



The University of Queensland  
School of Electrical Engineering and Computer Science

*BIOE6100 - Fundamentals of Neuroengineering*

***Final Report***

Investigating The Accuracy of EEGNet for  
Motor Movement and Motor Imagery  
Classification

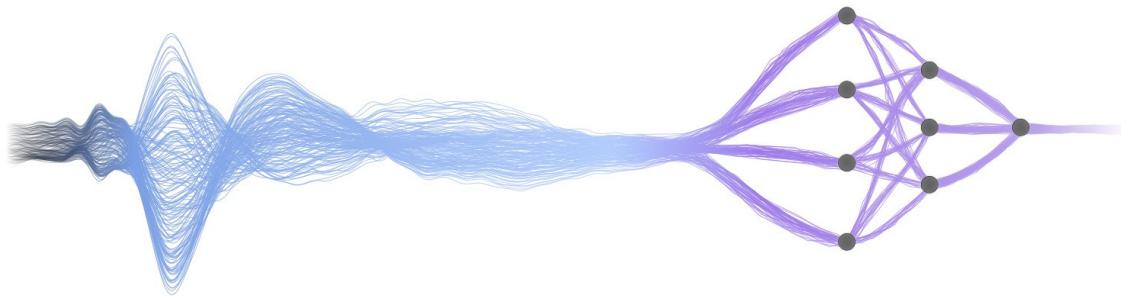


Image sourced from the Cognitive Computational Neuroscience (CCN) Lab, University of Bern [1]

Reuben Richardson

*Course Coordinator:* Dr Clarissa Whitmire

June 1, 2025

## Abstract

Brain-computer interfaces (BCIs) that utilize motor imagery (MI) offer intuitive and stimulus-free control schemes but often suffer from variable classification performance due to the subtlety of MI-induced EEG signals. This study investigates the classification performance of EEGNet (a compact convolutional neural network) when trained on motor movement (MM) versus MI EEG data. Two datasets were analyzed: Dataset A, a 64-channel wet-electrode dataset from PhysioNet involving 109 subjects, and Dataset B, an 8-channel dry-electrode dataset self-collected from a single subject using OpenBCI hardware. Preprocessing pipelines were standardized across datasets, and separate EEGNet models were trained for MM and MI conditions. Results showed a statistically significant classification advantage for MM over MI in the 64-channel Dataset A (mean accuracy difference of +3.74%,  $p = 0.0071$ ), whereas the difference was not significant in the 8-channel setting. In Dataset B, a pronounced performance gap was observed, with MM achieving 87.5% accuracy compared to 56.25% for MI. These findings support the hypothesis that motor movement yields more robust and classifiable EEG patterns than motor imagery, especially when using high-density recordings. However, a high standard-deviation (14.25% for motor movement, and 13.94% for motor imagery) in Dataset A models also indicates the viability of MI-based classification even under favorable conditions. The study highlights the importance of electrode configuration, signal fidelity, and individual variability in designing effective EEG-based BCI systems.

**Keywords:** Brain-Computer Interface, Motor Imagery, EEGNet, Electroencephalography, Deep Learning

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Background and Motivation . . . . .	3
1.2	Problem statement . . . . .	4
1.3	Solution approach . . . . .	4
<b>2</b>	<b>Methods</b>	<b>5</b>
2.1	Codebase overview . . . . .	5
2.2	Datasets . . . . .	5
2.2.1	Dataset A: PhysioNet EEG Motor Movement/Imagery Dataset . . . . .	5
2.2.2	Dataset B: Self-Collected Using OpenBCI Hardware . . . . .	6
2.3	Dataset A Preprocessing . . . . .	7
2.4	Dataset B Preprocessing . . . . .	8
2.5	Model - EEGNet . . . . .	8
2.6	Training . . . . .	9
<b>3</b>	<b>Results and Discussion</b>	<b>10</b>
3.1	Dataset A: Classification Performance for Motor Imagery vs. Motor Movement, 64-Channel vs 8-Channel Models . . . . .	10
3.1.1	Discussion . . . . .	12
3.2	Dataset B: Classification Performance for Motor Imagery vs. Motor Movement . . . . .	14
3.2.1	Discussion . . . . .	14
<b>4</b>	<b>Conclusions</b>	<b>17</b>
4.1	Conclusions . . . . .	17
4.1.1	Conclusions . . . . .	17
4.2	Improvements and Future work . . . . .	17
<b>Appendices</b>		<b>21</b>
<b>A</b>	<b>Methods</b>	<b>21</b>
A.1	Preprocessing . . . . .	21
A.1.1	Dataset A Preprocessing . . . . .	21
A.1.2	Dataset B Preprocessing . . . . .	24
A.1.3	Dataset A Training . . . . .	26
A.1.4	Dataset B Training . . . . .	27
<b>B</b>	<b>Results</b>	<b>28</b>
B.1	Dataset A . . . . .	28

# List of Abbreviations

BCI	Brain-Computer Interface
CNN	Convolutional Neural Network
Cyton	OpenBCI Cyton Biosensing Board
DL	Deep Learning
EEG	Electroencephalogram
ELU	Exponential Linear Unit
FIF	Functional Image File (used in MNE for EEG storage)
MI	Motor Imagery
Mel	Mental Imagery
MM	Motor Movement
MNE	Magnetoencephalography/EEG software (Python)
PSD	Power Spectral Density
SSVEPs	Steady-State Visual Evoked Potentials

# Section 1

## Introduction

### 1.1 Background and Motivation

Low-cost electroencephalogram (EEG) based Brain-Computer Interfaces (BCIs) have become gradually more viable in the last decade as electronics (integrated circuits, signal amplifiers, wireless microcontrollers etc.) continue to increase in capability and drop in price. EEG-compatible paradigms such as P300 BCI (based on event related potentials - ERPs) and steady state visual evoked potentials (SSVEPs) have shown high-reliability and perform well across individuals without prior training [2, 3]. However, these techniques function more as a detection of internal brain "reflexes" to external stimuli than as a detection of voluntary brain activity [4]. Additionally, they require the user to continuously focus on a computer screen for the interface to function.

These problems motivate the use of motor imagery (MI), and more generally mental imagery (MeI). These imagery techniques allow for a more intuitive interface for controlling movement-related peripherals such as a computer mouse, or a robotic arm [3, 5].

Given the inherent complexity and variability of EEG signals, deep learning models have gained traction in EEG-based research. A comprehensive systematic review of DL-EEG applications found that 40% of these studies used convolutional neural networks (CNNs) — highlighting CNNs as a dominant and effective architecture for EEG-based tasks [6]. EEGNet, in particular, is a lightweight CNN that has demonstrated high classification accuracy across various paradigms, including motor imagery [7, 6]. Its design allows for simultaneous temporal and spatial feature extraction [8], and is found to perform especially well with limited training datasets, as compared with other CNNs and traditional approaches [7]. As a compact and widely adopted deep learning model tailored for EEG-based BCIs, EEGNet can be chosen to serve as a representative benchmark for evaluating classification performance.

Previous research has shown that although real movement and motor imagery elicit similar patterns of EEG activation — most notably desynchronization in the  $\mu$  and  $\beta$  bands — these effects are not equally consistent across individuals [9]. Findings by Y. Höller Et al. suggest that motor imagery often results in weaker and less consistent EEG signatures than real movement, which may directly impact the classification performance of imagery-based BCI.

Comparing the accuracy of EEGNet on motor imagery versus real movement datasets could therefore help determine the suitability of a specific subject for MI-based BCI, and also provide insight into the broader differences in classification difficulty between these two paradigms.

## 1.2 Problem statement

Previous studies suggest that motor movement-based brain–computer interfaces (BCIs) should yield stronger classification performance than those based on motor imagery. This report aims to quantify this difference in BCI performance by using EEGNet, a compact convolutional neural network, as a benchmark model.

Importantly, this study investigates how the performance gap between motor movement (MM) and motor imagery (MI) varies across different EEG acquisition setups. Specifically, it compares high-density (64-channel), wet-electrode EEG data from the PhysioNet dataset (Dataset A) with low-density (8-channel), dry-electrode EEG data collected using OpenBCI hardware (Dataset B). This dual analysis aims to uncover not only whether MM outperforms MI, but also how this performance gap depends on electrode density and signal quality.

A research hypothesis is proposed based on reasoning outlined in Section 1.1.

**Research Hypothesis:** *EEGNet will show stronger classification performance when trained on a motor movement (MM) dataset than when trained on a motor imagery (MI) dataset, with this performance gap being more pronounced in high-density, low-impedance EEG recordings.*

## 1.3 Solution approach

To evaluate the research hypothesis, this study employed EEGNet (see Section 2.5) to classify EEG signals. Two datasets were analyzed: Dataset A (PhysioNet, 64-channel wet electrode) and Dataset B (self-collected, 8-channel dry electrode using OpenBCI consumer hardware). Each dataset was preprocessed using a consistent pipeline involving filtering, epoching, and normalization (see Sections 2.3 and 2.4). For Dataset A, MM and MI models were trained separately for each subject under both 64-channel and 8-channel configurations. For Dataset B, separate models were trained on MM and MI data from a single subject. Classification accuracy was used as the primary evaluation metric, and performance comparisons were made across task types (MM vs. MI) and electrode configurations (64ch vs. 8ch) to quantify the impact of recording quality on model performance (see Section 3).

# Section 2

## Methods

### 2.1 Codebase overview

The full codebase for this project can be found here: <https://github.com/Dooganar/bioe6100-MI-EEG-classification>. It has the following structure:

```
bioe6100-MI-EEG-classification
  ├── results
  │   ├── results_analysis.py
  │   ├── models-64ch-tasks12-200epoch-test-accuracy.csv
  │   └── models-8ch-tasks12-200epoch-test-accuracy.csv
  └── src
      ├── eegnet.py
      ├── train_eegnet.py
      ├── process_physionet_data.py
      ├── collect_openbci_data.py
      └── process_openbci_data.py
  README.md
```

The `src` folder contains all of the Python scripts associated with collecting and processing EEG data, and training the EEGNet model. The EEGNet CNN architecture is defined by `eegnet.py` using the PyTorch library (adapted from the original EEGNet TensorFlow implementation [7]). The training, and evaluation classes for EEGNet are also included in `eegnet.py`.

`train_eegnet.py` is the main script for training the EEGNet models. This script loads preprocessed data (`.fif` files), trains the models, and handles model evaluation (further detail provided in Section 2.6).

### 2.2 Datasets

#### 2.2.1 Dataset A: PhysioNet EEG Motor Movement/Imagery Dataset

The PhysioNet EEG Motor Movement/Imagery Dataset [10] consists of EEG recordings from 109 participants performing various motor execution and motor imagery tasks. For this project, only **Task 1** (left/right fist movement) and **Task 2** (left/right fist motor imagery) are utilized. Each task involved visual cues prompting subjects to either physically execute or mentally imagine the repetitive opening and closing of the corresponding fist.

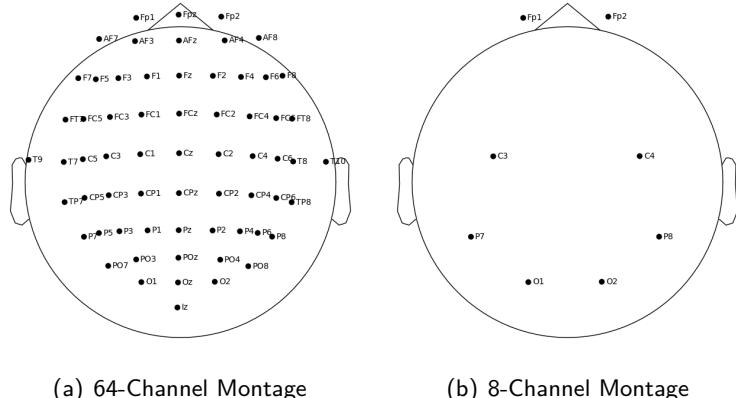


Figure 2.1: Visualisations from the MNE Python package for the Standard 10-10 Electrode Montage. Subfigure (a) depicts all 64 channels present in Dataset A. Subfigure (b) depicts the 8-channel montage of Dataset B, which is a subset of the channels present in (a).

Recordings were conducted using a 64-channel EEG cap with wet electrodes arranged according to the international 10-10 system, and collected at a sampling rate of 160 Hz using the BCI2000 platform. The electrode montage includes the channels Fp1, Fp2, C3, C4, O1, O2, P7, and P8 — which are also present in the OpenBCI Ultracortex headset used in Dataset B (see Figure 2.1).

## 2.2.2 Dataset B: Self-Collected Using OpenBCI Hardware

Dataset B consists of EEG recordings collected from a single subject (the report author) using an OpenBCI Cyton board equipped with the Ultracortex Mark IV headset. The data was collected using an 8-channel dry electrode montage corresponding to the Fp1, Fp2, C3, C4, O1, O2, P7, and P8 positions—the standard configuration for this device (see Figure 2.1).

Data collection was performed using a custom Python script (`collect_data_openbci.py`) that utilizes the BrainFlow API to stream EEG signals from the Cyton board over a serial connection. Prompts were delivered sequentially according to a randomized yet structured order that ensured balanced sampling of each class. Specifically, the script generated sequences of alternating active (“Left Fist” or “Right Fist”) and “Rest” prompts, with interleaved “Nothing” markers for transitions. Each recording session began with a “Nothing” and “Rest” state, and concluded with a final “Rest” prompt.

Two distinct runs were conducted:

- In the **Motor Movement** run, the subject physically opened and closed their left or right fist in response to each active prompt.
  - In the **Motor Imagery** run, the subject mentally imagined performing the same movements without any physical motion.

For each run, 5 prompt batches were completed, where each batch consisted of 5 left and 5 right fist prompts (with associated rest intervals). This resulted in a total of 25 “Left Fist”, 25 “Right Fist”, and 55 “Rest” samples per run. The script saved the raw EEG signals and corresponding prompt markers to `eeg_samples.json` and `eeg_markers.json` files, which were subsequently renamed to distinguish between movement and imagery runs before preprocessing.

The resulting dataset provides two aligned data collections — one for motor execution and one for motor imagery — suitable for evaluating classification accuracy under controlled, reproducible conditions using the same subject and hardware.

## 2.3 Dataset A Preprocessing

The raw EEG data from Dataset A (PhysioNet EEG Motor Movement/Imagery Dataset) was preprocessed using the MNE-Python library. For each subject, recordings from three relevant runs were loaded and concatenated, corresponding to the selected task (Task 1 for motor movement or Task 2 for motor imagery). Additionally, both 64-channel and 8-channel configurations were processed to investigate the impact of reduced spatial resolution on model performance.

The following steps were performed to prepare the data for classification: Run Extraction, Filtering, Channel Selection, Event Extraction and Epoching.

**Filtering:** Each subject's data was bandpass filtered between 0.2 Hz and 40 Hz to remove slow drifts and high-frequency noise, retaining the EEG frequency bands relevant for motor activity analysis.

**Channel Selection:** To produce the processed 8-channel dataset, only the following channels were retained: Fp1, Fp2, C3, C4, O1, O2, P7, and P8. All other channels were dropped to match this configuration. These channels were selected to match the 8-Channel montage of Dataset B to allow a electrode-position-controlled comparison of model accuracy between wet, and dry electrode 8-channel data. No channels were dropped (this step was skipped) when producing the 64-channel configuration.

**Event Extraction and Epoching:** Events were extracted from stimulus annotations using MNE's built-in `events_from_annotations` function. The raw continuous EEG was then segmented into epochs from  $-1$  to  $4$  seconds relative to each event onset. No baseline correction was applied. The resulting epochs were saved in FIF format for downstream training using the EEGNet model. Figure A.5 in the Appendix illustrates the segmented EEG epochs and their corresponding class labels following event extraction.

**Normalisation:** Before training, each channel was z-score normalised ( $X' = \frac{X - \mu}{\sigma}$ ) across all epochs to reduce the effect of arbitrary channel scaling.

There is thus 4 different processed epoched datasets for each subject:

1. Task 1 (Motor Movement), 64 Channels
2. Task 2 (Motor Imagery), 64 Channels



Figure 2.2: Preprocessing Pipeline for Dataset A

3. Task 1 (Motor Movement), 8 Channels
4. Task 2 (Motor Imagery), 8 Channels

## 2.4 Dataset B Preprocessing

The preprocessing of Dataset B was performed using a custom Python script built on top of MNE-Python library. This procedure aims to be identical the preprocessing pipeline used for Dataset A, allowing for a controlled comparison across datasets.

First, the raw EEG signals and corresponding prompt markers were loaded from the `eeg_samples.json` and `eeg_markers.json` files. Each sample consisted of 8 EEG channels corresponding to the selected montage. The BrainFlow Cyton data, originally buffered and recorded in microvolts, was converted into a continuous unbuffered stream in volts to ensure compatibility with MNE-Python.

A synthetic stimulus channel was appended to the data stream to represent prompt transitions, from which discrete events (Rest, Left Fist, Right Fist) were detected using `mne.find_events()`. A standard 10-05 montage was applied to spatially register channel positions, followed by bandpass filtering between 0.2 Hz and 40 Hz—identical to the filtering applied to Dataset A.

Examples of the raw EEG signal and its frequency characteristics before and after bandpass filtering are shown in Appendix Figures A.6–A.9. As seen in Figure A.6 and Figure A.7, the unfiltered data contains strong 50 Hz and 100 Hz components, likely due to power line interference and its harmonic. After applying the 0.2–40 Hz bandpass filter, these noise artifacts are effectively suppressed (Figures A.8 and A.9), resulting in cleaner signals suitable for downstream classification.

The filtered continuous signals were segmented into epochs spanning from -0.5 to 2 seconds relative to event onset. This epoch window was selected to capture both pre- and post-stimulus activity, and to align closely with the temporal dynamics of the prompt presentation structure used during data collection. Each resulting epoch was labeled according to the associated motor class and saved in FIF format for downstream training.

Figure A.10 in the Appendix illustrates the segmented EEG epochs and their corresponding class labels following event extraction.

Normalization was deferred until training, where z-score normalization was applied per channel across all epochs to ensure consistent scale across sessions and datasets.

## 2.5 Model - EEGNet

EEGNet is a compact convolutional neural network architecture specifically designed for EEG-based brain-computer interface (BCI) tasks. Its architecture incorporates well-established neurophysiological principles, such as temporal filtering, spatial filtering, and frequency-band separation, while minimizing the number of trainable parameters. This makes it particularly well-suited for EEG classification problems where limited training data is available.

The EEGNet architecture, originally proposed by Lawhern et al. [7], consists of three primary blocks:

1. **Block 1 (Temporal and Spatial Filtering):** A temporal convolution layer captures frequency features (this could be thought of as a set of band-pass filters), followed by a depthwise convolution that acts as a spatial filter (across channel/electrode locations) for each frequency-specific feature map. Batch normalization and ELU activation are applied, followed by average pooling and dropout.

2. **Block 2 (Separable Convolution):** A separable convolution is employed to temporally summarize each feature map (depthwise) and then recombine them using a pointwise convolution. This structure further reduces parameters and captures cross-feature map interactions.
3. **Classifier Block:** The feature maps are flattened and passed through a fully connected softmax output layer for classification.

The model was adapted to support both 64-channel and reduced 8-channel configurations. The model hyperparameters used are:

- Number of temporal filters ( $F_1$ ): 16
- Depth multiplier ( $D$ ): 2
- Number of separable filters ( $F_2$ ): 32
- Temporal kernel length: 32
- Dropout rate: 0.5
- Pooling kernel sizes: 8 and 16

## 2.6 Training

For Dataset A, a total of 436 EEGNet models were trained and evaluated. This is comprised of 4 models per subject, corresponding to each of the four processed dataset types outlined in Section 2.3. Each subject was found to have roughly 42, 22, and 22 samples (labelled epochs) for the "Rest", "Left-Hand" and "Right-Hand" classes respectively.

For Dataset B, 2 EEGNet models were trained corresponding to the MM and MI datasets. Each dataset contained 55, 25, and 25 samples (labelled epochs) for the "Rest", "Left-Hand" and "Right-Hand" classes respectively.

These class imbalances were addressed by computing weights using scikit-learn's `compute_class_weight` and passing them to a weighted cross-entropy loss function. A 70%-30% train-test split was then performed.

Model training was performed for 200 epochs using the Adam optimizer, and the trained model weights were saved for evaluation. Example training loss, and training accuracy curves are shown in Appendix A.11 for Subject 1 of Dataset A. Dataset B training loss, and training accuracy curves for motor movement, and motor imagery can be found in Appendix A.12 and A.13 respectively. A separate evaluation module calculated classification accuracy and plotted normalized confusion matrices. These figures are shown and discussed in Section 3.

## Section 3

# Results and Discussion

### 3.1 Dataset A: Classification Performance for Motor Imagery vs. Motor Movement, 64-Channel vs 8-Channel Models

In this section, we evaluate the classification performance of the EEGNet model on EEG data from the 109 *Dataset A* subjects, comparing motor imagery and motor movement across two different electrode configurations: **64-channel** and **8-channel**. We assess accuracy trends, statistical significance, and the impact of reducing the electrode count. The raw 3-class accuracy results for each model type on each subject can be found in Appendix B.1.

The difference in classification accuracy (MM - MI) for each subject was calculated from the Appendix B.1 results. Figures 3.1 and 3.2 visualise the distribution of these differences, and a statistics summary is provided in Table 3.1.

Table 3.1: Summary statistics comparing 64 and 8-channel Dataset A EEGNet models.

Metric	64-channel	8-channel
Number of Subjects	109	109
Mean MM Accuracy	76.51%	69.23%
Mean MI Accuracy	72.77%	67.49%
Mean Accuracy Difference (MM-MI)	+3.74%	+1.74%
Standard Deviation of Differences	14.25%	13.94%
t-statistic of Differences	2.74	1.30
p-value of Differences	0.0071	0.1956

Figure 3.1 presents boxplots comparing the classification accuracy of 64-channel and 8-channel models across Task 1 and Task 2. In both tasks, the 64-channel configuration achieves higher median accuracy than the 8-channel configuration. The interquartile ranges are similar across configurations, although the 64-channel model noticeably grows in variability in Task 2 when compared to the Task 1 distribution. A low-performing outlier is identified for the 8-Channel Task 1 model, and for the 64-Channel Task 2 model.

Figure 3.2 presents the distribution of classification accuracy differences between motor movement (MM) and motor imagery (MI) tasks for both 64-channel and 8-channel EEGNet configurations. The histogram shows the accuracy differences (MM - MI) in 20 bins across 109 subjects, overlaid with smoothed kernel density estimation curves for each configuration. While both distributions are centered near zero, the 64-channel configuration exhibits a more positively skewed distribution with a peak slightly shifted to the right, suggesting a larger

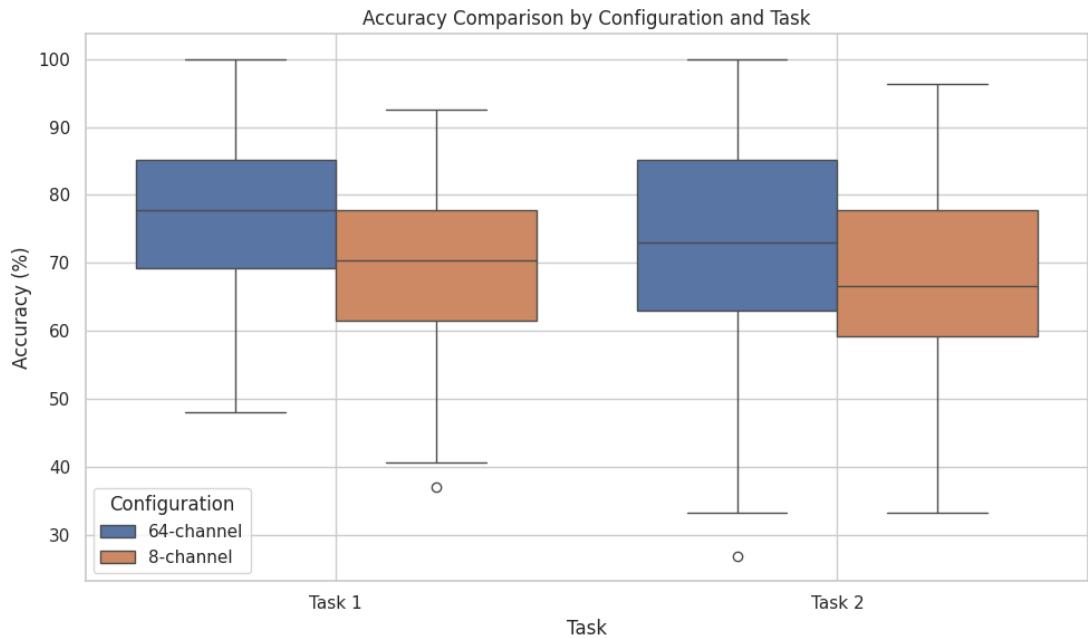


Figure 3.1: Boxplot comparison of classification accuracy by task and electrode configuration.

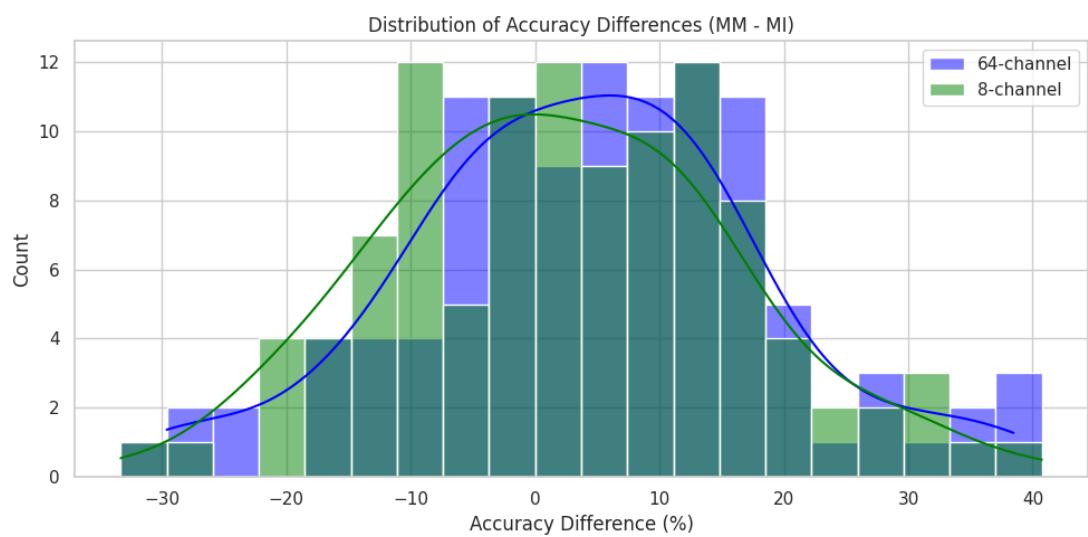


Figure 3.2: 20-bin histogram-plot of the distribution of accuracy differences (motor imagery - motor movement) for 64- and 8-channel configurations.

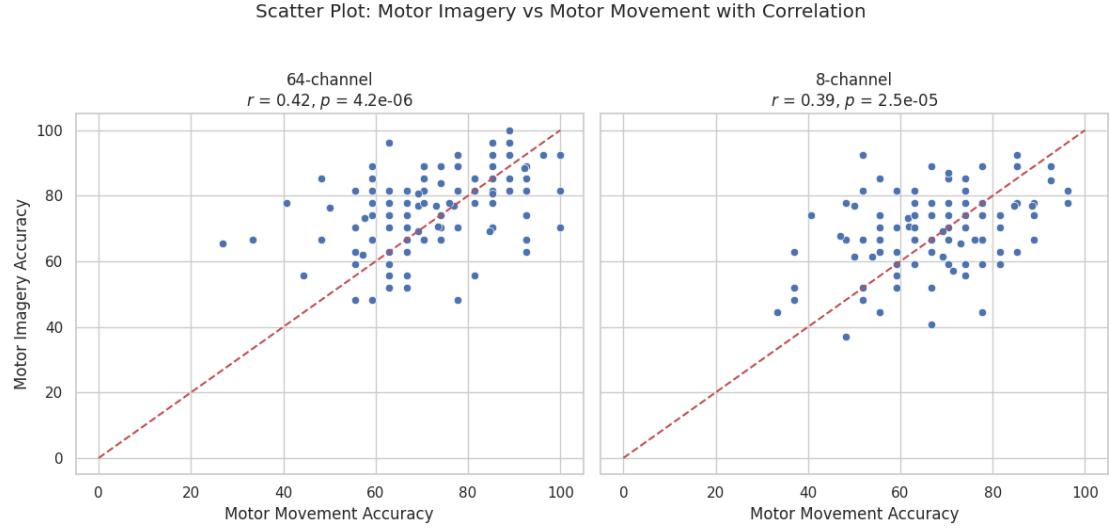


Figure 3.3: Per-subject classification accuracy scatter plots: Motor imagery vs. Motor movement. Red line denotes equality.

performance gain from MM over MI compared to the 8-channel setup.

Table 3.1 summarizes key statistical measures across the two configurations. Both configurations were evaluated on 109 subjects. The 64-channel model achieved a mean MM accuracy of 76.51% and mean MI accuracy of 72.77%, resulting in a mean difference of +3.74% (standard deviation 14.25%). In contrast, the 8-channel model had lower mean accuracies—69.23% for MM and 67.49% for MI—with a smaller mean difference of +1.74% (standard deviation 13.94%). A paired t-test revealed that only the 64-channel model's difference was statistically significant ( $t = 2.74, p = 0.0071$ ), while the 8-channel result was not ( $t = 1.30, p = 0.1956$ ).

Figure 3.3 shows scatter plots comparing motor imagery (MI) and motor movement (MM) classification accuracies for each subject under the two EEGNet configurations: 64-channel (left) and 8-channel (right). Each point represents an individual subject, with MM accuracy on the x-axis and MI accuracy on the y-axis. A reference line ( $y = x$ ) is plotted in red to indicate equal performance between the two tasks.

In the 64-channel plot, the data shows a moderate positive correlation between MM and MI accuracies (Pearson's  $r = 0.42, p = 4.2 \times 10^{-6}$ ). A similar trend is observed in the 8-channel configuration with a slightly lower correlation ( $r = 0.39, p = 2.5 \times 10^{-5}$ ). The presence of statistically significant positive correlations in both configurations suggests that individuals who perform well in MM classification tend to also perform relatively well in MI classification.

### 3.1.1 Discussion

**Impact of Electrode Count** Reducing the number of electrodes from 64 to 8 leads to a general decrease in classification accuracy, particularly for motor imagery. The mean imagery accuracy drops from 72.77% to 67.49%, and the advantage of movement over imagery shrinks by more than half. This highlights a trade-off in BCI system design: fewer electrodes offer practicality, affordability, and faster setup but come at the cost of reduced classification performance. The effect is especially pronounced in tasks relying on more subtle, and potentially less localised neural signals such as motor imagery.

**Subject Variability** These findings suggest high inter-subject variability in the neural signatures of motor imagery. While some individuals exhibit strong classification potential under motor imagery conditions, others show significant performance degradation. This highlights the need for individualized calibration and adaptive models in BCI systems.

**Statistical Significance and Correlation** The 64-channel configuration showed a statistically significant difference in classification accuracy between motor movement (MM) and motor imagery (MI) tasks ( $p = 0.0071$ ), whereas the 8-channel configuration did not ( $p = 0.1956$ ). This suggests that higher spatial resolution may be crucial in capturing the fine-grained differences needed to distinguish between MM and MI EEG signals. Additionally, the positive Pearson correlation between MM and MI accuracies ( $r = 0.42$  for 64ch,  $r = 0.39$  for 8ch) suggests that subjects who are good at one task are more likely to perform relatively well in the other, albeit with some variance.

**Interpreting the Accuracy Gap** Although motor movement tends to yield higher classification accuracy, the relatively modest performance gap (mean difference of +3.74% for 64ch and +1.74% for 8ch) indicates that MI-based classification is still feasible, especially when high-density EEG recordings are available. This supports the viability of motor imagery paradigms for BCI applications, provided that model tuning and signal quality are sufficient.

## Limitations

Several limitations must be acknowledged in light of the Dataset A model results:

- **Single Model Architecture:** EEGNet was used as the sole benchmark classifier. While it is a strong performer and widely adopted, other architectures—such as RNNs or hybrid models—may yield different accuracy profiles, particularly for motor imagery.
- **Fixed Training Parameters:** All models were trained with the same number of epochs (200) and default hyperparameters. This may not represent the optimal configuration for all subjects, especially given the variability in EEG signal quality and inter-subject performance.
- **Test-Set Size and Label Balance:** Each model's test set consisted of only 27 labelled records (30% of total records). A single test classification changes the overall model accuracy metric by 3.7%. This results in significant instability in a given model's accuracy metric. Although class weighting was applied, the inherent imbalance between "Rest", "Left", and "Right" classes may still bias the learning process.
- **Electrode Downsampling Approach:** The 8-channel configuration was derived by selecting a subset of electrodes rather than using an optimized electrode selection algorithm. This may not reflect the best possible 8-electrode layout for BCI use.
- **Task Design Constraints:** The analysis was restricted to only Tasks 1 and 2 of the PhysioNet dataset. Other tasks (e.g., bilateral movements) or additional baselines (e.g., eyes open/closed) may offer different insights.
- **Real-World Applicability:** While the data was collected in controlled conditions, real-world EEG-BCI use cases involve motion artifacts, electrode drift, and attention variability—all of which were not assessed in this study.

## 3.2 Dataset B: Classification Performance for Motor Imagery vs. Motor Movement

Table 3.2 presents summary statistics for EEGNet model performance on Dataset B, which contains motor movement (MM) and motor imagery (MI) EEG recordings collected from a single subject using OpenBCI hardware. The MM model achieved a classification accuracy of 87.50%, while the MI model reached a notably lower accuracy of 56.25%, suggesting a potential disparity in signal clarity or separability between the two paradigms for this subject.

Figure 3.4 displays the normalized confusion matrices for both the motor movement (Figure 3.4a) and motor imagery (Figure 3.4b) models. In the MM confusion matrix, the model demonstrates perfect classification of the "rest" class (1.00) and strong performance on both movement classes, with 75% accuracy for 'left-fist' and 71% for 'right-fist'. In contrast, the MI model shows more diffuse predictions. While the "right-fist" class retains relatively high classification accuracy (71%), the "rest" class drops to 53%, and "left-fist" is correctly classified only 50% of the time, with considerable confusion among all three classes.

Table 3.2: Summary Statistics For Dataset B Models

		Metric	Value
		Number of Subjects	1
		MM Model Accuracy	87.50%
		MI Model Accuracy	56.25%

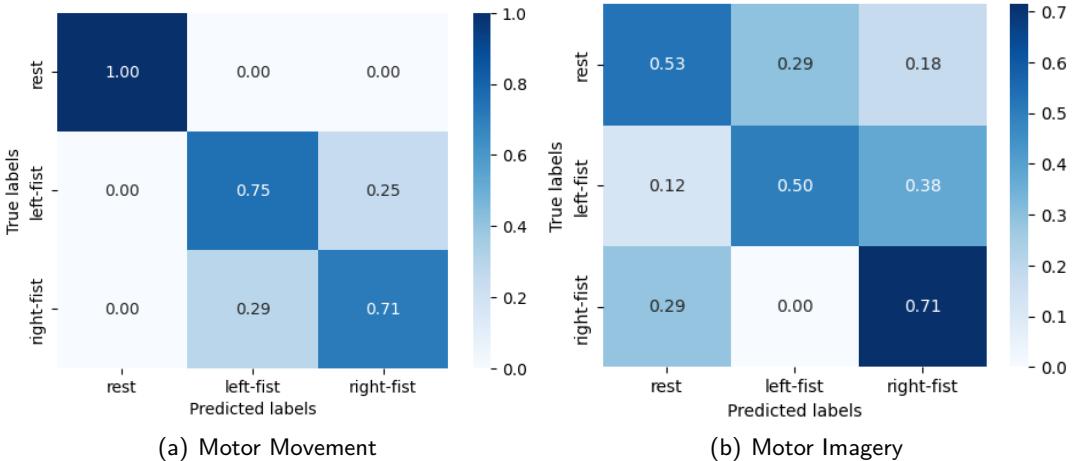


Figure 3.4: Normalised Confusion Matrices for Dataset B Models

### 3.2.1 Discussion

The results from Dataset B highlight a stark contrast in classification performance between motor movement (MM) and motor imagery (MI) tasks for a single subject using an 8-channel OpenBCI system. The MM model achieved high accuracy (87.50%), indicating that the EEGNet architecture was able to reliably detect and differentiate neural signatures associated with actual fist movements. The confusion matrix for MM further supports this, showing near-perfect classification for the "rest" class and strong performance for both "left-fist" and "right-fist" actions.

In contrast, the MI model achieved a much lower accuracy (56.25%), with substantial confusion between the three classes. Most notably, "left-fist" imagery was often misclassified as either "rest" or "right-fist", and "rest" states were frequently confused with active motor imagery. This aligns with existing literature suggesting that motor imagery generates weaker and more variable EEG signals than real movement, especially in untrained or novice users.

These findings reinforce the hypothesis that motor execution tasks produce more distinct and classifiable EEG features than motor imagery tasks. The performance gap observed here (+31.25%) is much larger than the average gap observed in the larger Dataset A (which showed a +3.74% advantage for MM using 64-channel data), possibly due to reduced spatial resolution, dry electrode limitations, and subject-specific variability. The results suggest that while consumer-grade EEG devices may support robust classification of motor movements, additional training, signal enhancement, or algorithmic tuning may be necessary to achieve reliable imagery-based classification.

## Limitations

Several limitations must be considered when evaluating the results derived from Dataset B:

- **Single Subject and Session:** The dataset was collected from only one subject (the report author) during a single session. The subject had minimal prior training in motor imagery tasks, likely contributing to weak or inconsistent MI-related neural patterns. This restricts the generalizability of the findings and makes it difficult to disentangle subject-specific factors from broader trends.
- **Limited Dataset Size:** With only 105 labeled records, of which 32 were used for testing, the classification results are sensitive to specific data points. Each test sample contributes over 3% to the overall accuracy metric, introducing significant volatility and limiting confidence in performance conclusions.
- **Dry Electrode Limitations:** The OpenBCI Ultracortex uses dry electrodes, which are more prone to signal degradation due to poor skin contact, motion artifacts, and higher impedance. These factors may disproportionately impair motor imagery signals, which are already subtle and less distinct than those from motor execution.
- **Absence of Artifact Rejection:** No explicit artifact rejection (e.g., independent component analysis or bad trial exclusion) was applied during preprocessing. Consequently, noisy segments or artifacts (e.g., eye blinks, muscle tension) may remain in the data and negatively influence classification accuracy.
- **Epoch Window and Timing Sensitivity:** The fixed epoch window (-0.5 to 2.0 s) assumes consistent cognitive response timing across trials. However, motor imagery onset is often more variable than motor movement, and fixed windows may fail to capture peak neural activity.
- **No Cross-Validation:** The evaluation was based on a single train-test split. Without k-fold cross-validation or repeated trials, the performance metric may reflect dataset partitioning rather than robust generalization.
- **Hardware and Environmental Noise:** The use of a consumer-grade, wireless EEG system introduces potential latency and environmental noise. Factors such as wireless transmission delay, ambient electrical interference, and unshielded surroundings could degrade signal fidelity and event alignment.

These limitations suggest that while the results are promising for movement-based BCI using OpenBCI, further validation with larger, multi-subject datasets and more rigorous pre-processing is necessary for motor imagery applications.

# Section 4

## Conclusions

### 4.1 Conclusions

#### 4.1.1 Conclusions

This report investigated the classification performance of EEGNet on motor movement (MM) and motor imagery (MI) tasks using both high-density (64-channel) wet-electrode EEG data and low-density (8-channel) dry-electrode EEG data. Two datasets were utilized: Dataset A, derived from the PhysioNet database with recordings from 109 participants, and Dataset B, a self-collected dataset using OpenBCI hardware from a single subject.

The results provide quantitative evidence supporting the hypothesis that motor movement yields stronger and more consistent EEGNet classification accuracy than motor imagery. For Dataset A, a statistically significant advantage in MM classification was observed using 64-channel data, with an average performance gain of +3.74% over MI ( $p = 0.0071$ ). However, this difference was not significant in the 8-channel condition ( $p = 0.1956$ ), suggesting that spatial resolution plays a critical role in distinguishing between MM and MI EEG features. Dataset B further emphasized this disparity, with a 31.25% performance gap between MM and MI tasks for the single subject using dry electrodes.

These findings highlight the impact of electrode configuration, signal quality, and subject variability on BCI performance. While MI-based BCIs remain a viable avenue, particularly with high-density EEG setups, their success is contingent on subject proficiency and hardware capability. Consumer-grade, low-density systems may suffice for MM-based applications but face limitations when applied to MI classification.

The study also faced several limitations, including the use of a single model architecture (EEGNet), fixed training parameters, limited test set sizes, and the absence of advanced artifact rejection techniques. These constraints affect the generalizability and stability of the reported accuracy metrics.

In summary, EEGNet served as a benchmark for evaluating EEG classification across paradigms. The comparative analysis supports the notion that MM produces more classifiable EEG patterns than MI, particularly when signal fidelity is high. This insight can guide future BCI system designs, emphasizing the importance of task selection, electrode configuration, and individualized calibration.

### 4.2 Improvements and Future work

Several avenues remain open for improving experimental design and expanding the scope of analysis:

- **Use Different Available Task Variants:** Future study could incorporate Task 3 (bilateral hand or foot movement) and Task 4 (bilateral motor imagery) from the PhysioNet dataset. These paradigms may produce more distinctive EEG patterns due to increased motor cortex activation, potentially leading to improved classification performance or new insights into task complexity and separability.
- **Eye Movement Correction:** Baseline recordings from the PhysioNet dataset (“eyes open” and “eyes closed”) could be leveraged to model and subtract ocular artifacts during preprocessing. A similar approach could be applied to Dataset B using dedicated calibration runs. Incorporating eye movement correction may enhance signal quality, particularly for subtle MI-related signals.
- **Test MM Trained Models on MI Datasets:** An immediate direction is to investigate how well a model trained on motor movement generalizes to motor imagery test data. This could help evaluate the transferability of motor-related features and inform strategies for subject calibration in MI-based BCIs.
- **Improved Cross-Validation:** Given the limited size of the datasets, future work should adopt k-fold cross-validation or repeated random sub-sampling to produce more stable and generalizable performance metrics. These methods would mitigate the volatility introduced by small test sets and offer a better estimate of model reliability, but come at the cost of significantly increased compute time/energy requirements.
- **Multi-Subject OpenBCI Collection:** Dataset B currently includes data from only a single subject. Expanding this dataset to include multiple participants would allow for the analysis of inter-subject variability using dry electrode consumer-grade systems, and improve the generalizability of findings related to MI-based classification performance in low-cost consumer BCI applications.

Collectively, these improvements could enhance both the reliability of classification results and the real-world applicability of EEGNet-based BCI systems.

# Bibliography

- [1] A. Tzovara, F. Aellen, C. Mignardot, and Riccardo Cusinato Smiley face, "Cognitive Computational Neuroscience." [Online]. Available: <https://aath0.github.io/>
- [2] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, "P300 brain computer interface: current challenges and emerging trends," *Frontiers in Neuroengineering*, vol. 5, p. 14, Jul. 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3398470/>
- [3] S. Park, J. Ha, D.-H. Kim, and L. Kim, "Improving Motor Imagery-Based Brain-Computer Interface Performance Based on Sensory Stimulation Training: An Approach Focused on Poorly Performing Users," *Frontiers in Neuroscience*, vol. 15, Nov. 2021, publisher: Frontiers. [Online]. Available: <https://www.frontiersin.orghttps://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2021.732545/full>
- [4] J. Frey, "OpenBCI crossing swords with motor imagery," Mar. 2015, section: Uncategorized. [Online]. Available: <https://openbci.com/community/openbci-crossing-swords-with-motor-imagery/>
- [5] C. Gonzalez Visiedo, "Deep Learning EEG-based Motor Imagery System for Robot Control using 3D Printed Headset and Electrodes," 2023, accepted: 2023-06-08T08:33:52Z. [Online]. Available: <http://www.theseus.fi/handle/10024/803563>
- [6] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, p. 051001, Aug. 2019.
- [7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," May 2018, arXiv:1611.08024 [cs]. [Online]. Available: <http://arxiv.org/abs/1611.08024>
- [8] S. Lee, S. Jang, and S. C. Jun, "Exploring the Ability to Classify Visual Perception and Visual Imagery EEG Data: Toward an Intuitive BCI System," *Electronics*, vol. 11, no. 17, p. 2706, Jan. 2022, number: 17 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/2079-9292/11/17/2706>
- [9] Y. Höller, J. Bergmann, M. Kronsbichler, J. S. Crone, E. V. Schmid, A. Thomschewski, K. Butz, V. Schütze, P. Höller, and E. Trinka, "Real movement vs. motor imagery in healthy subjects," *International Journal of Psychophysiology*, vol. 87, no. 1, pp. 35–41, Jan. 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167876012006319>

- [10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
- [11] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "EEG Motor Movement/Imagery Dataset," 2009. [Online]. Available: <https://physionet.org/content/eegmmidb/>

# Appendix A

## Methods

### A.1 Preprocessing

#### A.1.1 Dataset A Preprocessing

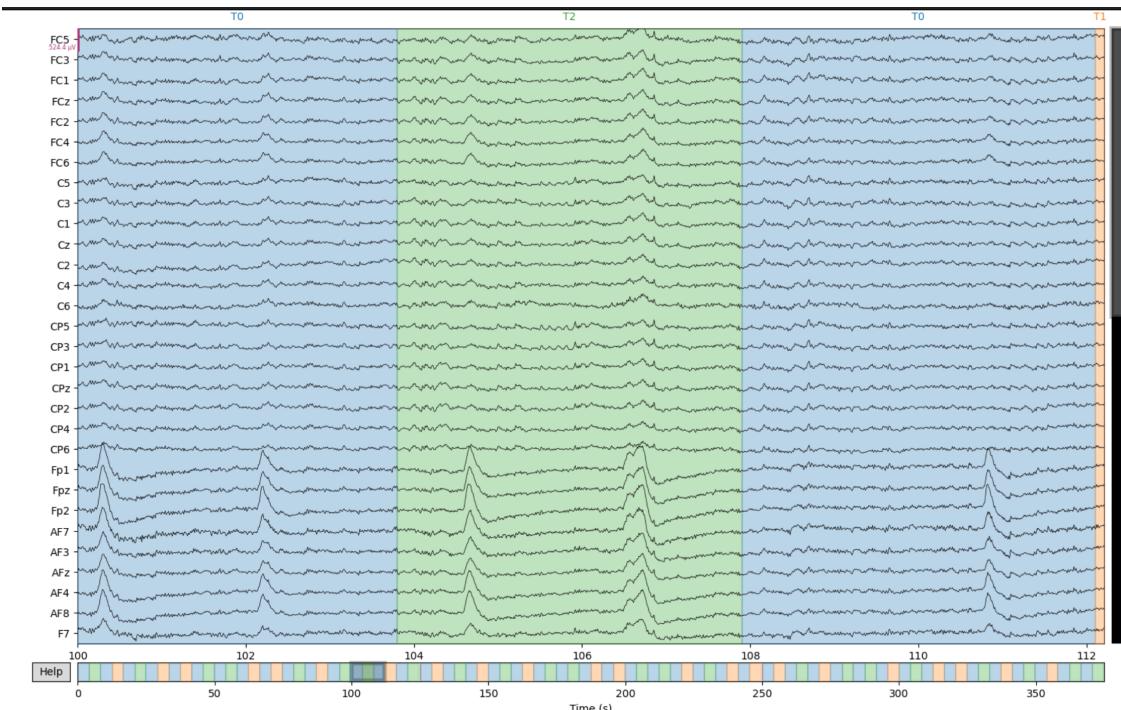


Figure A.1: Dataset A raw EEG signal before bandpass filtering (Using example data from Subject 3, Motor Movement).

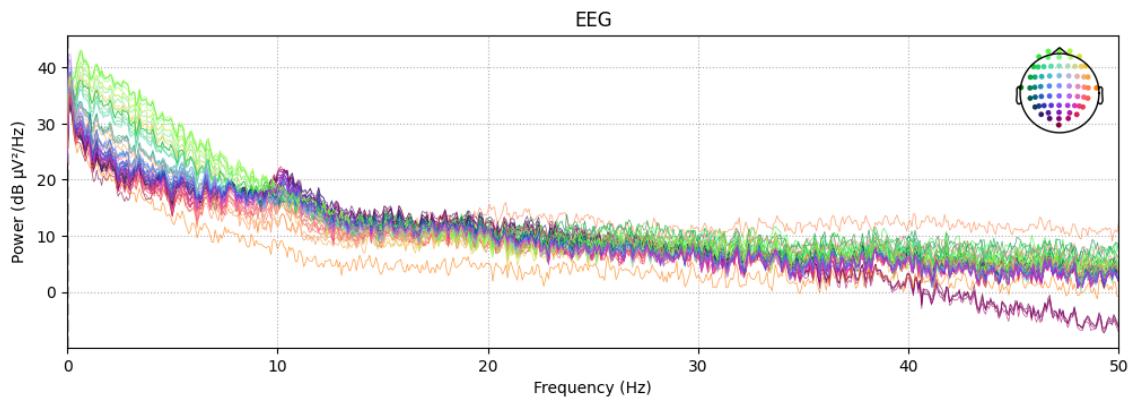


Figure A.2: Power spectral density (PSD) before filtering for Dataset A (Using example data from Subject 3, Motor Movement).

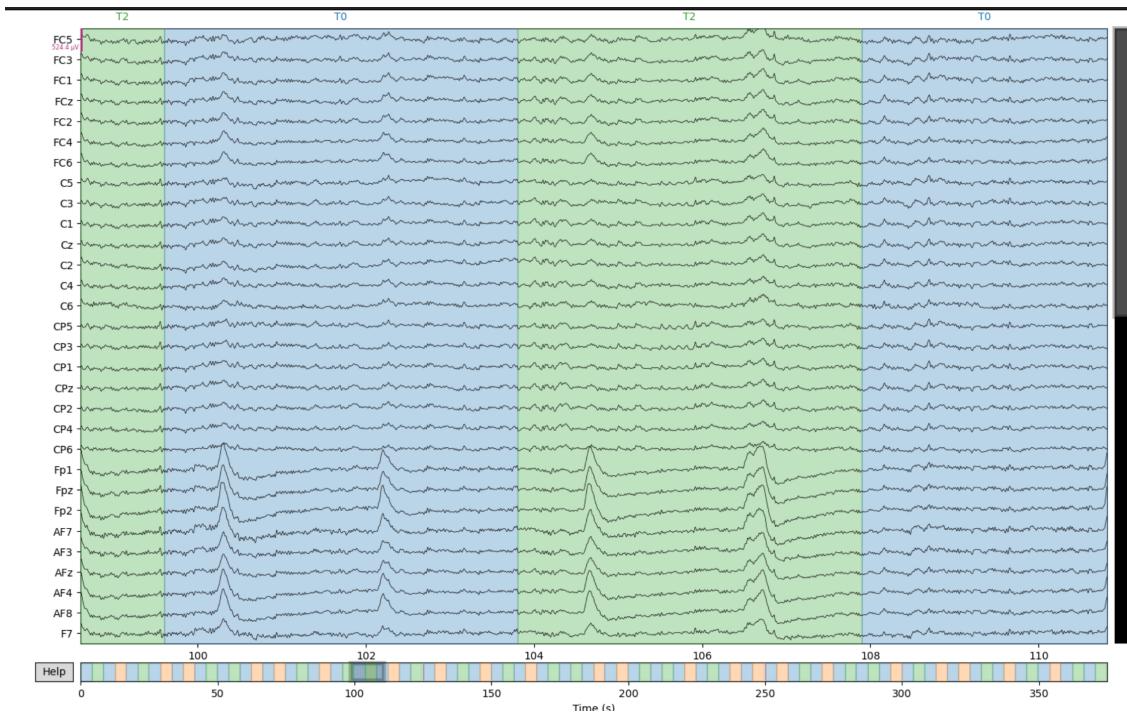


Figure A.3: Dataset A example of raw EEG signal after 0.2-40Hz bandpass filtering (Using example data from Subject 3, Motor Movement).

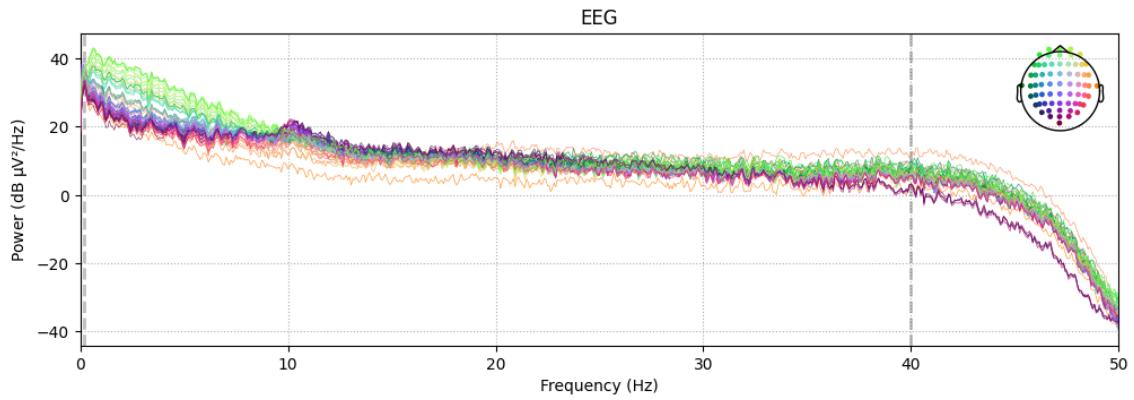


Figure A.4: Power spectral density (PSD) after applying the 0.2-40 Hz bandpass filter (Using example data from Subject 3, Motor Movement).

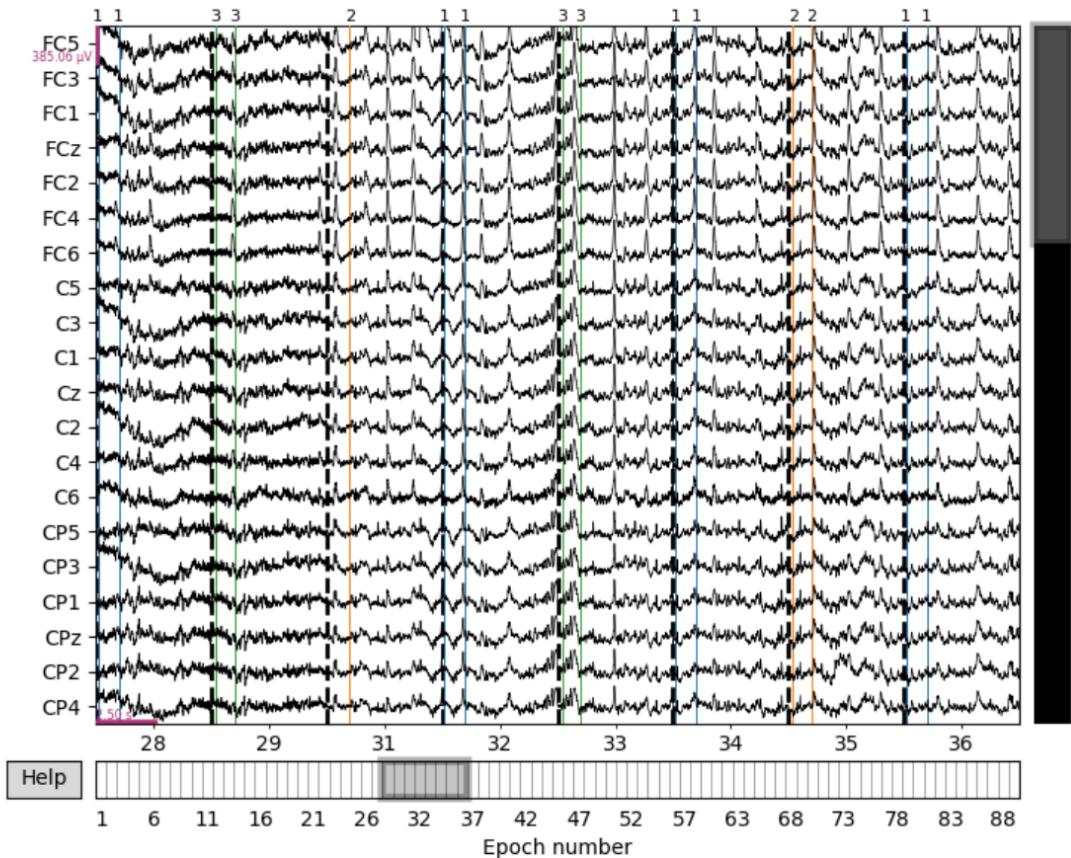


Figure A.5: Segmented and labeled EEG epochs for Dataset A following preprocessing. Each vertical dashed line denotes an event onset, and the corresponding class label (1 = Rest, 2 = Left Fist, 3 = Right Fist) is shown above.

### A.1.2 Dataset B Preprocessing

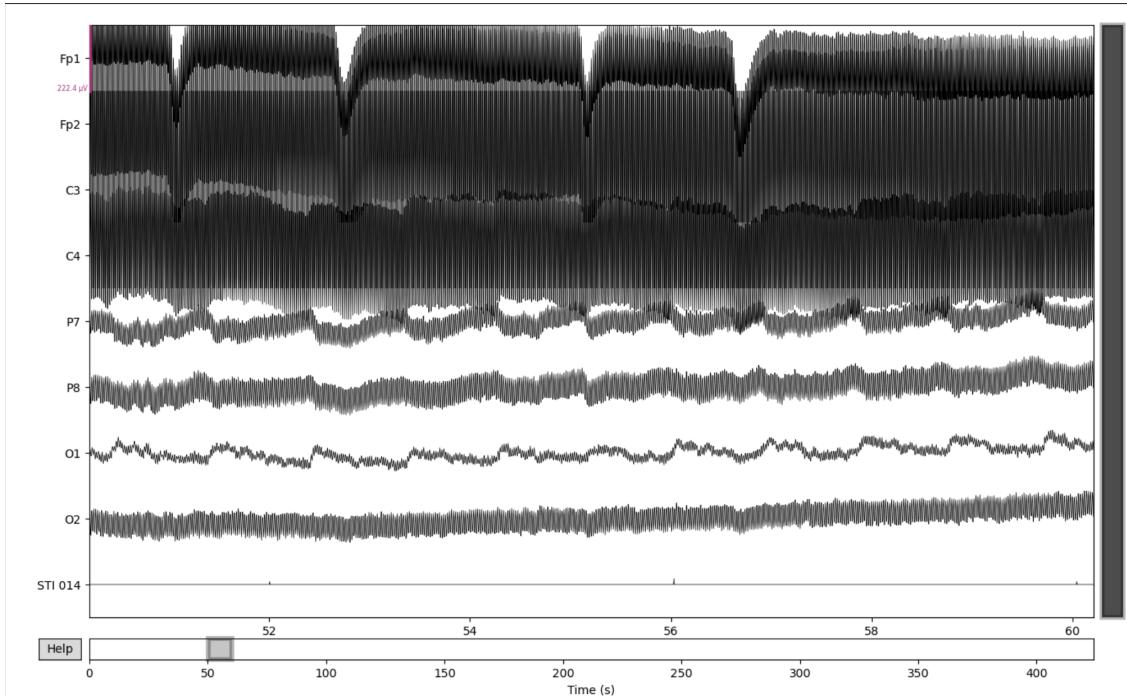


Figure A.6: Raw EEG signal before bandpass filtering (0.2–40 Hz). Prominent 50 Hz and 100 Hz noise components are visible, particularly in frontal channels (Fp1 and Fp2), likely due to power line interference and its harmonic.

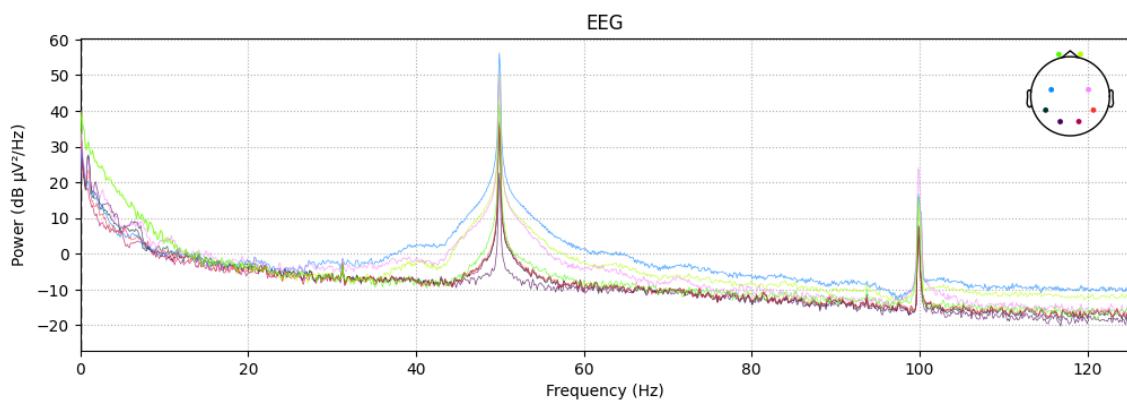


Figure A.7: Power spectral density (PSD) before filtering. Peaks are clearly observed around 50 Hz and 100 Hz, indicating strong power line interference and its harmonic, which contaminate the signal spectrum.

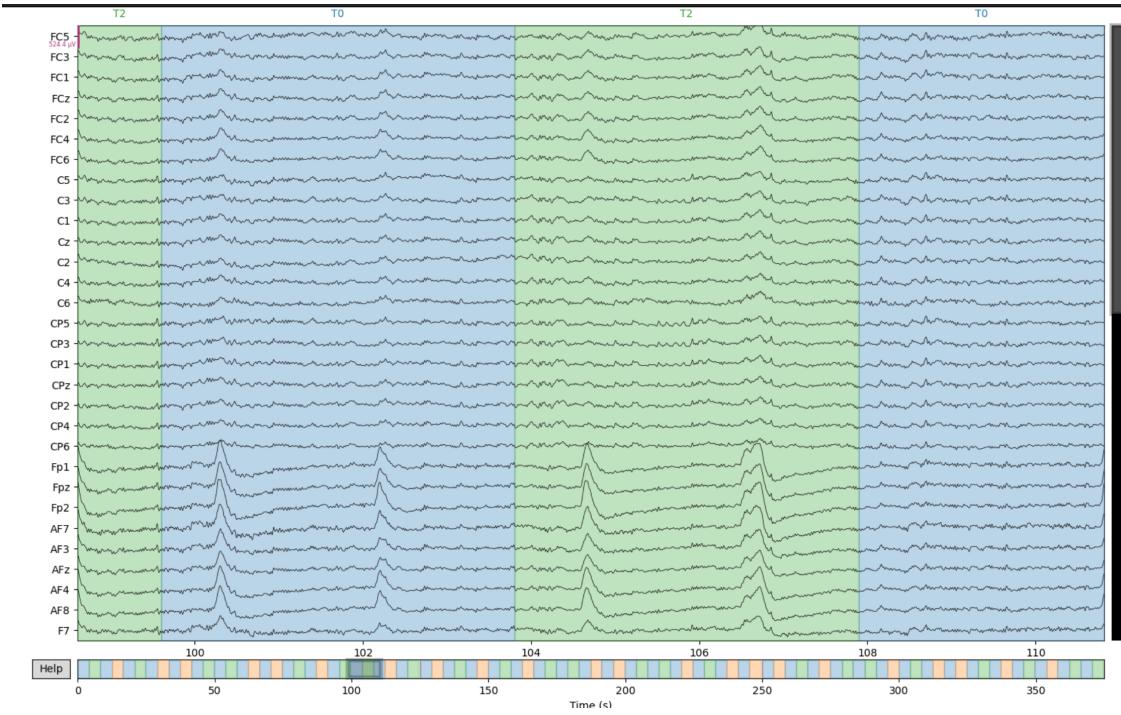


Figure A.8: Raw EEG signal after applying the 0.2–40 Hz bandpass filter. The periodic noise components at 50 Hz and 100 Hz are effectively attenuated, resulting in cleaner signals across all channels.

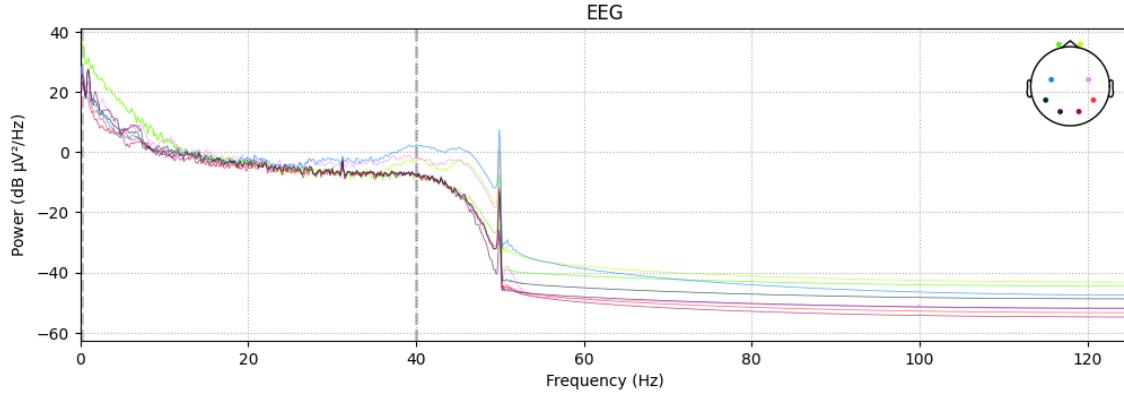


Figure A.9: Power spectral density (PSD) after applying the 0.2-40 Hz bandpass filter. The high-frequency components have been successfully removed, and the signal power is now concentrated in physiologically relevant bands.

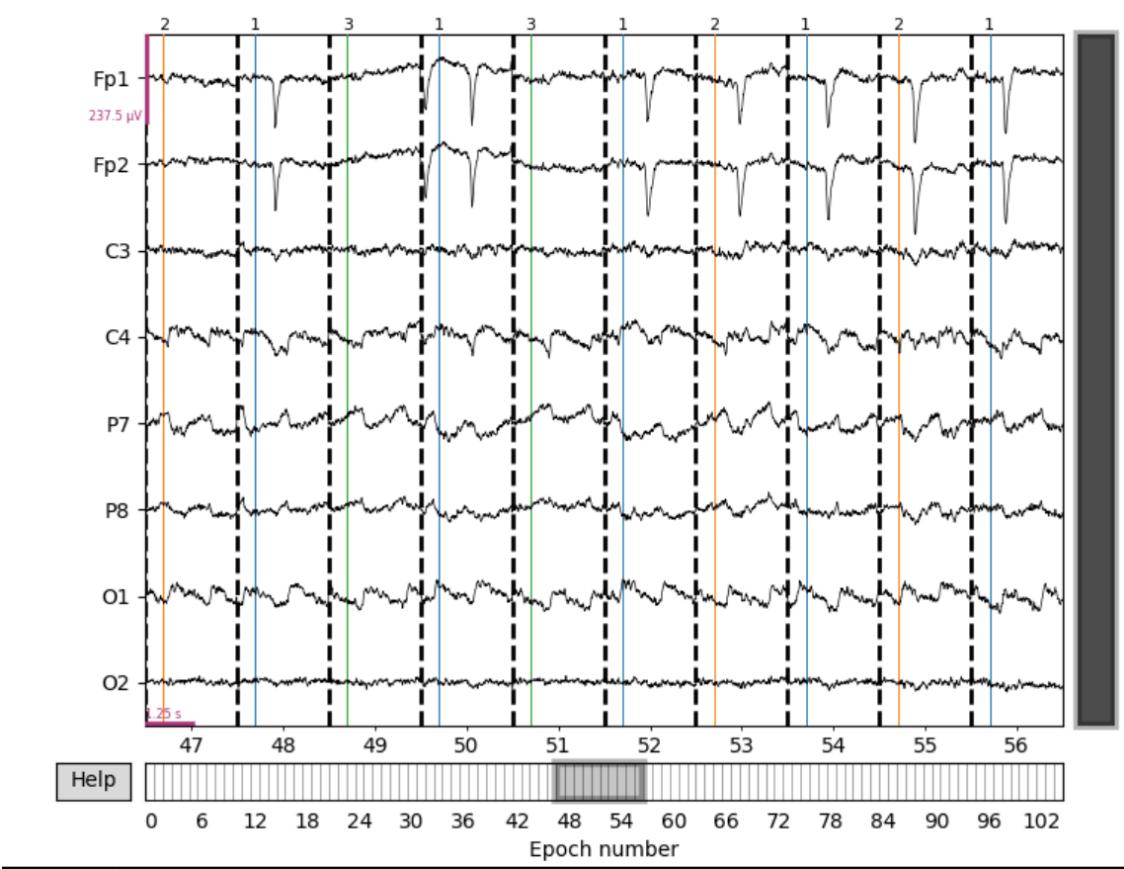


Figure A.10: Segmented and labeled EEG epochs for Dataset B following preprocessing. Each vertical dashed line denotes an event onset, and the corresponding class label (1 = Rest, 2 = Left Fist, 3 = Right Fist) is shown above.

### A.1.3 Dataset A Training

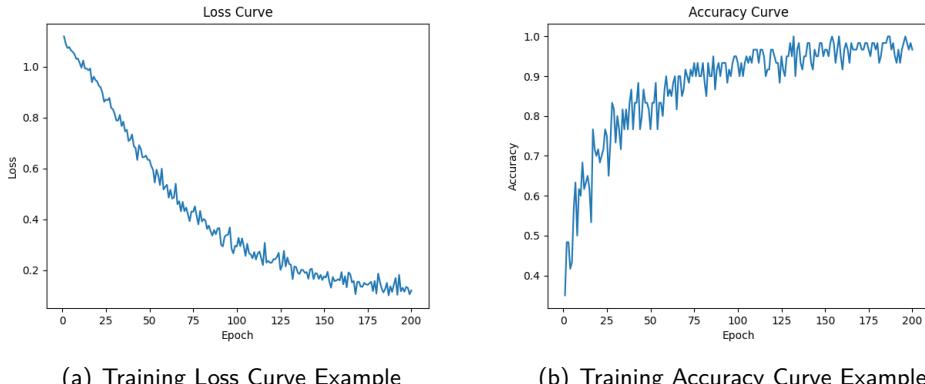


Figure A.11: Loss and accuracy curves produced when training EEGNet on Dataset A (Subject 1, 8-channel MM training shown as an example)

#### A.1.4 Dataset B Training

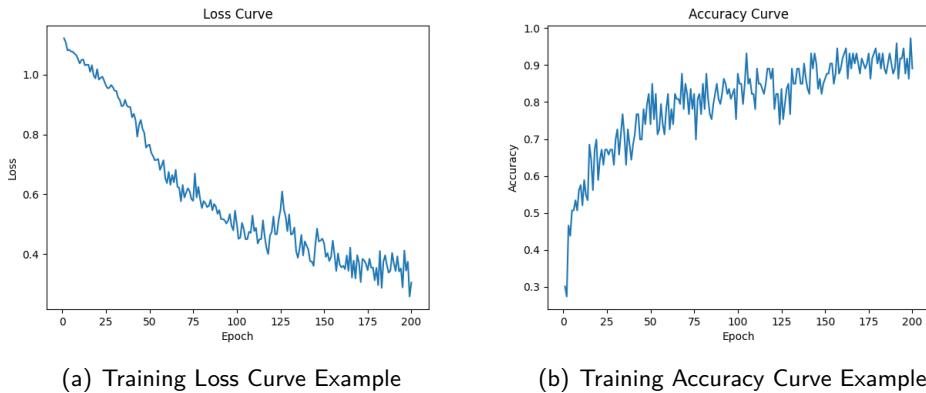


Figure A.12: Loss and accuracy curves produced when training EEGNet on Dataset B, Motor Movement

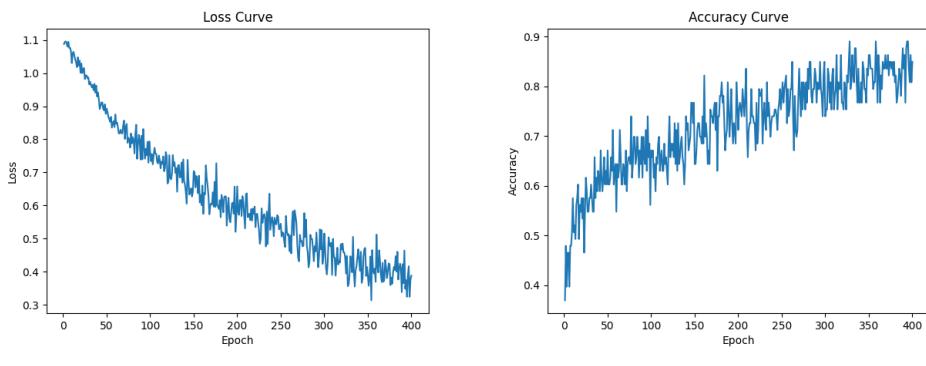


Figure A.13: Loss and accuracy curves produced when training EEGNet on Dataset B, Motor Imagery

## Appendix B

# Results

### B.1 Dataset A

Table B.1: EEGNet classification accuracy comparison for 64-channel and 8-channel tasks across the 109 subjects of the PhysioNet dataset [11]

Subject	Task 1 (64ch)	Task 2 (64ch)	Task 1 (8ch)	Task 2 (8ch)
1	74.07	92.59	59.26	81.48
2	74.07	62.96	59.26	59.26
3	81.48	85.19	81.48	62.96
4	70.37	74.07	74.07	77.78
5	74.07	62.96	62.96	81.48
6	88.89	92.59	66.67	66.67
7	92.59	100.00	88.89	92.59
8	81.48	59.26	74.07	66.67
9	70.37	66.67	59.26	74.07
10	66.67	48.15	44.44	33.33
11	59.26	55.56	66.67	62.96
12	85.19	81.48	74.07	77.78
13	92.59	85.19	66.67	88.89
14	88.89	74.07	77.78	48.15
15	70.37	100.00	77.78	96.30
16	55.56	62.96	40.74	66.67
17	70.37	85.19	66.67	77.78
18	81.48	100.00	81.48	96.30
19	81.48	70.37	66.67	62.96
20	77.78	85.19	59.26	81.48
21	81.48	66.67	74.07	77.78
22	85.19	85.19	77.78	85.19
23	59.26	62.96	66.67	51.85
24	74.07	74.07	88.89	77.78
25	81.48	62.96	77.78	66.67
26	48.15	59.26	37.04	48.15
27	88.89	59.26	85.19	55.56

Continued on next page

Table B.1: EEGNet classification accuracy comparison for 64-channel and 8-channel tasks across the 109 subjects of the PhysioNet dataset [11] (continued).

Subject	Task 1 (64ch)	Task 2 (64ch)	Task 1 (8ch)	Task 2 (8ch)
28	85.19	92.59	48.15	51.85
29	85.19	70.37	85.19	70.37
30	81.48	88.89	66.67	77.78
31	70.37	85.19	66.67	62.96
32	96.30	85.19	85.19	70.37
33	85.19	85.19	85.19	74.07
34	69.23	69.23	65.38	73.08
35	74.07	92.59	74.07	66.67
36	81.48	88.89	77.78	77.78
37	73.08	57.69	61.54	53.85
38	48.15	55.56	44.44	55.56
39	92.59	77.78	77.78	70.37
40	81.48	77.78	66.67	55.56
41	88.46	92.31	61.54	69.23
42	88.89	77.78	66.67	62.96
43	66.67	70.37	59.26	59.26
44	51.85	66.67	59.26	70.37
45	77.78	62.96	70.37	74.07
46	96.30	88.89	88.89	85.19
47	81.48	70.37	77.78	66.67
48	81.48	92.59	74.07	81.48
49	81.48	77.78	74.07	70.37
50	66.67	92.59	74.07	88.89
51	80.77	85.19	84.62	92.59
52	70.37	77.78	70.37	70.37
53	74.07	59.26	59.26	62.96
54	88.89	70.37	85.19	70.37
55	62.96	92.59	44.44	77.78
56	81.48	81.48	70.37	81.48
57	100.00	88.89	92.59	51.85
58	88.89	77.78	88.89	66.67
59	70.37	66.67	81.48	51.85
60	85.19	59.26	66.67	48.15
61	85.19	92.59	51.85	59.26
62	74.07	66.67	81.48	59.26
63	55.56	81.48	59.26	77.78
64	76.92	69.23	73.08	61.54
65	92.59	77.78	77.78	74.07
66	70.37	62.96	70.37	59.26
67	66.67	66.67	51.85	66.67
68	96.30	62.96	62.96	59.26
69	88.89	85.19	70.37	55.56
70	92.59	88.89	77.78	88.89

Continued on next page

Table B.1: EEGNet classification accuracy comparison for 64-channel and 8-channel tasks across the 109 subjects of the PhysioNet dataset [11] (continued).

Subject	Task 1 (64ch)	Task 2 (64ch)	Task 1 (8ch)	Task 2 (8ch)
71	77.78	70.37	74.07	74.07
72	69.23	84.62	76.92	88.46
73	76.92	76.92	76.92	50.00
74	65.38	26.92	69.23	69.23
75	51.85	62.96	48.15	37.04
76	80.77	69.23	61.54	50.00
77	66.67	59.26	74.07	62.96
78	77.78	40.74	74.07	40.74
79	66.67	33.33	51.85	51.85
80	66.67	59.26	55.56	59.26
81	62.96	62.96	66.67	77.78
82	66.67	70.37	62.96	66.67
83	81.48	66.67	77.78	62.96
84	77.78	66.67	66.67	70.37
85	77.78	81.48	74.07	77.78
86	85.19	88.89	77.78	77.78
87	81.48	77.78	74.07	62.96
88	76.47	50.00	67.65	47.06
89	83.87	74.07	87.10	70.37
90	92.59	96.30	92.59	85.19
91	66.67	74.07	74.07	55.56
92	70.59	73.53	70.59	61.76
93	85.19	48.15	62.96	55.56
94	88.89	77.78	77.78	77.78
95	70.37	55.56	74.07	66.67
96	85.19	81.48	59.26	59.26
97	62.96	66.67	62.96	85.19
98	88.89	70.37	77.78	74.07
99	62.96	55.56	55.56	59.26
100	61.90	57.14	57.14	71.43
101	96.30	88.89	81.48	74.07
102	76.92	73.08	76.92	84.62
103	55.56	66.67	51.85	59.26
104	77.78	76.00	66.67	76.00
105	48.15	77.78	70.37	62.96
106	77.78	59.26	59.26	59.26
107	55.56	44.44	62.96	37.04
108	81.48	55.56	55.56	74.07
109	77.78	62.96	51.85	37.04
<b>Average</b>	76.51	72.77	69.23	67.49