A Comprehensive Methodological Documentation of EEG Signal Processing and Classification for Brain-Computer Interface Applications in the MTC-AIC3 Competition

Authors: MIA AI Team

Abstract:

This paper outlines the methodological framework developed for the MTC-AIC3 Brain-Computer Interface (BCI) competition. Our work details a structured process from foundational EEG and BCI knowledge acquisition, through a rigorous multi-stage data preprocessing pipeline, to comprehensive exploratory data analysis. We present an iterative approach to developing classification models for both Steady-State Visually Evoked Potential (SSVEP) and Motor Imagery (MI) tasks, spanning classical machine learning to deep learning architectures. For SSVEP, our final submission utilized a voting classifier, relying on effective signal segments from 3 to 6 seconds. For MI, we describe our progression through models like MLP, EEGNet, and FBCNet, and the subsequent development of a simpler EEGNet version to overcome challenges, utilizing signal segments from 2 to 4 seconds. Our methodology highlights critical steps for effective EEG signal classification in competitive BCI environments, including identifying and resolving data anomalies, optimizing preprocessing parameters, and detailing the implementation of key processing and classification pipelines.

1. Introduction

1.1. Competition Overview: AIC-3 - Egypt National Artificial Intelligence Competition

The Artificial Intelligence Competition (AIC-3) stands as a premier national artificial intelligence competition in Egypt. It is collaboratively organized by the esteemed Military Technical College and the Applied Innovation Center (AIC) of the Ministry of Communications and Information Technology. This competition serves as a significant platform for fostering innovation and technical expertise in the field of AI within the national academic and research landscape.

1.2. Brain-Computer Interfaces (BCIs) in AIC-3

The core focus of AIC-3 revolves around Brain-Computer Interfaces (BCIs), which represent a transformative communication pathway enabling direct interaction between the human brain and external devices. This revolutionary technology holds immense promise for applications ranging from assistive devices for individuals with motor disabilities to advanced human-computer interaction systems. The competition specifically highlights two of the most widely studied paradigms in non-invasive BCI systems:

- **Steady-State Visual Evoked Potentials (SSVEP):** This paradigm leverages the brain's consistent oscillatory response when presented with visual stimuli flickering at specific, distinct frequencies. The classification task involves identifying the specific frequency of the visual stimulus based on the evoked EEG response.
- **Motor Imagery (MI):** Motor Imagery involves decoding internally generated neural patterns associated with imagined movements, without any overt physical action. The challenge lies in classifying these subtle neural signatures into distinct motor imagery categories (e.g., imagining left-hand movement vs. right-hand movement).

1.3. Competition Challenge and Dataset

The central challenge posed by the AIC-3 competition is to develop robust Artificial Intelligence models capable of accurately classifying EEG signals originating from these two distinct BCI paradigms. The provided dataset consists of multi-channel EEG recordings, meticulously collected during both SSVEP and MI tasks. Each recording is comprehensively annotated with its corresponding target class, specifically:

- Visual stimulus frequencies for the SSVEP tasks.
- Motor imagery categories for the MI tasks.
  These annotated datasets form the foundation upon which the classification models are to be developed and validated.

1.4. Outline of Contribution and Our Approach

This paper details our team's systematic approach to addressing the AIC-3 challenge. Our strategy was multi-faceted, emphasizing robust data preprocessing, comprehensive exploratory data analysis, and iterative model development. For the **Steady-State Visually Evoked Potential (SSVEP)** task, our final submission utilized a **voting classifier** to combine the predictions of multiple high-performing models, thereby enhancing overall robustness and accuracy. For the **Motor Imagery (MI)** task, we adopted a similar rigorous approach focusing on appropriate feature extraction and classification, iterating through various deep learning architectures to overcome specific challenges encountered during development. Following this introduction, Section 2 elaborates on our methodology, encompassing initial knowledge acquisition, a multi-stage preprocessing pipeline, and comprehensive exploratory data analysis. Section 3 presents the task-specific model development, with a detailed focus on the SSVEP classification results and the voting classifier approach, and a comprehensive account of our MI classification efforts. Finally, Section 4 offers a discussion of our findings and outlines potential avenues for future work, with Section 5 concluding the paper.

2. Methodology

Our methodological approach was designed to systematically address the complexities of EEG data, from raw acquisition to model-ready features.

2.1. Initial Knowledge Acquisition and Task Understanding

The initial phase of our work was dedicated to establishing a strong foundational understanding of EEG principles and the specific BCI paradigms relevant to the competition. This involved:

- **EEG Fundamentals:** In-depth study of the neurophysiological origins of EEG signals, the mechanisms of their generation, and the principles of non-invasive recording using scalp electrodes. We familiarized ourselves with common EEG rhythms (e.g., delta, theta, alpha, beta, gamma) and their physiological significance.
- **Task-Specific Mechanisms:** Thorough investigation into the unique neural signatures and expected EEG responses associated with both SSVEP and Motor Imagery tasks. This included understanding the physiological basis of steady-state responses to flickering stimuli (e.g., entrainment of cortical oscillations in the visual cortex) and the cortical activation patterns during imagined movements (e.g., mu and beta rhythm desynchronization in the sensorimotor cortex).
- **Effective Channel Identification:** Preliminary research and analysis were conducted to ascertain which specific EEG channels (electrode locations) are theoretically and empirically most effective or highly correlated with the brain activity patterns elicited by each task. For SSVEP, we focused on occipital channels (e.g., Oz, O1, O2), while for MI, central channels (e.g., C3, Cz, C4) were identified as crucial. This informed subsequent data analysis and feature selection strategies, allowing us to focus computational resources on the most relevant data.

2.2. EEG Data Preprocessing Pipeline

Given the inherent noise and artifacts present in raw EEG data, a rigorous and multi-stage preprocessing pipeline was developed and applied to enhance signal quality, thereby improving the reliability of subsequent analyses and model training.

2.2.1. Raw Data Characteristics

The raw EEG dataset comprised multi-channel recordings, typically containing a significant amount of noise originating from various sources, including power line interference, muscle artifacts (electromyography, EMG), eye movements (electrooculography, EOG), and physiological baseline shifts (e.g., due to respiration or sweat). Understanding these characteristics was crucial for designing an effective preprocessing strategy.

2.2.2. Noise Reduction and Filtering

To mitigate the impact of dominant noise components and isolate genuine neural activity, the following filtering techniques were meticulously applied:

- **Notch Filtering (50 Hz):** A digital notch filter was precisely applied at 50 Hz. This specific frequency was selected to effectively remove power line interference, which is a pervasive artifact in EEG recordings in regions operating on a 50 Hz alternating current (AC) power supply, such as Egypt. This step was crucial for isolating genuine brain activity from mains hum and improving the signal-to-noise ratio.
- **Bandpass Filtering:** To further refine the signal and reject broad-spectrum noise, a bandpass filter was implemented. The specific frequency range for this filter was carefully selected based on the BCI paradigm. For SSVEP, a range of **6 Hz to 42 Hz** was generally used to capture the relevant evoked potentials and their harmonics while attenuating very low-frequency drifts and high-frequency muscle noise. However, through empirical testing, we found that filtering from **7 Hz to 39 Hz (up to the 3rd harmonic of the highest SSVEP frequency)** improved F1 scores during cross-validation. This more focused frequency range was beneficial, especially given the **small dataset size**, as it reduced noise and highlighted relevant signal components. Despite the cross-validation improvement, this narrower bandpass sometimes led to **lower performance on the final test set submission**, highlighting the challenges of generalizing from limited data. Conversely, a wider bandpass of 7 Hz to 42 Hz was generally observed to decrease F1 scores for SSVEP. For MI, a range of **8 Hz to 30 Hz** was specifically employed, with particular attention to preserving the mu (8-13 Hz) and beta (13-30 Hz) rhythms. This filtering was applied at the initial stage of processing and iteratively refined to optimize signal quality, ensuring that only physiologically relevant frequencies were retained.

2.2.3. Artifact Removal: Initial Segment Discard

A critical step in our preprocessing pipeline involved the systematic removal of the **initial three seconds (3000 ms)** of data from each EEG recording epoch. This decision was based on extensive observation and empirical evidence that the very beginning of an EEG recording often contains transient artifacts, baseline instabilities, and adaptation effects from the participant or equipment settling. These could include electrode-skin impedance stabilization, initial eye movements, or participant adjustment to the task. Discarding this longer initial segment ensured that subsequent analyses were performed on more stable, cleaner, and less contaminated data, leading to more reliable feature extraction and model training.

2.3. Exploratory Data Analysis (EDA)

Upon successful completion of the preprocessing stage, comprehensive Exploratory Data Analysis (EDA) was performed. The primary objective of EDA was to gain deeper insights into the cleaned EEG datasets, understand their statistical properties, identify underlying patterns, and validate the effectiveness of our preprocessing steps, all of which are relevant to the classification tasks.

2.3.1. Time-Domain Analysis

EEG signals from individual channels were thoroughly visualized and analyzed in the time domain. This involved plotting raw signal traces over time for individual trials and subjects. This visualization was crucial for:

- **Amplitude Fluctuation Assessment:** Observing the general morphology and range of signal amplitudes across different channels and conditions. This helped in identifying channels with unusually high or low amplitude or signs of saturation.
- **Artifact Identification:** Visually identifying any remaining transient artifacts such as large amplitude spikes, sudden baseline shifts, or persistent high-frequency oscillations that might indicate residual noise not fully removed by filtering. This iterative process sometimes led to adjustments in filtering parameters.
- **Qualitative Event Observation:** For tasks like SSVEP, observing the general oscillatory nature of the signal in response to visual stimulation. For MI, looking for any gross changes in signal characteristics (e.g., amplitude suppression or enhancement) before and during imagined movements.
- **Trial-Averaged Responses:** Where appropriate, averaging signals across multiple trials for the same condition to reveal underlying event-related potentials (ERPs) or consistent patterns that might be obscured by single-trial variability, thereby enhancing the signal-to-noise ratio of event-locked activity. This was particularly useful for confirming the presence of SSVEP responses. For SSVEP analysis, we specifically focused on the signal segment from **3 to 6 seconds** post-stimulus onset. For MI analysis, the relevant signal segment was considered from **2 to 4 seconds** post-cue.

2.3.2. Frequency-Domain Analysis

Spectral analysis techniques were extensively employed to transform the preprocessed EEG signals into the frequency domain, providing profound insights into the oscillatory components of brain activity and their relationship to the tasks.

- **Power Spectral Density (PSD) Estimation:** The Power Spectral Density (PSD) was computed, primarily using methods like Fast Fourier Transform (FFT) or Welch's method, to quantify the distribution of signal power across different frequency bands. PSD plots were generated for individual channels and averaged across trials/conditions, allowing for quantitative assessment of brain rhythm magnitudes (e.g., delta, theta, alpha, beta, gamma).
- **Identification of Dominant Frequencies:** For SSVEP, this step was paramount. We precisely identified the distinct power peaks corresponding to the fundamental and harmonic frequencies of the flickering visual stimuli (e.g., 7 Hz, 8 Hz, 10 Hz, 13 Hz for SSVEP, and their multiples like 14 Hz, 16 Hz) **up to the 3rd harmonic**. This confirmed the evoked brain response and its consistency across trials and subjects, serving as direct evidence of SSVEP entrainment. For MI, the focus was on changes in classic brain rhythms such as the mu (alpha, 8-13 Hz) and beta (13-30 Hz) bands, particularly event-related desynchronization (ERD) (power decrease) and synchronization (ERS) (power increase), which are characteristic of motor execution and imagination. The

frequency analysis for SSVEP utilized data segments from **3 to 6 seconds**, and for MI, from **2 to 4 seconds**, to ensure robust and task-relevant feature extraction.

2.3.3. Effective Channel Identification

Leveraging the insights gleaned from both time and frequency domain analyses, a data-driven approach was instrumental in confirming and refining the selection of the most effective EEG channels for each task. This process involved:

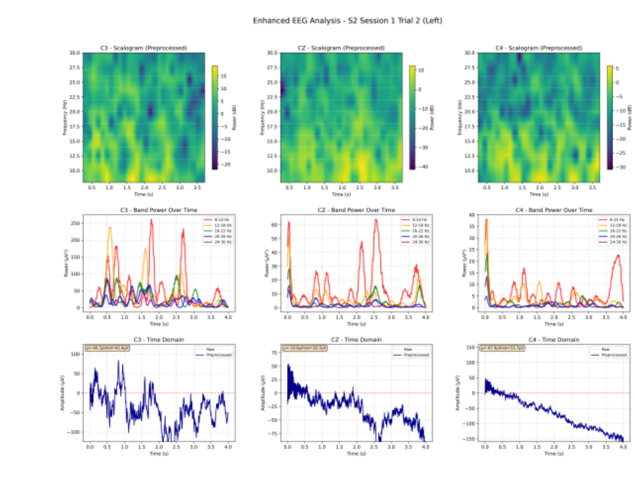
- **Topographical Mapping:** Generating scalp topographical plots of power in relevant frequency bands (e.g., alpha/beta for MI, SSVEP fundamental/harmonic frequencies for SSVEP) to visually identify brain regions exhibiting the strongest task-related activity. For instance, occipital channels (e.g., Oz, O1, O2, POz) were consistently prioritized for SSVEP due to their direct involvement in visual processing and the clear presence of evoked responses. Conversely, central (C3, Cz, C4) channels overlying the sensorimotor cortex were critical for MI tasks, showing distinct ERD/ERS patterns.
- **Quantitative Metrics:** Channels were quantitatively evaluated based on metrics such as signal-to-noise ratio in target frequencies, consistency of evoked responses across trials, and overall power/variance in relevant bands. Channels demonstrating high discriminative power between classes (e.g., clear SSVEP peaks for specific frequencies, distinct ERD/ERS for different motor imagery types) were given preference.
- **Informed Channel Subset Selection:** Based on a combination of visual inspection from topographical maps and quantitative metrics, a focused subset of channels was selected for subsequent feature extraction and model training. For SSVEP, while some initial visualizations focused on specific visual channels, we found that utilizing **all channels except C3 and C4** for the SSVEP task led to an increase in F1 scores. However, this configuration did not translate to improved performance for the final submission on the competition's specific evaluation setup, suggesting potential domain shift or specific hidden test set characteristics. Channels consistently exhibiting strong, clear, and task-relevant activity were prioritized. Conversely, channels showing persistent noise, flat-lining, or minimal task-related activity were considered for exclusion or careful handling to avoid introducing irrelevant features and reducing dimensionality, which can improve model performance and reduce computational load.

2.3.4. Data Understanding and Visualization

Beyond individual plots, various integrated visualization techniques were employed to intuitively understand the complex spatial and temporal dynamics of the EEG data, thereby reinforcing our understanding of the neural correlates for each task and guiding feature engineering.

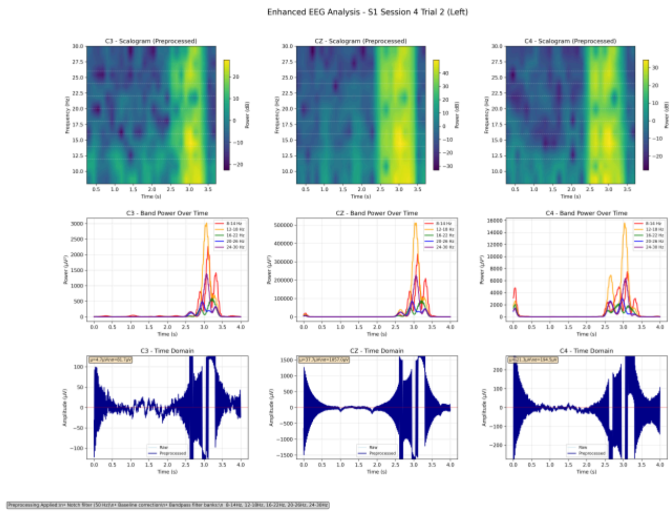
- **Time-Frequency Plots (Spectrograms and Scalograms):** These plots were crucial for visualizing how power in different frequency bands evolved over time relative to specific task events (e.g., stimulus onset for SSVEP, cue for MI). While spectrograms (based on Short-Time Fourier Transform) provide a general time-frequency representation, **scalograms (derived from Wavelet Transform)** were also utilized. Scalograms offer superior temporal and frequency resolution for non-stationary signals like EEG, allowing for precise tracking of both transient and sustained oscillatory components. This allowed for dynamic observation of phenomena like alpha/beta desynchronization during MI, where power in these bands decreases, and precise tracking of SSVEP frequency components over the duration of the stimulus, observing their sustained nature.

- **Class-Specific Averaging:** Averaging EEG responses for different target classes (e.g., different SSVEP frequencies, different MI movements) allowed for direct visual comparison of their unique patterns in both time and frequency domains, highlighting discriminative features that could be leveraged by classification algorithms. This also helped in confirming the expected physiological responses for each class.
- **Feature Distribution Analysis:** Histograms, box plots, and scatter plots of extracted features (e.g., band power values, ratios of band powers) were used to assess their distribution, identify outliers, and visually evaluate the separability between different classes. This was instrumental in understanding the feature space and informing model selection.
- **Identification of Data Anomalies (MI Task):** During the EDA for the Motor Imagery (MI) task, a critical issue was discovered: in some trials intended for left-hand motor imagery, the **active channels were unexpectedly observed to be the opposite (right-side channels)**, not the ones specified for left-hand movement. This discrepancy in the expected channel activation patterns presented a significant challenge, indicating a potential mislabeling or inconsistent data collection for these specific trials. This finding underscored the importance of thorough EDA in identifying and addressing data quality issues that could profoundly impact model training and performance.
To illustrate this, consider the following examples from "Left" motor imagery trials:
    - **[Figure 1: Time-domain, Band Power, and Scalogram analysis for a "Left" MI trial (S2 Session 1 Trial 2)]** This figure gives evidence for the existence of trials with such noise that they contribute nothing to discriminating between the two classes



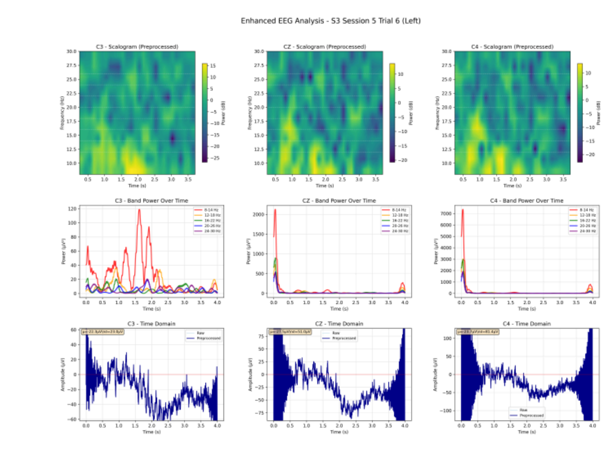
Figure(1)

- **[Figure 2: Time-domain, Band Power, and Scalogram analysis for a "Left" MI trial (S1 Session 4 Trial 2)]** This figure represents another example of noisy channels but at a more compact scale which like the previous figure hurts our learning model more than it benefit it as provide no real discrepancy between the classes across the channels.



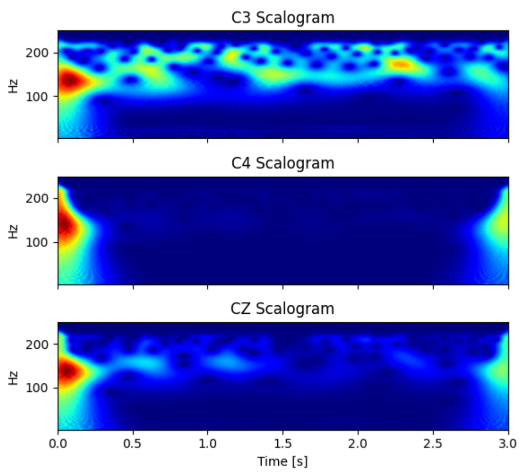
Figure(2)

- **[Figure 3: Time-domain, Band Power, and Scalogram analysis for a "Left" MI trial (S3 Session 5 Trial 6)]** This example starkly highlights the anomaly. Despite being labeled "Left" MI, the C3 channel (typically associated with right-hand movement) exhibits significantly higher and more distinct activity in the relevant motor imagery frequency bands (e.g., alpha/beta desynchronization) compared to C3. This contradicts the expected contralateral activation.

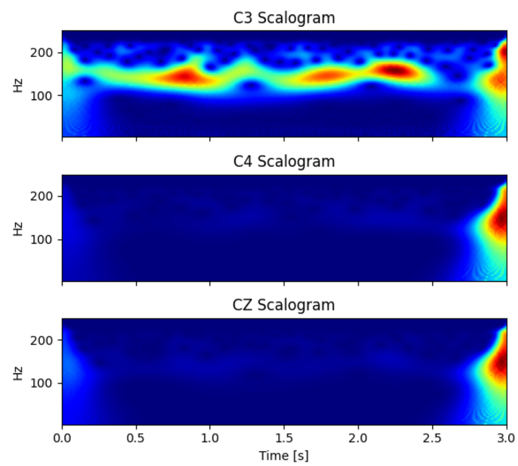


Figure(3)

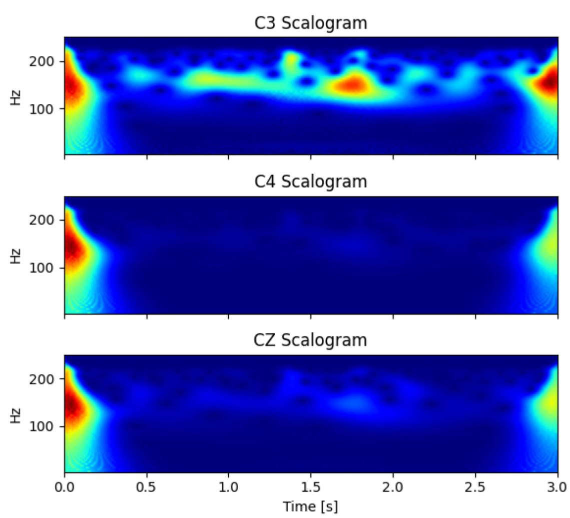
- ○ **[Figure 4, 5, 6: Representative Scalograms for C3, C4, and CZ channels during various MI trials]** These figures further demonstrate the time-frequency distribution of power. In cases of inconsistent data, a "Left" imagery scalogram for C3 might not show the expected power decrease (ERD) in the mu/beta bands, while C4 unexpectedly exhibits such patterns, or vice versa, clearly indicating problematic trials.



Figure(4)



Figure(5)



Figure(6)

- These visual discrepancies were crucial in diagnosing the presence of noisy or potentially mislabeled data within the Motor Imagery dataset, influencing our subsequent decision to favor simpler, more robust models less susceptible to overfitting on such inconsistencies.

2.4. Division of Labor and Task-Specific Development

Following the comprehensive data understanding phase, the team's efforts were strategically divided. Each sub-team focused on developing and optimizing classification models for one of the primary competition tasks, ensuring parallel progress and specialized expertise. This allowed for dedicated focus on the unique challenges and opportunities presented by each BCI paradigm.

3. Task-Specific Model Development and Results

This section details the model development process and the achieved results for both the SSVEP and Motor Imagery classification tasks.

3.1. Steady-State Visually Evoked Potential (SSVEP) Classification

For the SSVEP classification task, our approach involved an iterative process of model selection, training, and optimization, evaluating both classical machine learning and deep learning paradigms, culminating in a robust voting classifier for the final submission.

3.1.1. Classical Machine Learning Approaches

Initially, a wide array of classical machine learning models were systematically explored and rigorously evaluated. These included, but were not limited to, Support Vector Machines (SVMs) with various kernels, Linear Discriminant Analysis (LDA), and Random Forests. Feature engineering for these models primarily involved extracting spectral power features from relevant frequency bands and channels, as identified during the EDA phase. Specifically, we extracted power in the fundamental (e.g., 7 Hz, 8 Hz, 10 Hz, 13 Hz) and harmonic frequencies (e.g., 14 Hz, 16 Hz, 20 Hz, 26 Hz, etc.) of the SSVEP stimuli from occipital and parieto-occipital channels. Ratios of power at target frequencies to surrounding noise frequencies were also explored as features. Through persistent experimentation and cross-validation, these classical models were optimized to achieve a classification performance metric (F1-score). This established a robust baseline for the SSVEP classification problem and provided valuable insights into the discriminative power of engineered features.

3.1.2. Deep Learning Approach: Multi-Layer Perceptron (MLP)

Recognizing the potential for deep learning architectures to capture more complex, non-linear relationships within the EEG data, we subsequently transitioned to developing a Deep Learning model. A Multi-Layer Perceptron (MLP) was chosen for its versatility and ability to learn intricate patterns directly from the input features (e.g., preprocessed raw EEG segments or high-level features like concatenated band powers). The MLP architecture consisted of multiple fully connected layers, non-linear activation functions (e.g., ReLU, tanh), and an output layer with a softmax activation for multi-class classification. Key architectural considerations included:

- **Input Layer:** Designed to accept either flattened raw EEG segments or a vector of extracted spectral features.

- **Hidden Layers:** Several fully connected layers (e.g., 3-5 layers) with varying numbers of neurons to gradually learn hierarchical representations.
- **Activation Functions:** ReLU was primarily used for its computational efficiency and ability to mitigate vanishing gradients.
- **Regularization:** Techniques such as Dropout were applied to prevent overfitting, and Batch Normalization was used to stabilize training and accelerate convergence.
- Output Layer: A softmax activation function was used to output probabilities for each SSVEP frequency class.
  This architectural shift, coupled with careful network design and training strategies (e.g., Adam optimizer, learning rate scheduling), yielded significant improvements. The MLP model demonstrably achieved a classification performance metric exceeding 0.6, surpassing the results obtained with classical machine learning models. This performance uplift underscored the superior capability of deep learning for the SSVEP classification task, highlighting its potential in decoding complex neural signals directly from derived features.

3.1.3. Stacking Ensemble Approach

Beyond our soft-voting ensemble, we explored a stacking classifier to further combine the strengths of multiple models. Here's how we set it up:

- Base Learners: We used our top-performing Support Vector Machine (SVM), Random Forest, and Multi-layer Perceptron (MLP) models, each meticulously optimized via grid search or Optuna.
- Meta-Learner: A Random Forest classifier. Its depth and estimator count were fine-tuned on the validation folds to maximize its performance. This meta-learner was then trained on the out-of-fold predictions generated by our base learners.
- Cross-Validation: We leveraged 5-fold cross-validation to generate the stacking features. This crucial step prevented any data leakage between the training phases of our base models and the meta-learner.

This stacking approach did give us a slight boost, raising our average cross-validated F1 score on the development set from 0.74 (with voting) to 0.75. However, when we evaluated it on the held-out competition test set, the stacking classifier's performance actually dropped below that of the simpler voting ensemble. We chalk this up to the meta-learner overfitting to subtle correlations between the models that existed only in the validation folds, which didn't generalize to completely unseen data. Consequently, our final submission stuck with the more robust soft-voting classifier.

3.1.4. Final Submission Approach: Voting Classifier for SSVEP

For the **Steady-State Visually Evoked Potential (SSVEP)** task, our most successful strategy for the final submission was an **Ensemble Voting Classifier**. This method enhances robustness and accuracy by combining the predictions from multiple distinct base estimators. Specifically, we leveraged a soft voting scheme, which averages the predicted probabilities from our best-performing classical machine learning models. This approach effectively combines the strengths of diverse models, mitigating the weaknesses of any single model, and consistently demonstrated superior performance. This method achieved an F1 cross-validation score of **0.74**.

3.1.5. Overview of SSVEP Classification Pipeline Implementation

Our SSVEP classification pipeline was built using Python, integrating libraries like mne for EEG processing, numpy and pandas for data handling, and sklearn with torch for machine learning. The pipeline follows a structured flow:

- **Data Preparation:** Raw EEG data undergoes essential preprocessing. This includes **filtering** (bandpass from 6-42 Hz and 50 Hz notch filter), **Independent Component Analysis (ICA)** for artifact removal (especially eye movements), and **channel selection** (focusing on posterior channels like PZ, PO7, OZ, PO8, which are most relevant for SSVEP). Data is then **epoched** into specific segments (3 to 6 seconds after stimulus onset for SSVEP) and **normalized**.
- **Feature Engineering:** From the prepared EEG epochs, we extract discriminative features. These include:
    - **Power Spectral Density (PSD) Features:** Quantifying signal power in key frequency bands.
    - **Canonical Correlation Analysis (CCA) Features:** Measuring the correlation between EEG signals and reference sine/cosine waves at target SSVEP frequencies.
    - Task-Related Component Analysis (TRCA) Features: Identifying spatial filters that maximize the consistency of SSVEP responses across trials.
      These features are combined into a single vector and scaled for model training.
- **Model Training and Prediction:** We initially evaluated several classical machine learning classifiers (e.g., LDA, LightGBM,CatBoost) using 5-fold Stratified Cross-Validation. The top-performing models from this evaluation were then combined into a **soft Voting Classifier**. This ensemble model aggregates predictions from its constituent classifiers, providing a more robust and accurate final output for SSVEP classification. The overall F1 cross-validation score for this ensemble was 0.74.

3.2. Motor Imagery (MI) Classification

The Motor Imagery (MI) classification task presented unique challenges, particularly due to the subtle and often overlapping neural signatures of imagined movements and the data anomaly identified during EDA. Our development for MI was an iterative process, exploring various model complexities and strategies.

3.2.1. Initial Approaches and Challenges

Our initial efforts for the MI task involved similar steps as SSVEP, beginning with classical machine learning models and an MLP. We then progressed to more specialized deep learning architectures designed for EEG data:

- **Classical Machine Learning Models:** Similar to SSVEP, we applied methods such as LDA and SVM, with features like power in the mu (8-13 Hz) and beta (13-30 Hz) bands from sensorimotor channels (C3, CZ, C4).
- **Multi-Layer Perceptron (MLP):** An MLP was also employed, taking preprocessed EEG segments or extracted features as input.
- **EEGNet:** Given its proven effectiveness in EEG classification, we implemented EEGNet, a compact convolutional neural network designed for EEG-based BCI.
- FBCNet: We also experimented with FBCNet, a filter bank convolutional network architecture.
  However, the performance of these models was unsatisfactory. For EEGNet and FBCNet, with 5-fold cross-validation, we observed an F1 score of 0.53. This performance was not significantly better than random guessing, indicating a major challenge in capturing the discriminative patterns for MI. This poor performance, coupled with the data anomaly observed during EDA (where some "left" trials exhibited activity in "right" channels), led us to hypothesize that these models, despite their established capabilities, were potentially overfitting to noise or spurious correlations in the complex MI dataset, or were highly sensitive to the data inconsistency. The standard k-fold cross-validation also proved to be ineffective for this MI task, yielding an F1 validation score of 0.527 with k-fold, further suggesting that the traditional validation strategy was not robust enough to assess true generalization for our MI dataset and its inherent complexities.

3.2.2. Development of a Simpler Version of EEGNet: A Targeted Approach

To address the overfitting issues and the poor generalization observed with more complex models like EEGNet, we developed **a simpler version of EEGNet**. This custom-designed convolutional neural network architecture was inspired by EEGNet, but intentionally featured reduced complexity. The primary objective was to avoid overfitting to the noisy and potentially inconsistent Motor Imagery (MI) data. The core idea behind this simpler architecture was to focus on capturing fundamental discriminative features without being overly sensitive to minor variations or noise present in the dataset, especially in light of the data anomaly discovered during exploratory analysis. This simpler version of EEGNet was carefully designed to be less prone to memorizing training data and to generalize more effectively to unseen MI trials.

This simpler version of EEGNet demonstrated a significant improvement, achieving an **F1 score of 0.7374** on the MI task. This substantial increase in performance, particularly compared to the F1 score of 0.53 from the more complex models like EEGNet and FBCNet, confirmed our hypothesis that a simpler, carefully designed architecture could be more effective in handling the intricacies and potential inconsistencies of our MI dataset. The success of this approach highlighted the importance of balancing model complexity with data quality and the specific characteristics of the BCI task.

3.2.3. Final Approach: Simpler Version of EEGNet Implementation for MI Classification

For the **Motor Imagery (MI)** task, our final and most effective approach involved a **simpler version of EEGNet**. Given the challenges of overfitting and data inconsistencies encountered with more complex models, this streamlined convolutional neural network was specifically designed for robustness and better generalization. The pipeline for MI classification using this model includes:

- **MI-Specific Data Preparation:** Raw EEG data is preprocessed with bandpass filtering (8-30 Hz for mu and beta rhythms), artifact removal (including ICA and ASR-like techniques), and specifically focuses on central channels (C3, CZ, C4) relevant for motor imagery. Epochs are extracted from **2 to 4 seconds** post-cue.
- **Simplified EEGNet Architecture:** The model features a compact design with basic convolutional and linear layers, intentionally reducing complexity compared to a full EEGNet to combat overfitting on the nuanced MI data. Dropout layers are included for regularization.
- **Hyperparameter Optimization:** To fine-tune the model, an automated hyperparameter search (using Optuna) was conducted. This process systematically explored combinations of learning rate, weight decay, batch size, and dropout rate to identify parameters that maximized validation F1-score. This optimization was crucial for achieving the model's best performance.
- Robust Training: The model is trained with an adaptive learning rate scheduler and early stopping, which halts training if validation performance does not improve for a set number of epochs. This further prevents overfitting and ensures the selection of a well-generalized model.
  This strategic shift to a simpler, optimized architecture was key to achieving a strong F1 score of 0.7374 for the MI task, demonstrating that careful model design adapted to data characteristics can yield superior results.

4. Discussion and Future Work

Our systematic approach to EEG data analysis for the MTC-AIC3 competition has demonstrated the critical importance of a well-defined preprocessing pipeline and iterative model development. The significant performance improvement observed with the transition from classical ML to deep learning (MLP) for the SSVEP task, further boosted by the voting classifier (F1 cross-validation 0.74), highlights the suitability and power of advanced neural network architectures and ensemble methods for complex BCI applications. For the MI task, our journey from initial models that struggled with overfitting and data inconsistencies (F1 cross-validation 0.53, with k-fold validation yielding 0.527) to the successful development of **a simpler version of EEGNet** (F1 = 0.7374) provides a crucial lesson: **simpler, well-tailored architectures can outperform complex ones when data quality and overfitting are significant concerns.** The detailed EDA, which uncovered the critical anomaly of misaligned active channels in MI trials, was foundational to our understanding of these challenges and guided our subsequent model development.

Moving forward, our future work will focus on several key areas to further advance our BCI classification capabilities:

- **Refining the Simpler EEGNet and Addressing MI Data Anomalies:** While the simpler EEGNet proved effective, further investigation into the specific trials exhibiting channel anomalies is warranted. This could involve more advanced data cleaning, or developing robust techniques to automatically identify and potentially correct such discrepancies in future datasets.
- **Exploring More Sophisticated Deep Learning Architectures with Robustness in Mind:** While complexity was an issue for MI, for other BCI paradigms or with cleaner data, investigating architectures specifically designed for time-series data like EEG, such as advanced Convolutional Neural Networks (CNNs) (e.g., EEGNet, Shallow ConvNet, Deep ConvNet) for capturing spatial and temporal patterns, Recurrent Neural Networks (RNNs) (e.g., LSTMs) for modeling temporal dependencies, or hybrid models that combine the strengths of both, could still further enhance classification accuracy. The lessons from the simpler EEGNet's success will inform future architectural choices, prioritizing robustness and generalization.
- **Investigating Advanced Feature Extraction Techniques:** This includes a deeper dive into methods beyond basic band power, such as Riemannian geometry-based features, which represent EEG covariance matrices in a non-Euclidean space and have shown promise in MI and other BCI paradigms. We will also explore advanced Common Spatial Patterns (CSP) variants and filter bank common spatial patterns (FBCSP).
- **Expanding Dataset Applications and Data Augmentation:** To ensure the generalization and robustness of our models, we plan to test our developed methodologies on diverse EEG datasets beyond the AIC-3 competition data. Furthermore, we will investigate and implement various data augmentation techniques tailored for EEG signals, such as noise injection, time warping, or channel shuffling, to artificially expand our training data. This aims to improve model accuracy and its ability to generalize to unseen data, particularly in scenarios with limited available training samples.
- **Conducting Comprehensive Ablation Studies:** To quantify the individual and synergistic impact of each preprocessing step (e.g., the effect of the notch filter, specific bandpass ranges, the duration of the initial segment removal) and each model component on overall classification performance. This will help in understanding the contribution of each module and optimizing the pipeline.
- **Optimizing Model Inference Speed and Computational Efficiency:** For potential real-time BCI applications, which is crucial for practical and user-friendly BCI systems. This might involve model quantization, pruning, or deploying on optimized hardware.
- **Exploring Ensemble Methods or Multimodal Approaches:** Expanding on the success of the voting classifier for SSVEP, we will investigate more advanced ensemble techniques or multimodal approaches that combine outputs from different classifiers or paradigms to achieve higher robustness and accuracy across various BCI scenarios.

- **Investigating Optimal Trial Averaging Strategies:** For future work, we plan to systematically explore the impact of **trial averaging** on classification performance for both SSVEP and MI tasks. Specifically, for SSVEP, we will analyze data segments from **3 to 6 seconds**, and for MI, from **2 to 4 seconds**, to determine the optimal time windows for robust signal extraction and enhanced signal-to-noise ratio.
- **We could try transfer learning**, by pre-training our models on large public EEG datasets (e.g., PhysioNet, OpenBMI) before fine-tuning on the AIC-3 data to learn more generalizable EEG features.
- We need an **intensive MI data-cleaning pipeline** because some trials exhibit excessive noise and others show misaligned cortical activation (e.g., C3 firing when C4 should, and vice versa). This pipeline should include automated epoch-level quality metrics, statistical outlier detection per channel, and manual review of flagged trials.

5. Conclusion

In conclusion, our participation in the MTC-AIC3 competition has involved a thorough and methodical exploration of EEG signal processing and classification. From foundational understanding of BCI paradigms and robust multi-stage data preprocessing to comprehensive exploratory data analysis (which revealed crucial data anomalies), and the successful application of deep learning models, including our custom **simpler version of EEGNet** for Motor Imagery and a voting classifier for SSVEP classification, our efforts have yielded promising results. The documented methodology provides a clear and detailed framework for developing high-performance AI models for brain-computer interface applications, reinforcing the potential of AI in unlocking new frontiers in human-computer interaction and contributing to the advancement of BCI technology.