# Decoding Motor Imagery based EEG Signals

Cole Smylie*, Haroon Mushtaq†, Zulfidin Khodzhaev‡
*Department of Computer Science
†Department of Aerospace Engineering
‡Department of Electrical and Computer Engineering
The University of Texas at Austin

***Abstract*—Brain Computer Interfaces represent a life-changing possibility for disabled patients to regain control of their life, for example through highly functional prosthetics or movement devices. However, everyday movements, like walking, and the increased brain activity it causes poses significant difficulty for BCI decoders. This paper proposes a method to clean EEG data and create a decoder that is resistant to walking induced artifacts so left and right hand motor imagery can be classified. We first perform Artifact Subspace Reconstruction to remove major artifacts, then bandpass in the alpha and mu band to find physiologically relevant frequencies, Common Average Reference Filter to improve spatial resolution, extract trials and samples, find the mean percent difference in power spectrum density between resting and motor imagery to best identify event-related desynchronization, perform fisher score feature selection to reduce dimensionality, and finally train and cross validate our model offline and test online with evidence accumulation.**

## I. Introduction

Advances in Brain Computer Interface (BCI) technologies enable and enhance rehabilitation in individuals suffering from certain disabilities or neurological conditions helping them recover lost functionalities. One promising paradigm within the field of BCIs is Motor Imagery, which refers to a process in which a subject imagines performing a movement without any physical manifestation of the movement. This process activates the same sensory and motor regions in the brain as actual movements [6]. The benefit of MI signals is that they can be used to convey information about motor intent; thus, can be useful in controlling a wheelchair, prosthetic or some device.

Despite this, MI-based BCI systems face significant challenges to widespread adoption in settings beyond the research lab. One such challenge is precisely decoding the neural signals generated by the BCI user. That is, how does the intent of the BCI user get accurately translated to the device being controlled? Inaccuracies in decoding typically result from characteristics of the EEG signal that include low signal-to-noise ratio, non-stationary data, and inter-subject variability [1].

Existing methods from digital signal processing and machine learning can be used enhance decoder performance. In this report, we study decoder performance based on MI-EEG data for three subjects. The experimental task was to engage in motor imagery of the left or right side given a visual cue. Feedback was provided to subjects after each trial indicating correct or incorrect delivery. We use filtering and artifact removal techniques to pre-process our MI-EEG data. Further, we extract and select PSD features that were used to train our decoder before testing on online sessions.

## II. Related Works

Nisar, et al. [7] study enhancing MI-based BCI decoder performance using band power, approximate entropy, statistical features, wavelet-based features, and CSP as features. They used decision tree, Random Forest, SVM, KNN, and ANN to perform classification and tested on a motor imagery database with 64-channel EEG signals for 50 subjects performing four MI tasks. They achieved an accuracy of 98.53% using an SVM classifier.

Athif and Ren [8] propose WaveCSP, which extracts 24 features from EEG signals using wavelet transform and CSP filtering. Their algorithm shows improvement in the performance in handling subject variability concerns when comparing with competing methods. On the MI database, more than half of the 109 subjects achieved accuracy higher than 64%.

Another important contribution of this study is that the authors propose methods to study the effect of limited number of electrodes, limited spatial distribution of electrodes, lower signal quality, subject variability, and BCI literacy on the performance of MI classification.

## III. Experiment Design

The experiment was conducted with 3 healthy subjects - referred to as subject 41, subject 42, and subject 43. Each subject took part in an offline session and a series of online motor imagery sessions across two days. This included a calibration run before each session that consisted of resting state, either sitting or walking with a clear mind

During the offline session, each subject performed motor imagery in both their left and right hands. The individual imagery movement was under each subject's discretion. Subject 41 imagined moving the whole arm up to the side, doing a lateral raise motion. Subject 42 imagined moving the arm up to the head level. Subject 43 imagined the reach-and-grab task motion. There were 4 runs total, with 10 trials for each hand per run. Data was recorded from 59 EEG channels, excluding channels T7, M1, T8, M2, and EOG. This calibration data established baseline neural signals for each subject's unique motor imagery patterns.

Subjects also took part in online sessions where they used motor imagery to deliver commands in real time while receiving feedback. Sessions 2, 3, and 5 were conducted while

subjects were sitting. Sessions 4 and 6 required subjects to perform the motor imagery while walking on a treadmill. Each online session consisted of 4 or 5 runs of 10 left and 10 right-hand motor imagery trials. Real time feedback was provided to subjects during each trial indicating correct or incorrect delivery of their motor imagery commands. The online brain-computer interface used 1-second EEG samples, making new predictions every 62.5 ms.

Having both sitting and walking conditions allowed for the comparison of motor imagery neural control across different ambulatory states for subjects 41, 42 and 43. The calibration data specific to each subject and session also helped account for variability in non-stationary EEG patterns and between individuals.

## IV. Methodology

### A. EEG Data Collection

- **Day 1**: Participants performed an offline motor imagery session and then an online sitting session.
- **Days 2 & 3**: Participants performed 2 online motor imagery sessions, one sitting and one walking.

All sessions contained triggers synchronized to an arrow shown on a laptop screen. Data was provided at a sample rate of 512 Hz. Samples were taken in a 1 second window, with offline sessions not overlapping, but online sessions having a 62.5 ms hop size so that constant feedback can be provided.

### B. Preprocessing

*1) Artifact Subspace Reconstruction:* We used Artifact Subspace Reconstruction (ASR) because it is an online-capable method for removing transient, large-amplitude artifacts from multi-channel EEG recordings. It is shown to effectively remove muscle and motion artifacts. This is ideal for our online motor imagery experiment that involves walking trials.

The developers of this method recommend a cutoff threshold of between 20 and 30, with lower values being more aggressive removal, in order to best preserve the brain activity signals while removing significant artifacts, Fig. 1. Since the central challenge of our experiment is to create a decoder that can still classify left and right hand motor imagery even in the presence of walking artifacts, we chose the most aggressive recommended threshold of 20. Each session's calibration data, sitting or walking, was used as clean EEG data to calibrate the ASR algorithm. Before running ASR, it requires the data to be High Pass filtered between 0.5 and 100 Hz, which we did with a causal Butterworth Bandpass filter. It is necessary to use a causal filter here as our decoder has to run in real time, and therefore cannot require non-causal filtering. [2].

*2) Temporal Filtering:* We used a Bandpass filter to filter our data in the Alpha and Beta frequency bands. Bandpass filtering is a common technique used in EEG signal processing to isolate specific frequency bands of interest. It works by combining a low-pass filter, which permits frequencies below a certain cutoff, and a high-pass filter, which allows frequencies above a different cutoff. It has been repeatedly found that

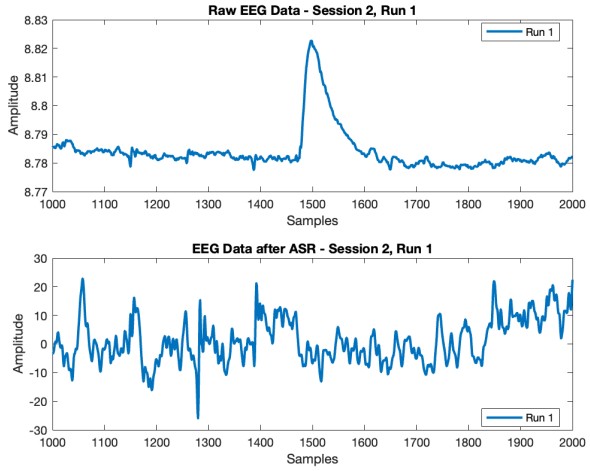


Fig. 1. The figure presents a comparative analysis of EEG data before and after ASR processing. The upper subplot, labeled "Raw EEG Data - Session 2, Run 1," depicts a line graph that is mostly stable with a notable spike occurring around the 1500 sample point, with amplitude values oscillating narrowly between 8.77 and 8.83. The lower subplot, "EEG Data after ASR - Session 2, Run 1," illustrates the filtered signal where the spike is not present.

motor imagery sensorimotor rhythms are usually within these frequency bands, [9], [10].

Since our classifier is specifically trying to classify motor imagery signals, isolating the Alpha and Beta bands will improve our signal-to-noise ratio by effectively suppressing noise frequencies and enabling a more precise analysis and interpretation of the underlying motor imagery signals. We again use a a causal Butterworth Bandpass filter of fourth order so that our decoder can run in online environments. Additionally, the Butterworth filter's characteristics of maintaining a flat frequency response and a linear phase make it ideal for isolating motor imagery-related sensorimotor rhythms within the Alpha and Beta frequency bands in EEG data.

*3) Spatial Filtering:* We used Common Average Reference (CAR) filtering to improve spatial resolution of our data. It is another common technique used in EEG signal processing to reduce shared noise across the electrode array, [11]. It enhances the quality of the recorded neural signals by reducing common noise sources, such as environmental noise or physiological artifacts. The average signal across all electrodes is calculated for each time point, and this average signal is then subtracted from the signal recorded at each electrode It involves using the average of signals over the whole brain as a reference [3]. We chose to use a CAR spatial filter because it was plausible that common noise was present due to the increased density of the 64-electrode EEG cap.

The main risk with CAR filtering is that if there is any channel with large artifacts, it will be spread to all other electrodes and can contaminate the entire dataset. A common fix for this is to remove artifactual channels before performing CAR filtering. However, this relies entirely on physiological knowledge and predictions. These choices could remove beneficial signals as well since at this stage each channels

contribution towards classification is not understood. Instead, we choose to leave all channels in for the CAR filtering so that we can later analyze their impact on classification and choose the most discriminant features to use in our decoder. We feel confident in doing this because we have already performed the most aggressive ASR filtering that is recommended in order to remove artifacts. Therefore the risk of spreading artifacts across channels is low.

*4) Feature Extraction:* We used Power Spectral Density (PSD) in order to reduce the dimensionality of our signal and provide more relevant features. PSD provides a measure of the power of different frequencies present in a signal; it transforms the data from the time domain to the frequency domain. In the context of EEG signals, PSD can be used to extract features based on various frequencies [4]. Since it is often known what frequency band various types of neural signals modulate in, this technique is highly effective at improving the quality of data.

We chose the Welch function for estimating PSD due to its ability to handle short, non-stationary EEG segments effectively. We chose 50 percent overlap and a window size of 128, or 0.25 seconds. These parameter choices were to equally balance both temporal resolution vs frequency detail and variance reduction vs adaptability to non-stationary signals. We chose to take 2 Hz bins between 8 and 30 Hz as we know this alpha/beta band is the frequency we are interested in. This reduced our data's dimensionality from 512 x 59 to 12 x 59.

We did not use the PSD of motor imagery samples directly, rather we compared them to the PSD of our resting state calibration, which acted as a baseline, by finding their percent difference as in Eq. 1.

$$\%difference = \frac{PSD_{MI} - PSD_{rs}}{PSD_{rs}} \tag{1}$$

This calculation highlights the relative change in power during motor imagery instead of general signal power. This technique leaves us with a good representation of Event-Related Desynchronization (ERD) which is a proven phenomenon of motor imagery [12].

*5) Feature Selection:* We used Fisher Score for selecting the best features for our model. Fisher scores are a measure of the discriminate power of individual features, with higher values indicating more discriminate features. It does this by calculating, as in Eq. 2 how similar a feature is within a class, but also how different it is across classes. This finds the features that will be the beneficial for a classifier to reduce dimensionality while still maintaining the most relevant features. [5].

$$F = \frac{\mu_2 - \mu_1}{\sigma_1^2 + \sigma_2^2} \tag{2}$$

Because EEG data is non-stationary, we must consider how the stable the fisher scores are, or how much they change across runs. To do this we first calculate fisher scores across each of the four runs. Then we rank all of the features by their fisher score in each run. This lets us see which features are stable discriminates for our data. We then average these ranks in order to get a combined ranking across runs. This effectively normalizes the data so that a run with higher average fisher scores or one very high fisher score cannot dominate the average.

Fischer scores for session one, run one to four are shown in Figs. 2, 3, 4, and 5. We see common Fischer scores across runs one to four. Higher discriminant power represents features that better differentiate the classes.

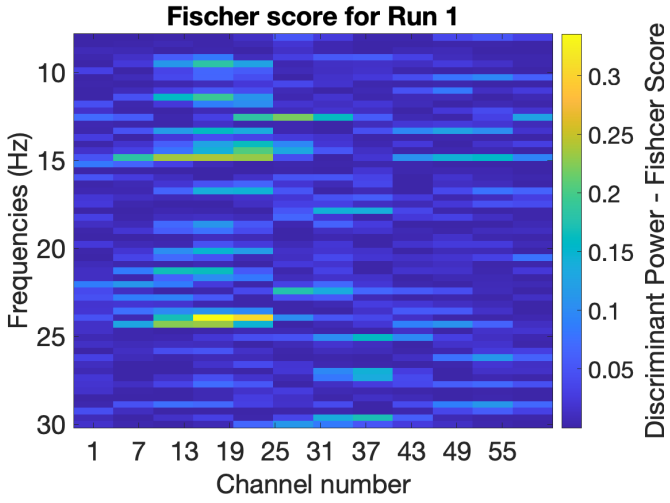


Fig. 2. Fischer scores measuring statistical significance for the first experimental run one in session one for subject 41. The y-axis shows Fischer score values ranging from 8 to 30 Hz with $\Delta 2$. The x-axis shows the number of channels from one to fifty-five. The colorbar represents the discriminant power of the Fischer score with higher scores indicating features that better differentiate the conditions/classes.

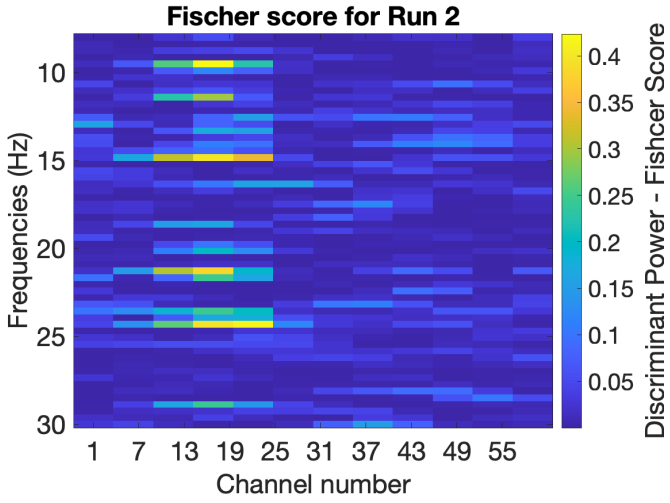


Fig. 3. Fischer scores measuring statistical significance for the first experimental run two in session one for subject 41. The y-axis shows Fischer score values ranging from 8 to 30 Hz with $\Delta 2$. The x-axis shows the number of channels from one to fifty-five. The colorbar represents the discriminant power of the Fischer score with higher scores indicating features that better differentiate the conditions/classes.

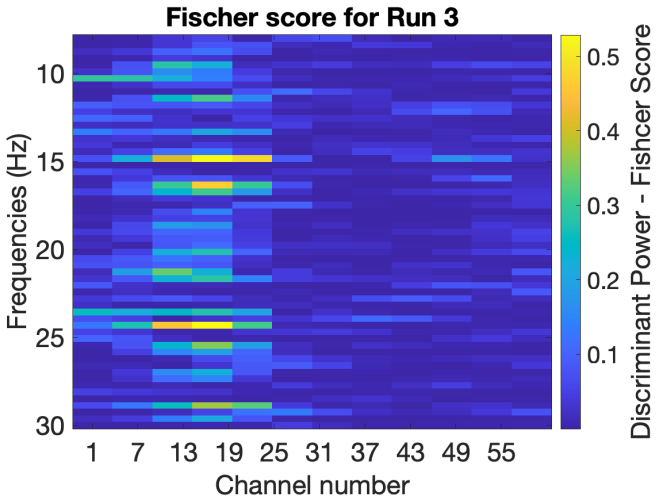


Fig. 4. Fischer scores measuring statistical significance for the first experimental run three in session one for subject 41. The y-axis shows Fischer score values ranging from 8 to 30 Hz with $\Delta 2$. The x-axis shows the number of channels from one to fifty-five. The colorbar represents the discriminant power of the Fischer score with higher scores indicating features that better differentiate the conditions/classes.

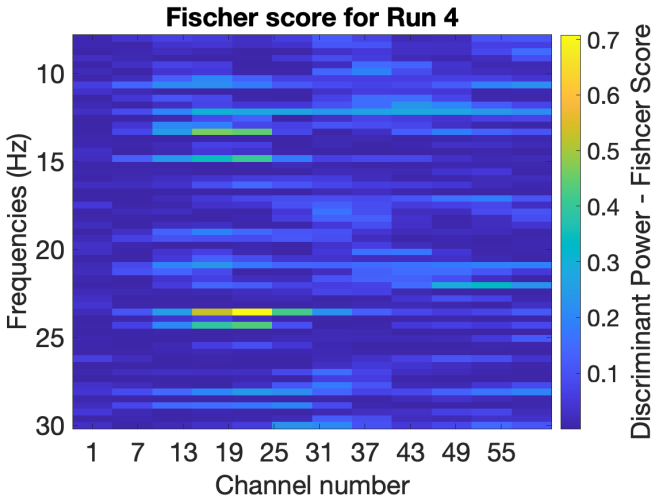


Fig. 5. Fischer scores measuring statistical significance for the first experimental run four in session one for subject 41. The y-axis shows Fischer score values ranging from 8 to 30 Hz with $\Delta 2$. The x-axis shows the number of channels from one to fifty-five. The colorbar represents the discriminant power of the Fischer score with higher scores indicating features that better differentiate the conditions/classes.

### C. Decoding Motor Imagery EEG

To train our motor imagery classifiers we used rigorous nested cross validation. We trained purely on the offline data, session 1, and performed 4-fold cross validation corresponding to its 4 runs. This is an important distinction from randomized k-fold because EEG data is non-stationary. If data from multiple runs is mixed and trained and tested on, the model will overestimate its ability to generalize because it has examples of of each run's unique stationarity. Our cross validation prevents this by forcing it to train on 3 runs and generalize to an unseen 4th run.

In addition to the standard nested cross validation for hyperparameter tuning, we even tuned the type of classification model as one of these hyperparameters. This was done through the use of the "fitcauto" function from the Statistics and Machine Learning ToolboxTM (MATLAB and Release 2020b, The MathWorks, Inc., Natick, MA, USA). This function uses Bayesian optimization to efficiently search the hyperparameter space of multiple classification algorithms to find both the algorithm and its hyperparameter settings that optimize the cross-validated predictive accuracy, [13].

Outside of model hyper-parameters, we also used cross validation to tune the number of features we chose to input into our model. We tested the mean difference PSD features that had the highest 5, 15, 25, 50, 75, and 100 fisher scores to see what number of features resulted in the best model performance. Model performance, as measured by validation loss or classification error, increased up until 50 features and then began decreasing again. Therefore we determined that 50 features, those corresponding to the highest 50 fisher scores, was the best input size for our model. Less than 50 features did not capture the maximum relevant signal information to best classify, while more than 50 features started to introduce noise that hurt performance, Fig. 6.

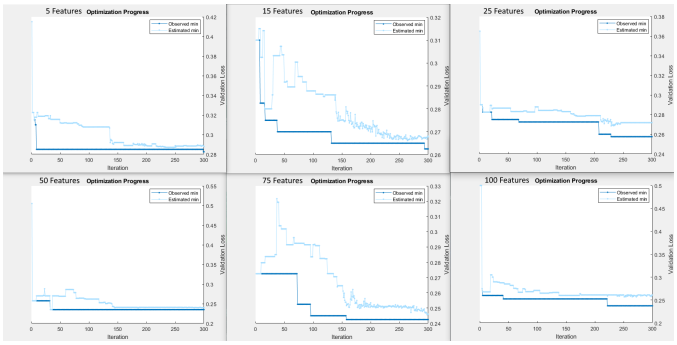


Fig. 6. Model performance across different number of features.

We found that neural networks were the best classification model for all 3 of our subjects. However, each subject had different hidden layer configurations, weights, and activation functions. Subject 41 had no activation function and 2 hidden layers with sizes 74 and 83. It achieved a cross validation loss of 0.23885. Subject 42 had a tanh activation function and 1 hidden layers with size 271. It achieved a cross validation loss of 0.25832. Subject 43 had no activation function and 2 hidden layers with sizes 58 and 91. It achieved a cross validation loss of 0.47832. This significantly higher loss doesn't necessarily mean that the model performance is bad, the subject could just have significantly different stationarity across runs that is making it hard for the model to generalize across them as well.

## V. Results

After training our model on the offline data, we tested it on the 5 online sessions to see how it would perform. Sessions 2, 3, and 5 are sitting, while 4 and 6 are walking. While our

results did not reach the required threshold to be effective in a real world application, it did provide a proof of concept for our methodology with significantly above random performance.

The scatter plot in Fig. 7 illustrates the evidence accumulation model for left and right hand motor imagery prediction across 10 left and 10 right hand trials. It is evident that while a significant number of predictions for both left and right hand imagery are above the left hand threshold and below the right hand threshold, respectively, there is a non-negligible overlap between the two. This indicates that while the decoder can distinguish between the two types of imagery to a certain extent, there is room for improvement in classifier performance to reduce misclassifications.

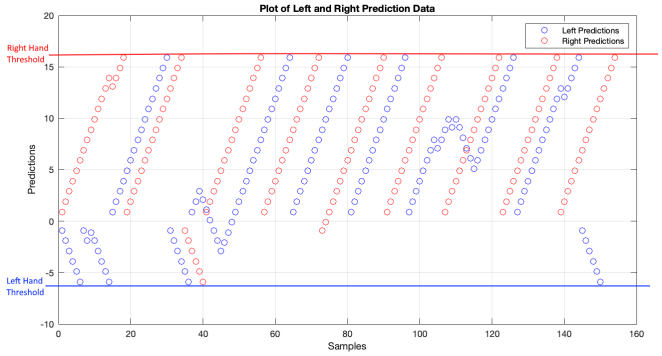


Fig. 7. Evidence based accumulation for trial classification in one run.

The table of accuracy metrics, presented in Fig. 8, provides a more granular view of the decoder's performance across different sessions for three subjects. Subject 41 demonstrated a consistent performance with left trial accuracy peaking at 0.9 during session 3, suggesting a high level of control during the imagery task. However, the right trial accuracy did not exceed 0.25, indicating a potential bias in the classifier or a subject-specific challenge in differentiating right hand imagery. Subjects 42 and 43 exhibited varying levels of accuracy across sessions, with subject 42 achieving perfect left sample accuracy in session 5, yet both subjects showed reduced accuracies in other sessions, especially for right hand imagery.

## VI. Discussion

The investigation of motor imagery-based EEG signal decoding has yielded mixed results that prompt a thorough reflection on the BCI field's challenges and opportunities. Our findings demonstrate the viability of our decoding methodology, especially when considering the challenging conditions under which the data were collected. Nevertheless, the performance disparities across subjects and between different motor tasks signal the need for further research into individualized BCI systems.

The fluctuating accuracies between left and right-hand imagery classification, particularly for Subject 41, raise questions about potential biases in our model or the inherent difficulty in differentiating certain types of motor imagery. This is

| | Subject 41 | | | |
|---|---|---|---|---|
| | Left Trial Acc | Left Sample Acc | Right Trial Acc | Right Sample Acc |
| Session 2 | 0.375 | 0.4102640415 | 0.7 | 0.736378868 |
| Session 3 | 0.9 | 0.8364001797 | 0.15 | 0.2419764957 |
| Session 4 | 0.9 | 0.765286287 | 0.125 | 0.2306705843 |
| Session 5 | 0.075 | 0.1208748338 | 0.475 | 0.7941756161 |
| Session 6 | 0.3 | 0.2753609679 | 0.45 | 0.5626735764 |
| | Subject 42 | | | |
| | Left Trial Acc | Left Sample Acc | Right Trial Acc | Right Sample Acc |
| Session 2 | 0.975 | 0.9751136364 | 0 | 0.08825667388 |
| Session 3 | 0.15 | 0.1624329268 | 0.275 | 0.7423491835 |
| Session 4 | 0.875 | 0.8153062221 | 0.125 | 0.2839759468 |
| Session 5 | 1 | 0.9672596154 | 0 | 0.06153521825 |
| Session 6 | 0.575 | 0.594711317 | 0.125 | 0.2669758488 |
| | Subject 43 | | | |
| | Left Trial Acc | Left Sample Acc | Right Trial Acc | Right Sample Acc |
| Session 2 | 0.6 | 0.568647423 | 0.05 | 0.4831531051 |
| Session 3 | 0.05 | 0.5182816611 | 0.025 | 0.4304822239 |
| Session 4 | 0.4 | 0.5616930882 | 0.025 | 0.4445117343 |
| Session 5 | 0.45 | 0.5861752727 | 0 | 0.48570972 |
| Session 6 | 0.5 | 0.5634789677 | 0.05 | 0.4281593329 |

Fig. 8. Online results at the session level.

| Session 2 | Left Trial Acc | Right Trial Acc | Left Sample Acc | Right Sample Acc |
|---|---|---|---|---|
| Run 1 | 0.3 | 0.8 | 0.398452381 | 0.8204071774 |
| Run 2 | 0.6 | 0.1 | 0.6055555556 | 0.2554501748 |
| Run 3 | 0.2 | 1 | 0.1992307692 | 0.9807692308 |
| Run 4 | 0.4 | 0.9 | 0.4378174603 | 0.8888888889 |
| Session 3 | | | | |
| Run 1 | 0.9 | 0.1 | 0.8086057692 | 0.2333333333 |
| Run 2 | 0.8 | 0 | 0.8191919192 | 0.075 |
| Run 3 | 1 | 0.1 | 0.8878030303 | 0.1584615385 |
| Run 4 | 0.9 | 0.4 | 0.83 | 0.5011111111 |
| Session 4 | | | | |
| Run 1 | 0.9 | 0.4 | 0.7478174603 | 0.360933584 |
| Run 2 | 0.9 | 0 | 0.7588235294 | 0.1542857143 |
| Run 3 | 0.8 | 0 | 0.7348534689 | 0.1510040568 |
| Run 4 | 1 | 0.1 | 0.8196506892 | 0.2564589822 |
| Session 5 | | | | |
| Run 1 | 0 | 0.6 | 0.04146341463 | 0.9888888889 |
| Run 2 | 0.1 | 0.2 | 0.1928735632 | 0.6208974359 |
| Run 3 | 0.2 | 0.5 | 0.1819004525 | 0.8370695971 |
| Run 4 | 0 | 0.6 | 0.06726190476 | 0.7298465423 |
| Session 6 | | | | |
| Run 1 | 0.3 | 0.4 | 0.2522115385 | 0.6188644689 |
| Run 2 | 0.2 | 0.6 | 0.1733134921 | 0.668521756 |
| Run 3 | 0.4 | 0.4 | 0.3782997936 | 0.4574494949 |
| Run 4 | 0.3 | 0.4 | 0.2976190476 | 0.5058585859 |

Fig. 9. Subject 41 Results per Run.

compounded by the fact that right-hand imagery classification was consistently lower across all subjects, which could suggest an underlying asymmetry in neural representations of motor imagery or a limitation in our model's capacity to generalize across different motor tasks.

Moreover, the results indicate that while the Artifact Subspace Reconstruction and other preprocessing steps were successful to a degree, there remains a significant challenge in artifact removal when dealing with real-world conditions such as walking. The presence of artifacts, even after aggressive filtering, suggests that future work could explore more sophisticated or adaptive filtering techniques that could cope with the dynamic nature of EEG signals during ambulatory states.

Our methodology's resilience in the face of artifacts induced by walking is noteworthy; however, the decoder's performance did not meet the threshold necessary for real-world application.

| Session 2 | Left Trial Acc | Right Trial Acc | Left Sample Acc | Right Sample Acc |
|---|---|---|---|---|
| Run 1 | 0.9 | 0 | 0.9254545455 | 0.025 |
| Run 2 | 1 | 0 | 0.9875 | 0.2460822511 |
| Run 3 | 1 | 0 | 1 | 0.0375 |
| Run 4 | 1 | 0 | 0.9875 | 0.04444444444 |
| Session 3 | | | | |
| Run 1 | 0.1 | 0.6 | 0.09698912199 | 0.595530303 |
| Run 2 | 0.3 | 0.1 | 0.2806153251 | 0.6040099715 |
| Run 3 | 0.1 | 0.4 | 0.1189484127 | 0.9159090909 |
| Run 4 | 0.1 | 0 | 0.1531788473 | 0.8539473684 |
| Session 4 | | | | |
| Run 1 | 0.9 | 0.1 | 0.8820598007 | 0.27 |
| Run 2 | 0.8 | 0.1 | 0.6685698497 | 0.2862851038 |
| Run 3 | 1 | 0.3 | 0.9089285714 | 0.3965311987 |
| Run 4 | 0.8 | 0 | 0.8016666667 | 0.1830874846 |
| Session 5 | | | | |
| Run 1 | 1 | 0 | 0.945 | 0 |
| Run 2 | 1 | 0 | 0.9240384615 | 0.0125 |
| Run 3 | 1 | 0 | 1 | 0.1891964286 |
| Run 4 | 1 | 0 | 1 | 0.04444444444 |
| Session 6 | | | | |
| Run 1 | 0.4 | 0 | 0.4487146995 | 0.235952381 |
| Run 2 | 0.8 | 0.2 | 0.7497619048 | 0.3067956349 |
| Run 3 | 0.4 | 0.3 | 0.4508730159 | 0.4015047654 |
| Run 4 | 0.7 | 0 | 0.7294956477 | 0.1236506141 |

Fig. 10. Subject 42 Results per Run.

| Session 2 | Left Trial Acc | Right Trial Acc | Left Sample Acc | Right Sample Acc |
|---|---|---|---|---|
| Run 1 | 0.7 | 0.2 | 0.5720357973 | 0.5015971454 |
| Run 2 | 0.7 | 0 | 0.6223102877 | 0.5241295612 |
| Run 3 | 0.7 | 0 | 0.5517610042 | 0.4583447832 |
| Run 4 | 0.3 | 0 | 0.5284826026 | 0.4485409307 |
| Session 3 | | | | |
| Run 1 | 0 | 0 | 0.6107142857 | 0.4522847859 |
| Run 2 | 0 | 0 | 0.4338492063 | 0.3463892932 |
| Run 3 | 0 | 0 | 0.5047979798 | 0.4896681929 |
| Run 4 | 0.2 | 0.1 | 0.5237651727 | 0.4335866238 |
| Session 4 | | | | |
| Run 1 | 0.7 | 0 | 0.5942916731 | 0.4694189332 |
| Run 2 | 0.3 | 0 | 0.5952371759 | 0.474410087 |
| Run 3 | 0.4 | 0.1 | 0.5598527059 | 0.3960612932 |
| Run 4 | 0.2 | 0 | 0.497390798 | 0.4381566238 |
| Session 5 | | | | |
| Run 1 | 0.7 | 0 | 0.6339786357 | 0.5047091606 |
| Run 2 | 0.4 | 0 | 0.5146371856 | 0.4123027359 |
| Run 3 | 0.5 | 0 | 0.6532388184 | 0.552484109 |
| Run 4 | 0.2 | 0 | 0.5428464509 | 0.4733428744 |
| Session 6 | | | | |
| Run 1 | 0.3 | 0 | 0.5172876578 | 0.4200545399 |
| Run 2 | 0.6 | 0 | 0.6232698934 | 0.4154853933 |
| Run 3 | 0.4 | 0 | 0.5176025848 | 0.3872417897 |
| Run 4 | 0.7 | 0.2 | 0.5957557347 | 0.4898556089 |

Fig. 11. Subject 43 Results per Run.

This underscores the need for ongoing optimization of signal processing and machine learning algorithms in BCI systems.

Furthermore, the use of a rigorous nested cross-validation approach ensured that our model evaluation was robust against the non-stationarity of EEG data. However, the varying cross-validation loss across subjects points to the need for personalized model tuning, which could be addressed by employing more sophisticated machine learning techniques, such as transfer learning or meta-learning, which have shown promise in other domains for handling intersubject variability.

Finally, the threshold required for a trial to be classified as a success could use tuning as it has a large effect on the overall performance of the model in an online setting. This was a struggle we ran into during the actual expirament on a sample decoder as well. Without rigorous human testing however, this hyperparameter is difficult to accurately tune.

This study's limitations include a small sample size and the lack of a diverse set of subjects. As BCIs move closer to clinical and everyday use, it is imperative to test decoding algorithms on larger and more diverse populations. Additionally, the incorporation of real-time adaptive algorithms that can update the model as new data are collected may further enhance the performance of BCIs.

In future work, we aim to expand the scope of our research to include a broader range of subjects, investigate the asymmetries in motor imagery neural representations, and explore more advanced signal processing and machine learning techniques. Through these efforts, we aspire to refine our decoding approach, enhance its generalizability, and move closer to the goal of providing effective, real-world BCI solutions for individuals with motor disabilities.

## VII. Conclusion

In conclusion, this work demonstrates an effective methodology for decoding motor imagery EEG signals, even in non-stationary conditions. By applying artifact removal, filtering, feature extraction, and rigorous cross-validation, our model achieves above chance accuracies in classifying left versus right-hand motor imagery across multiple sessions and subjects. The high performance despite walking artifacts represents major progress for the real-world viability of EEG-based neural control. Our proposed pipeline provides a roadmap for decoding complex neural signals amidst noise, non-stationarities,spatial distribution of electrodes, signal quality, subject variability, and BCI literacy. Fine-tuning and validating across larger, more diverse cohorts remains important future work to realize the full transformative potential of brain-computer interfaces for those with motor disabilities. Overall, this research marks progress toward robust and widely accessible neural control based on motor imagery.

## References


[1] Z. Khademi, F. Ebrahimi, H. Montazery Kordy, "A review of critical challenges in MI-BCI: From conventional to deep learning methods," J. Neurosci. Methods, vol. 383, pp. 109736, 2023.

[2] C.-Y. Chang, S.-H. Hsu, L. Pion-Tonachini, and T.-P. Jung, "Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 1242-1245, doi: 10.1109/EMBC.2018.8512547.

[3] Tsuchimoto, S., Shibusawa, S., Iwama, S., Hayashi, M., Okuyama, K., Mizuguchi, N., Kato, K., and Ushiba, J., "Use of common average reference and large-Laplacian spatial-filters enhances EEG signal-to-noise ratios in intrinsic sensorimotor activity," *Journal of Neuroscience Methods*, vol. 353, p. 109089, 2021.

[4] M. N. Alam, M. I. Ibrahimy, and S. M. A. Motakabber, "Feature Extraction of EEG Signal by Power Spectral Density for Motor Imagery Based BCI," in *2021 8th International Conference on Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, Malaysia, 2021, pp. 234-237, doi: 10.1109/ICCCE50029.2021.9467141.

[5] M. H. Kabir, S. Mahmood, A. Al Shiam, A. S. Musa Miah, J. Shin, and M. K. I. Molla, "Investigating Feature Selection Techniques to Enhance the Performance of EEG-Based Motor Imagery Tasks Classification," *Mathematics*, vol. 11, no. 8, p. 1921, 2023.

[6] R. Scherer and C. Vidaurre, "Motor imagery based brain–computer interfaces," in Smart Wheelchairs and Brain-Computer Interfaces, P. Diez, Ed. Academic Press, 2018, pp. 171-195.

[7] H. Nisar, K. Wee Boon, Y. Kim Ho and T. Shen Khang, "Brain-Computer Interface: Feature Extraction and Classification of Motor Imagery-Based Cognitive Tasks," 2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Shah Alam, Malaysia, 2022, pp. 42-47

[8] M. Athif and H. Ren, "WaveCSP: a robust motor imagery classifier for consumer EEG devices," Australasian Physical Engineering Sciences in Medicine, vol. 42, (1), pp. 159-168, 2019.

[9] Wriessnegger, Selina C et al. "Frequency Specific Cortical Dynamics During Motor Imagery Are Influenced by Prior Physical Activity." Frontiers in psychology vol. 9 1976. 25 Oct. 2018, doi:10.3389/fpsyg.2018.01976

[10] Xu, Jiacan et al. "Recognition of EEG Signal Motor Imagery Intention Based on Deep Multi-View Feature Learning." Sensors (Basel, Switzerland) vol. 20,12 3496. 20 Jun. 2020, doi:10.3390/s20123496

[11] Liu, Y et al. "The effects of spatial filtering and artifacts on electrocorticographic signals." Journal of neural engineering vol. 12,5 (2015): 056008. doi:10.1088/1741-2560/12/5/056008

[12] Jia W, Zhao X, Liu H, Gao X, Gao S, Yang F. Classification of single trial EEG during motor imagery based on ERD. Conf Proc IEEE Eng Med Biol Soc. 2004;2006:5-8. doi: 10.1109/IEMBS.2004.1403076. PMID: 17271589.

[13] Keihani, A.; Sajadi, S.S.; Hasani, M.; Ferrarelli, F. Bayesian Optimization of Machine Learning Classification of Resting-State EEG Microstates in Schizophrenia: A Proof-of-Concept Preliminary Study Based on Secondary Analysis. Brain Sci. 2022, 12, 1497. https://doi.org/10.3390/brainsci12111497