limited to 70.2%.

The study by Ma et al. [49] evaluated several spatial feature extraction strategies and their combinations, and observed that the optimal method varied across subjects. Similar findings were reported by Lotte [48] and Moufassih [53], reinforcing the notion that subject-specific variability significantly influences classification outcomes. Notably, these prior studies typically involved fewer than 11 subjects. While they reported high individual accuracies exceeding 90%, their average performance remained around 80

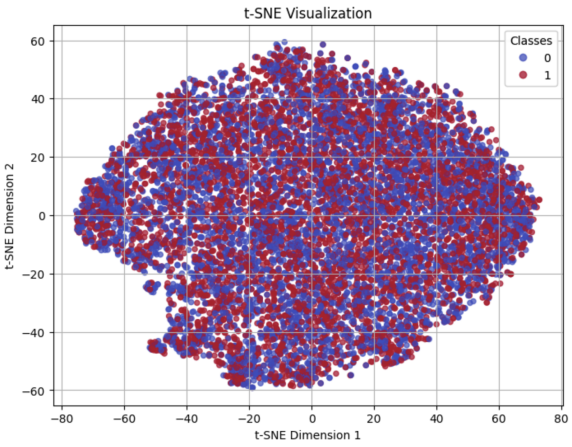
In contrast, the present study utilizes data from 103 subjects, offering a more generalized and representative evaluation. Although the highest average classification accuracies in this study remain below 80%, the alignment of precision, recall, and F1 scores indicates robust and consistent classifier behavior.

Lastly, in the study by Lazcano et al. [45], a subset of the same PhysioNet dataset was used, focusing on 10 subjects and data from a *single electrode* (FC3). Their classification task aimed to distinguish between ME, MI and rest, achieving a perfect 100% accuracy using a Quadratic Discriminant classifier. This outcome reinforces the reliability of the PhysioNet EEG dataset. Their results indicate that the dataset itself does not present major limitations. Instead, the real challenge lies in distinguishing between Motor Execution and Motor Imagery, which share overlapping brain activation patterns and are therefore more difficult to classify.

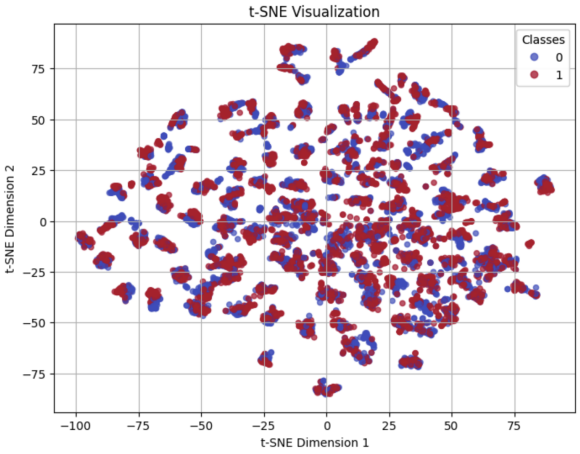
5.7 Dimensionality Reduction and Accuracy

This study evaluated four feature matrices derived from different feature extraction methods: Shannon Entropy (36 features), Common Spatial Patterns (4144 features), time-domain features (72 features), and time–frequency domain features (48 features). Among these, only the CSP feature matrix can be considered high-dimensional. The best classification performance was achieved using CSP with eight spatial filters ($m = 8$). Reducing the number of filters ($m = 2, 4, 6$) resulted in decreased accuracy, and increasing the number to $m = 12$ did not lead to further improvement. Thus, the optimal performance was obtained using the full CSP feature set with $m = 8$, and additional dimensionality reduction did not enhance classification performance.

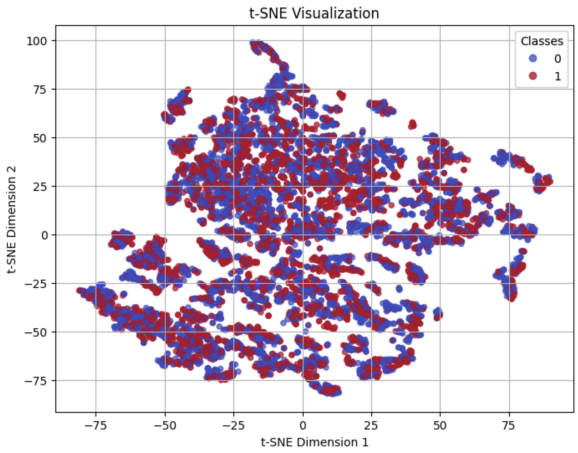
Visual analysis using t-SNE plots (Figure 5.14) reveals that the data is inherently difficult to separate. Shannon Entropy features show significant overlap between classes, while time-domain and time–frequency features provide slightly better class separation but still lack well-defined clusters. CSP features produced more distinguishable clusters, although class mixing within clusters remained evident. This indicates that it is challenging to separate the two classes using dimensionality reduction alone. While CSP with band power extraction demonstrated relatively better class separability, further performance improvement may require combining it with other feature extraction methods or exploring alternative strategies for extracting CSP features.



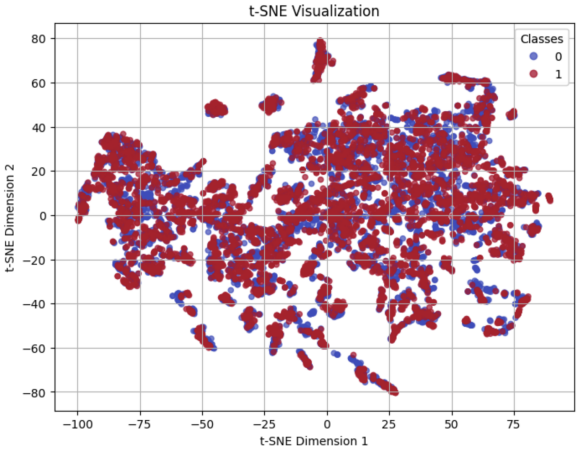
(a) Shannon Entropy Features



(b) CSP Features



(c) Time Domain Features



(d) Time Frequency Domain Features

Figure 5.14: t-SNE Visualizations for Different Feature sets.

Chapter 6

CONCLUSIONS AND FUTURE WORK

This chapter highlights general findings and outlines potential future directions that could contribute to enhancing classification accuracy for distinguishing Motor Imagination in subsequent studies.

Identifying and/or isolating Motor Imagination signals in EEG data remains a challenging problem in neuroscience. In this research, we proposed combining Independent Component Analysis (ICA) with a range of feature extraction techniques to enable better classification between ME and MI in EEG data. Our results confirm that this multimodal feature strategy, particularly when leveraging Common Spatial Pattern (CSP) features and Band Power enables for the discrimination of these two signals. The highest average accuracy of 79.34% was achieved using XGBoost with CSP+Band Power features.

The hybrid approach proposed in this work is a first of its kind, providing a means to classify MI and ME with comparable accuracy to that presented in previous work.

The findings validate the assumption that the integration of ICA and diverse feature spaces can enhance signal separability between ME and MI tasks, despite their inherent physiological and cognitive overlap. The results also highlight the role of spatial-domain transformations in maximizing class-discriminative information, especially when used in conjunction with efficient channel selection.

Working with Imagery tasks poses a key challenge, given the subjective nature of motor imagery tasks. Unlike motor execution, where the physical movement provides observable confirmation, motor imagery relies solely on the participant's internal cognitive process. Only the subject truly knows whether they were able to imagine the movement accurately and consistently. This introduces variability in the quality of the recorded signals for Motor

Imagery, as the effectiveness of imagination can differ significantly across individuals. Consequently, this variability may affect the performance of classification models and presents a fundamental challenge in distinguishing MI from ME EEG signals.

6.1 Contributions

The main contributions of this work involve the use of a *robust* dataset and the design of a *novel* classification strategy that combines the use of feature extraction methods.

6.1.1 Large Dataset for Analysis and Baseline

We introduced the use of a large dataset (Physionet [31]) comprising 103 subjects, totaling over 1,500 recordings and approximately 42 hours of data (detailed in Section 4.1). This allows to overcome issues with the small in-house collected dataset, which also potentially hides artifacts resulting from device issues that cannot easily be differentiated from noise or relevant signals. Furthermore, this larger dataset enhances the generalizability of the findings and provides balanced classes.

In addition, this dataset was used to feed into our classification pipeline, presenting a more reliable way to study its behavior.

6.1.2 Integration of Multiple Feature Extraction and Classification Techniques

The core aspect of this work was to research the use of ***multiple*** feature extraction methods (time-domain, frequency-domain, time-frequency, and spatial-domain techniques) and test them in combination with various commonly used classifiers (SVM, Random Forest, KNN, XGBoost, and MLP) to evaluate results and create comparative baseline for classifying EEG signals.

Various combinations were tested, such as Shannon entropy with Band Power, and spatial-domain features combined with band power. Multiple sets of time-domain features were extracted, and different combinations were evaluated to determine which one offers best

classification opportunity. While the classification efforts focused around MI and ME, our strategy can be extended to be used on other classes of EEG signals.

6.1.3 Effective Channel Selection

During exploratory data analysis, it was observed that not all 64 EEG channels contributed significantly during MI and ME tasks. Based on the association between Brodmann areas and EEG electrode placement, and informed by prior EEG studies on effective channel selection, the number of electrodes was first reduced to 46. This was further narrowed down to 12 electrodes by referencing additional research. This effective channel selection strategy helped reduce computational complexity without compromising classification accuracy.

Moreover, processing data from 64 channels (translating to 64 features) can be computationally prohibitive for researchers without access to high performance computing hardware. At the time of this study there was no previous work to classify between these two types of signals leveraging feature extraction.

The only other study that aims to differentiate MI from ME signals starting with 64 channels (same dataset) is the recently published work in [76] using Riemannian geometry-based decoding algorithms for ME and MI classification, achieving a maximum accuracy of 78.97% with 64 channels, and 74.24% using 29 channels. In comparison, our channel selection strategy results in 12 electrodes, achieving an average accuracy of 79.34%. This highlights the effectiveness of the proposed feature engineering and channel selection approach in achieving high accuracy with lower compu4tational cost, supporting its practical applicability in real-world brain-computer interface systems.

6.1.4 Foundations for Future EEG Data Collection

We have identified key factors for EEG-based experiment designs, which will serve as a basis for future data acquisition efforts under the Smart NeuroRehab initiative: the outcomes of this work help inform robust and scalable data collection strategies for future research.

6.2 Limitations

While the study offers valuable insights into EEG-based classification of motor tasks, it also has several limitations, as listed below:

- **Dataset:** Although the dataset was collected by the developers of the BCI2000 system under controlled experimental conditions and is publicly available via PhysioNet, it was not acquired in our own lab or within a real-time clinical setting.

- **Binary Classification Scope:** The study is limited to a binary classification problem—distinguishing between motor execution (ME) and motor imagery (MI)—and does not explore more complex, multi-class motor task scenarios.

- **Limited to Traditional ML Approaches:** The study does not incorporate deep learning models or automated end-to-end pipelines, which are increasingly prominent in EEG-based research.

- **Healthy Subject Data Only:** The dataset consists solely of EEG recordings from healthy individuals performing upper-limb movements (specifically, "making a fist"). As such, the findings may not generalize to clinical populations with neurological disorders or to tasks involving lower-limb movements.

6.3 Future Work

The work presented in this thesis paves the road for future work opportunities, which we can organize into the following categories

Dataset

Wile our focus has been MI an ME signals, the dataset also includes recordings for lower-limb (foot movement) tasks, which remain unexplored. Future work could incorporate these additional classes to develop multi-class classification models, enabling more granular analysis

of motor-related brain activity. Additionally, this dataset is relatively underutilized in the literature, with only two known published studies employing it. One focused on action-vs-rest classification with a single electrode [45], and the aforementioned work in [76] with 64 electrodes. The promising results achieved in this study suggest that further improvements are possible through the integration of hybrid or advanced feature extraction methods.

Classification Strategies and Model Improvements

While this study employed classical machine learning models (SVM, RF, KNN, and MLP) and four hand-engineered feature extraction techniques, future research could explore more advanced methods. Deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are well-suited to capture spatial and temporal dependencies in EEG data and could lead to performance gains. Additionally, automatic feature extraction using deep learning or symbolic representation methods such as One Dimension-Aggregate Approximation (1D-AX) [85] could further enhance classification accuracy. Personalized classification strategies—such as subject-specific models or calibration mechanisms—may also improve model generalizability across individuals. Furthermore, transitioning from offline classification to real-time deployment would be a critical next step toward practical BCI applications in neurorehabilitation and assistive technologies. Finally, integrating EEG with other physiological signals such as EMG may provide a more robust and multimodal understanding of motor intent.

Motor Planning and Its Overlap with Execution in EEG Data

This study does not consider removing motor planning (MP) before motor execution (ME). Several studies [39][55][46] suggest that MP occurs prior to the actual action. Before performing the action, the subject thinks about it, which is called motor planning. This process happens during data collection after the cue is given to perform the action. If the data timestamps start from the cue appearance and end at the completion of the action, the data will include MP activity. Since MP is not the same as ME, it is important to consider ways

to remove or separate MP from ME, ideally during data collection.

For example, the study by Jean-Jacques and Hardwick [59] concludes that it is challenging to separate MP from ME because the transition is instantaneous. However, it is still worth considering a time window to exclude MP from ME data.

Another study by Bernardo et al. collected EEG data from 13 subjects and found that MP occurred approximately 250 ms before movement onset. The exact start of MP may vary between individuals but generally occurs between the cue appearance and actual movement. Similarly, Sommer et al. [80] used Transcranial Magnetic Stimulation (TMS) data from six subjects and reported that MP lasts about 200 ms before movement execution.

Appendix A

APPENDIX A: SUPPLEMENTAL DATA

Shannon Entropy Confusion Matrices

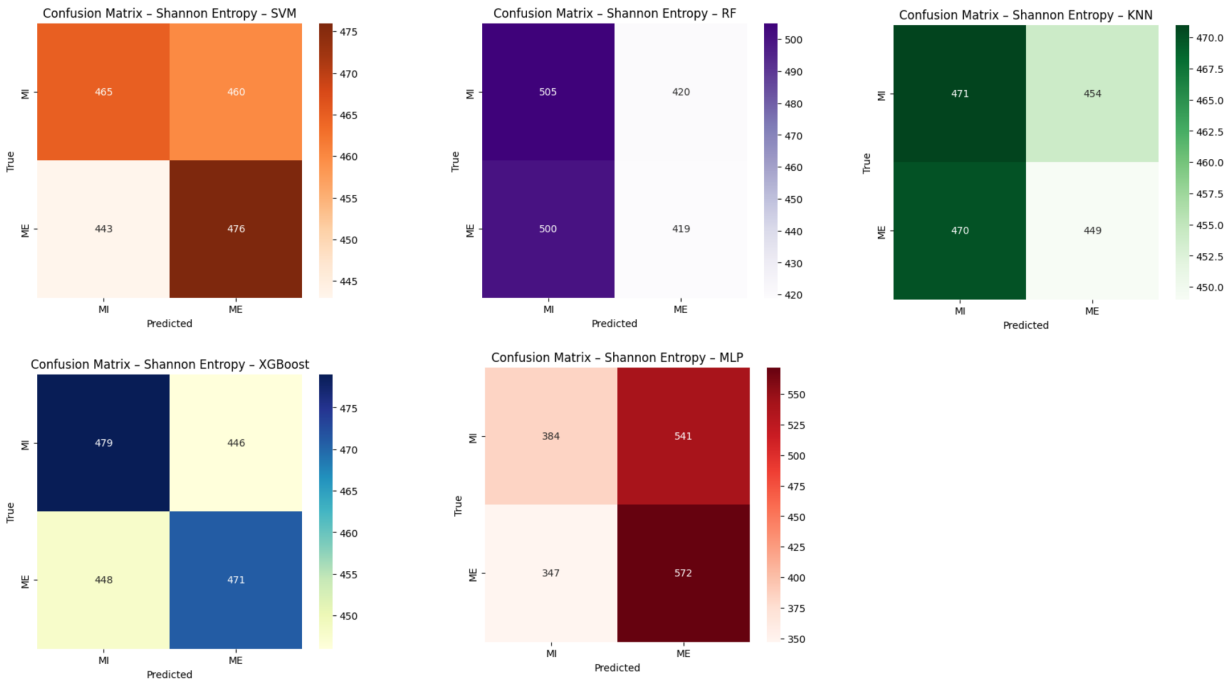


Figure A.1: Confusion Matrices for Shannon Entropy

Common Spatial Pattern (CSP) Confusion Matrices

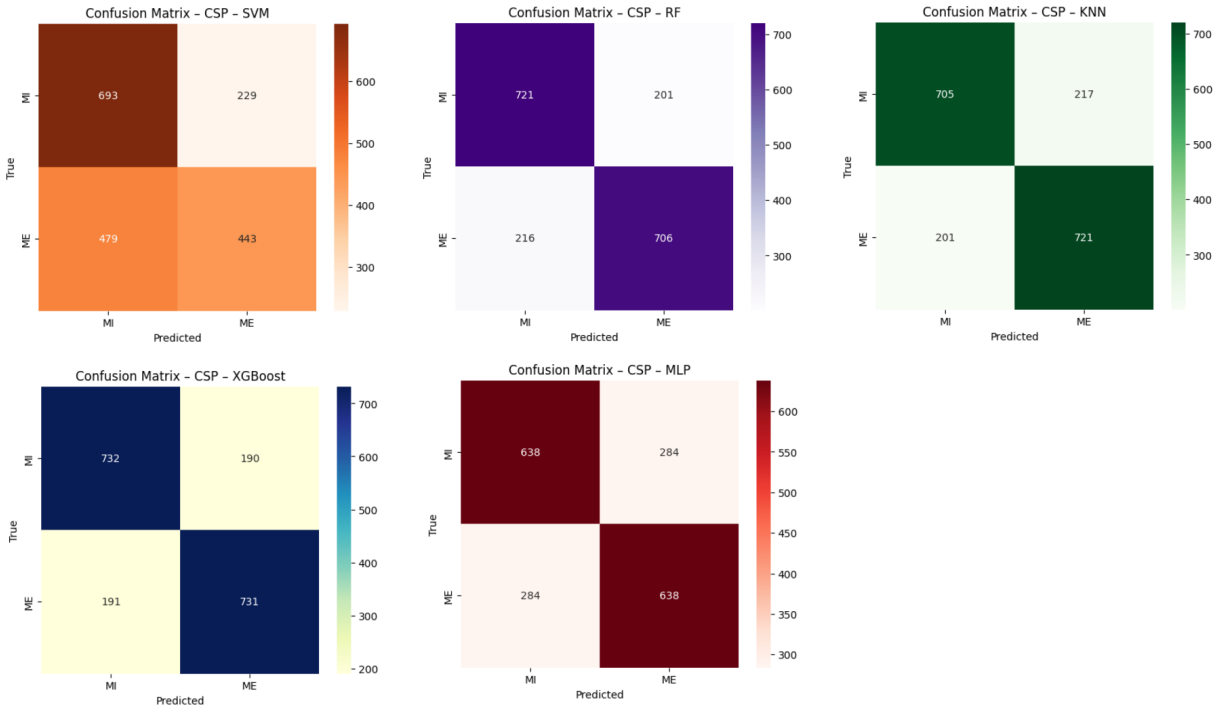


Figure A.2: Confusion Matrices for CSP

BIBLIOGRAPHY

[1] Brain-computer interface. Accessed: March 2024.

[2] Shannon entropy measures for eeg signals in tinnitus. *Neuroscience Letters*, 762:136153, 2021.

[3] A. Marushima H. Watanabe A. Zaboronok S. Watanabe A. Matsumura K. Suzuki Y. Matsumaru E. Ishikawa A. Anastasiev, H. Kadone. Supervised myoelectrical hand gesture recognition in post-acute stroke patients with upper limb paresis on affected and non-affected sides. page 8733, 2013.

[4] M. C . Brady P. Langhorne G. E. Mead J. Mehrholz F. van Wijck A. Pollock, S. E. Farmer. Interventions for improving upper limb function after stroke. page CD010820, 2014.

[5] Sarah N Abdulkader, Ayman Atia, and Mostafa-Sami M Mostafa. Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal*, 16(2):213–230, 2015.

[6] Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed A Bencherif, and M Shamim Hossain. Multilevel weighted feature fusion using convolutional neural networks for eeg motor imagery classification. *Ieee Access*, 7:18940–18950, 2019.

[7] Sarwar Ayesha and Emmady Prabhu. Spatial neglect. 5 2022.

[8] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2011.

[9] Berlin Brain-Computer Interface (BBCI) Competitions. Bci competition iv, 2008. Accessed: 2025-05-21.

[10] BrainyCalc Insights. Eeg filters: Types and examples, December 2024. Accessed: 2025-05-08.

[11] Brmlab. Closest 10-10 electrode position to each brodmann area. https://brmlab.cz/project/brain_hacking/broadmannarea, 2011. Accessed: 2025-05-09.

[12] C. Browne. Evaluating the effectiveness of preprocessing methods on motor classification accuracy in eeg data. *A Thesis, University of Washington Bothell, CSSE,*, 2023.

[13] C. Browne. Evaluating the effectiveness of preprocessing methods on motor classification accuracy in eeg data. *A Thesis, University of Washington Bothell, CSSE,*, 2024.

[14] Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Chan Jun. Eeg datasets for motor imagery brain–computer interface. *GigaScience*, 6(7):gix034, 2017.

[15] Jean Decety, Marc Jeannerod, Daniel Durozard, and Gabriel Baverel. Central activation of autonomic effectors during mental simulation of motor actions in man. *The Journal of physiology*, 461(1):549–563, 1993.

[16] Peter Dechent, Klaus-Dietmar Merboldt, and Jens Frahm. Is the human primary motor cortex involved in motor imagery? *Cognitive Brain Research*, 19(2):138–144, 2004.

[17] Arnaud Delorme. Ica for dummies, n.d. Accessed: 2025-05-08.

[18] Ruth Dickstein and Judith E Deutsch. Motor Imagery in Physical Therapist Practice. *Physical Therapy*, 87(7):942–953, 07 2007.

[19] Guido Dornhege, Benjamin Blankertz, Gabriel Curio, and K-R Muller. Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms. *IEEE transactions on biomedical engineering*, 51(6):993–1002, 2004.

[20] Hauke Dose, Jakob Skadkær Møller, Sadasivan Puthusserypady, and Helle K Iversen. A deep learning mi-eeg classification model for bcis. In *26th European Signal Processing Conference*, pages 1690–93. IEEE, 2018.

[21] H Henrik Ehrsson, Stefan Geyer, and Eiichi Naito. Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. *Journal of neurophysiology*, 2003.

[22] National Center for Adaptive Neurotechnologies. National center for adaptive neurotechnologies, 2005.

[23] S. Gahatraj. Exploring neurological rehabilitation enhancements: Combining machine learning and robotics in an edge computing environment. *A Thesis, University of Washington Bothell, CSSE,*, 2024.

[24] GeeksforGeeks. Principal component analysis with python. https://www.geeksforgeeks.org/principal-component-analysis-with-python/, 2021. Accessed: 2025-06-10.

[25] GeeksforGeeks. K-nearest neighbor(knn) algorithm, 2025. Accessed: 2025-05-10.

[26] GeeksforGeeks. Ml — xgboost (extreme gradient boosting), 2025. Accessed: 2025-05-10.

[27] GeeksforGeeks. Multi-layer perceptron learning in tensorflow, 2025. Accessed: 2025-05-10.

[28] GeeksforGeeks. Random forest algorithm in machine learning, 2025. Accessed: 2025-05-10.

[29] GeeksforGeeks. Support vector machine algorithm, 2025. Accessed: 2025-05-10.

[30] P Geethanjali, Y Krishna Mohan, and Jinisha Sen. Time domain feature extraction and classification of eeg data for brain computer interface. In *2012 9th international conference on fuzzy systems and knowledge discovery*, pages 1136–1139. IEEE, 2012.

[31] Ary L. Goldberger, Luis A. Nunes Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, C-K. Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals, 2000. Accessed: 2025-05-16.

[32] Google Developers. Classification: Roc and auc, 2025. Accessed: 2025-05-19.

[33] Bo Hjorth. Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 29(3):306–310, 1970.

[34] IBM. What is linear discriminant analysis (lda)? hhttps://www.ibm.com/think/topics/linear-discriminant-analysis, 2023. Accessed: 2025-06-10.

[35] Tomohiko Igasaki, Yugo Kuramura, and Junya Takemoto. Motor imagery, execution, and observation classification using small amount of eeg data with multiple two-class cnns. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 443–446. IEEE, 2021.

[36] Ivan Palomares Carrascosa. A complete guide to cross-validation. https://www.statology.org/complete-guide-cross-validation/, 2025. Accessed: 2025-06-09.

[37] Pari Jahankhani, Vassilis Kodogiannis, and Kenneth Revett. Eeg signal classification using wavelet feature extraction and neural networks. In *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06)*, pages 120–124. IEEE, 2006.

[38] John Hopkins Medicine. Neurological Rehabilitation.

[39] Scott H Johnson. Thinking ahead: the case for motor imagery in prospective judgements of prehension. *Cognition*, 74(1):33–70, 2000.

[40] Vijay Kanade. What is dimensionality reduction? https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-dimensionality-reduction/, 2023. Accessed: 2025-06-10.

[41] Vipin Katara. Dimensionality reduction: Pca vs lda vs t-sne. https://medium.com/analytics-vidhya/dimensionality-reduction-pca-vs-lda-vs-t-sne-681636bc686, 2020. Accessed: 2025-06-10.

[42] Aida Khorshidtalab, Momoh-Jimoh E Salami, and Mahyar Hamedi. Robust classification of motor imagery eeg signals using statistical time–domain features. *Physiological measurement*, 34(11):1563, 2013.

[43] Teresa J Kimberley, Gauri Khandekar, Laura L Skraba, Jessica A Spencer, Emily A Van Gorp, and Sarah R Walker. Neural substrates for motor imagery in severe hemiparesis. *Neurorehabilitation and Neural Repair*, 20(2):268–277, 2006.

[44] J Satheesh Kumar and P Bhuvaneswari. Analysis of electroencephalography (eeg) signals and its categorization–a study. *Procedia engineering*, 38:2525–2536, 2012.

[45] Alicia Guadalupe Lazcano-Herrera, Rita Q Fuentes-Aguilar, and Mariel Alfaro-Ponce. Eeg motor/imagery signal classification comparative using machine learning algorithms. In *2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, pages 1–6. IEEE, 2021.

[46] Hanfei Li, Chenyu Fan, Ke Chen, Hongyu Xie, Guohui Yang, Haozheng Li, Xiangtong Ji, Yi Wu, and Meng Li. Brain activation during motor preparation and execution in patients with mild cognitive impairment: An fnirs study. *Brain Sciences*, 15(4):333, 2025.

[47] Sheng Li, Derek G Kamper, Jennifer A Stevens, and William Z Rymer. The effect of motor imagery on spinal segmental excitability. *Journal of Neuroscience*, 24(43):9674–9680, 2004.

[48] Fabien Lotte. A new feature and associated optimal spatial filter for eeg signal classification: Waveform length. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pages 1302–1305. IEEE, 2012.

[49] Zhen Ma, Kun Wang, Minpeng Xu, Weibo Yi, Fangzhou Xu, and Dong Ming. Transformed common spatial pattern for motor imagery-based brain-computer interfaces. *Frontiers in Neuroscience*, 17:1116721, 2023.

[50] Ingo G Meister, Timo Krings, Henrik Foltys, Babak Boroojerdi, Mareike Müller, Rudolf Töpper, and Armin Thron. Playing piano in the mind—an fmri study on music imagery and performance in pianists. *Cognitive Brain Research*, 19(3):219–228, 2004.

[51] Kai J Miller, Gerwin Schalk, Eberhard E Fetz, Marcel Den Nijs, Jeffrey G Ojemann, and Rajesh PN Rao. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*, 107(9):4430–4435, 2010.

[52] Minks. Eeg: electrode positions & broadmann atlas. https://www.cnblogs.com/minks/p/5388623.html, 2016. Accessed: 2025-05-09.

[53] Mustapha Moufassih, Oussama Tarahi, Soukaina Hamou, Said Agounad, and Hafida Idrissi Azami. Spectral feature extraction from eeg based motor imagery using common spatial patterns. In *2022 2nd international conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, pages 1–6. IEEE, 2022.

[54] Th Mulder. Motor imagery and action observation: cognitive tools for rehabilitation. *Journal of neural transmission*, 114:1265–1278, 2007.

[55] Bence Nanay. Motor imagery and action execution. 2020.

[56] National Center for Biotechnology Information. *Electroencephalography*. National Center for Biotechnology Information (US), 2018.

[57] Marc R. Nuwer. 10-10 electrode system for eeg recording. *Clinical Neurophysiology*, 129(5):1103, 2018.

[58] Seung-Hyeon Oh, Yu-Ri Lee, Hyoung-Nam Kim, et al. A novel eeg feature extraction method using hjorth parameter. *International Journal of Electronics and Electrical Engineering*, 2(2):106–110, 2014.

[59] Jean-Jacques Orban de Xivry and Robert M Hardwick. A control policy can be adapted to task demands during both motor execution and motor planning. *Journal of Neurophysiology*, 133(1):232–244, 2025.

[60] Ashish Panat, Anita Patil, and Gayatri Deshmukh. Feature extraction of eeg signals in different emotional states. In *IRAJ conference*, 2014.

[61] Carlo A Porro, Maria Pia Francescato, Valentina Cettolo, Mathew E Diamond, Patrizia Baraldi, Chiara Zuiani, Massimo Bazzocchi, and Pietro E Di Prampero. Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. *Journal of Neuroscience*, 16(23):7688–7698, 1996.

[62] A. Raj. Innovative rehabilitation approach for upper limb neurologic conditions using mixed-reality simulation and eeg/emg biofeedback. *A Thesis, University of Washington Bothell, CSSE,*, 2024.

[63] Ricardo Ramos-Aguilar, J Arturo Olvera-López, Ivan Olmos-Pineda, and Susana Sánchez-Urrieta. Feature extraction from eeg spectrograms for epileptic seizure detection. *Pattern Recognition Letters*, 133:202–209, 2020.

[64] Mamunur Rashid, Norizam Sulaiman, Anwar PP Abdul Majeed, Rabiu Muazu Musa, Ahmad Fakhri Ab Nasir, Bifta Sama Bari, and Sabira Khatun. Current status, challenges, and possible solutions of eeg-based brain-computer interface: a comprehensive review. *Frontiers in neurorobotics*, 14:515104, 2020.

[65] Ericka Janet Rechy-Ramirez and Huosheng Hu. Ces-513 stages for developing control systems using emg and eeg signals: A survey. 2011.

[66] Muriel Roth, Jean Decety, Monica Raybaudi, Raphael Massarelli, Chantal Delon-Martin, Christoph Segebarth, Stéphanie Morand, Angelo Gemignani, Michel Décorps, and Marc Jeannerod. Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study. *Neuroreport*, 7(7):1280–1284, 1996.

[67] G Roy, AK Bhoi, and S Bhaumik. A comparative approach for mi-based eeg signals classification using energy, power and entropy. *IRBM*, 43(5):434–446, 2022.

[68] Simanto Saha, Khondaker A Mamun, Khawza Ahmed, Raqibul Mostafa, Ganesh R Naik, Sam Darvishi, Ahsan H Khandoker, and Mathias Baumert. Progress in brain computer interface: Challenges and opportunities. *Frontiers in systems neuroscience*, 15:578875, 2021.

[69] Saeid Sanei and Jonathon A Chambers. *EEG signal processing*. John Wiley & Sons, 2013.

[70] Paulo Saraiva. On shannon entropy and its applications. *Kuwait Journal of Science*, 50(3):194–199, 2023.

[71] Gerwin Schalk. Bci competition iii, 2005.

[72] SciPy Developers. scipy.linalg.eigh — SciPy v1.11.3 Manual, 2024. Accessed: 2025-05-07.

[73] Abdullah Al Shiam, Kazi Mahmudul Hassan, Md Rabiul Islam, Ahmed MM Almassri, Hiroaki Wagatsuma, and Md Khademul Islam Molla. Motor imagery classification using effective channel selection of multichannel eeg. *Brain Sciences*, 14(5):462, 2024.

[74] Abdullah Al Shiam, Kazi Mahmudul Hassan, Md Rabiul Islam, Ahmed MM Almassri, Hiroaki Wagatsuma, and Md Khademul Islam Molla. Motor imagery classification using effective channel selection of multichannel eeg. *Brain Sciences*, 14(5):462, 2024.

[75] Zaid Shuqfa, Abdelkader Nasreddine Belkacem, and Abderrahmane Lakas. Decoding multi-class motor imagery and motor execution tasks using riemannian geometry algorithms on large eeg datasets. *Sensors*, 23(11), 2023.

[76] Zaid Shuqfa, Abdelkader Nasreddine Belkacem, and Abderrahmane Lakas. Decoding multi-class motor imagery and motor execution tasks using riemannian geometry algorithms on large eeg datasets. *Sensors*, 23(11):5051, 2023.

[77] Zaid Shuqfa, Abderrahmane Lakas, and Abdelkader Nasreddine Belkacem. Increasing accessibility to a large brain–computer interface dataset: Curation of physionet eeg motor movement/imagery dataset for decoding and classification. *Data in Brief*, 54:110181, 2024.

[78] A. Siddalingaswami. Architecting a cybersecurity framework for an edge computing medical application for stroke patient rehabilitation. *A Capstone Project, University of Washington Bothell, CSSE*, 2023.

[79] Jason Sleight, Preeti Pillai, and Shiwali Mohan. Classification of executed and imagined motor movement eeg signals. *Ann Arbor: University of Michigan*, 110, 2009.

[80] Martin Sommer, Joseph Classen, Leonardo G Cohen, and Mark Hallett. Time course of determination of movement direction in the reaction time task in humans. *Journal of Neurophysiology*, 86(3):1195–1201, 2001.

[81] Jiuxiang Song, Qiang Zhai, Chuang Wang, and Jizhong Liu. Eeggan-net: enhancing eeg signal classification through data augmentation. *Frontiers in Human Neuroscience*, 18:1430086, 2024.

[82] Igor Stancin, Mario Cifrek, and Alan Jovic. A review of eeg signal features and their application in driver drowsiness detection systems. *Sensors*, 21(11):3786, 2021.

[83] Cathy M Stinear, Winston D Byblow, Maarten Steyvers, Oron Levin, and Stephan P Swinnen. Kinesthetic, but not visual, motor imagery modulates corticomotor excitability. *Experimental brain research*, 168:157–164, 2006.

[84] Christoph Stippich, Henrik Ochmann, and Klaus Sartor. Somatotopic mapping of the human primary sensorimotor cortex during motor imagery and motor execution by functional magnetic resonance imaging. *Neuroscience letters*, 331(1):50–54, 2002.

[85] Ping Wang, Aimin Jiang, Xiaofeng Liu, Jing Shang, and Li Zhang. Lstm-based eeg classification in motor imagery tasks. *IEEE transactions on neural systems and rehabilitation engineering*, 26(11):2086–2095, 2018.

[86] Wearable Sensing. EEG Electrodes.

[87] Yu-Te Wu, Tzu Hsuan Huang, Chun Yi Lin, Sheng Jia Tsai, and Po-Shan Wang. Classification of eeg motor imagery using support vector machine and convolutional neural network. In *2018 international automatic control conference (CACS)*, pages 1–4. IEEE, 2018.

[88] Norashikin Yahya, Huwaida Musa, Zhong Yi Ong, and Irraivan Elamvazuthi. Classification of motor functions from electroencephalogram (eeg) signals based on an integrated method comprised of common spatial pattern and wavelet transform framework. *Sensors*, 19(22):4878, 2019.

[89] Jun Yang, Zhengmin Ma, and Tao Shen. Multi-time and multi-band csp motor imagery eeg feature classification algorithm. *Applied Sciences*, 11(21):10294, 2021.

[90] Florian Yger, Maxime Berar, and Fabien Lotte. Riemannian approaches in brain-computer interfaces: A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1753–1762, 2017.

[91] Aihua Zhang, Bin Yang, and Ling Huang. Feature extraction of eeg signals using power spectral entropy. In *2008 international conference on BioMedical engineering and informatics*, volume 2, pages 435–439. IEEE, 2008.

[92] Catherine Zuckerman. The human brain, explained, 2009.

[93] Catherine Zuckerman. The human brain, explained, 2009. Accessed: 21st March 2024.