

---

# Two Class Motor Imagery: Comparison of Recent Approaches

---

**George Hanna**

Institute of Biomaterials and Biomedical Engineering  
University of Toronto  
Toronto, ON  
georgesk.hanna@mail.utoronto.ca

**Behnam Reyhani**

Institute of Biomaterials and Biomedical Engineering  
University of Toronto  
behnam.reyhani@mail.utoronto.ca

## Abstract

Brain computer interfaces (BCI) are systems that decipher an individual's intents by analyzing their brain signals. These systems have shown the potential to be a solution for improving independent mobility in paralyzed individuals. Under the motor imagery paradigm, an individual imagines movement of a limb. The system subsequently decodes this signal and provides a class prediction. Most previous work has sought to improve feature extraction and selection methods but has relied on linear classifiers for class prediction. In this work, we examine the use of ensemble learning methods to improve classification accuracies in a BCI paradigm based on 2-class motor imagery. Our results demonstrate that majority voting schemes outperform Linear Discriminant Analysis (LDA) but that other ensemble techniques such as boosting and bagging perform worse. In the future, we aim to employ more robust feature extraction methods such as Filter-bank Common Spatial Patterns to improve our system's generalizability.

## 1 Introduction

Brain-Computer Interfaces (BCIs) are systems which translate an individual's intent into computerized commands. These systems offer individuals with severe motor impairments, such as those suffering from paralysis, a means for environmental control (e.g. TV, light, computer, wheelchair control)[8] [13]. A BCI system is composed of several sub-components as depicted in Figure 1.

Several types of inputs can be used as inputs to the system. Electroencephalography (EEG) is the most widely used modality during the signal acquisition phase as it is practical, non-invasive, and easy-to use [10]. EEG signals are recorded using an electrode cap according to standardized placements (10-20 system). Several types of control signals can be inferred from the raw data to allow for the decoding of user intent. One such type of signals are sensorimotor rhythms (SMRs). SMRs are modulation in the signals of the sensorimotor cortex of the brain; they arise as a result of actual movement, motor intent or motor imagery (MI). This causal relation between movement-related brain processes and SMRs has allowed motor imagery-based BCIs to be used for environmental control [19].

Under the motor imagery paradigm, the user imagines a specific limb's movement. This results in alterations in the rhythmic activities overlying that limb's location in the user's sensorimotor cortex

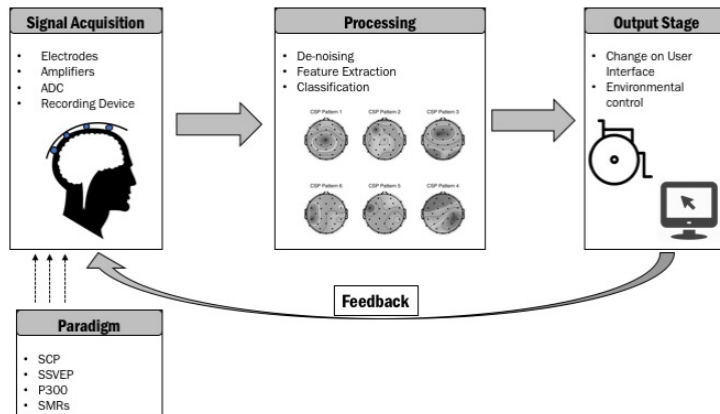


Figure 1: BCI Subcomponent Overview. Signal acquisition is done using an electrode cap according to a standardized system. Data then undergoes de-noising, feature extraction and output prediction. Several BCI paradigms are possible, each of which depend on a different control signal. SCP are slow cortical potentials, SSVEP are steady state visually evoked potentials, P300 are evoked resting potentials and SMRs are sensory motor rhythms.

[17]. Once recorded, the raw EEG data undergoes a feature extraction process. The features are then passed to a classification algorithm which is trained with the extracted features to discern between the different motor intents of the user. These intents can therefore serve as different control signals during environmental control tasks (e.g. right-hand movement corresponds to moving a computer cursor to the right).

Traditionally, MI classification accuracy has relied on extensive user training to maximize the power of the measured signals, and subsequently, the separability of the different classes. Users were instructed to use visual feedback signals to reinforce their imagined movements [3]. This placed the onus on users to adapt and hampered widespread use of the systems. Recent developments in Machine Learning techniques now place the task of improving classification accuracy on the machine side.

Several challenges still face BCI algorithms and are subject to much work. First and foremost, EEG-based BCI systems are non-stationary by nature, have limited signal-to-noise ratios and suffer from poor spatial resolutions due to volume conductor effects. These effects are worsened by varying electrode impedances, muscular activity, eye movements and changes in user mental states)[3][5]. Additionally, the datasets are highly dimensional, with the number of channels either equalling or exceeding the number of measured trials[8].

In this project, we seek to improve 2-class motor imagery classification accuracy using ensemble learning methods. Figure 2 depicts the typical data processing pipeline we will follow. More specifically, we outline the following aims: examining the use of boosting, bagging and majority voting schemes with popular BCI classification algorithms.

## 2 Problem Domain

### 2.1 Overview

Under the formulation of this classification problem,  $X \in \mathbb{R}^{C \times T}$  will correspond to a short segment of an EEG signal which, in turn, pertains to a motor imagery trial. In this case, C represents the number of data acquisition channels and T represents the sampled time points [5]. For correct prediction of the output, the raw signal must pass through several stages before classification. Figure 2 depicts the pipeline describing the data processing method followed in this work. The method employed follows the data processing pipeline based on the Common Spatial Patterns (CSP) algorithm [5].

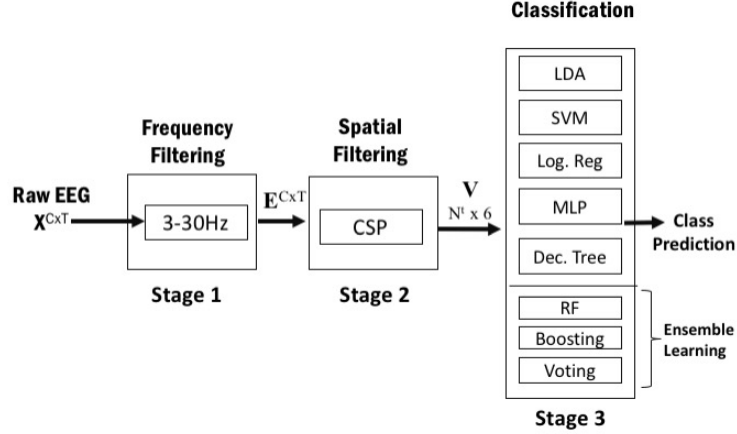


Figure 2: CSP based data processing pipeline employed in this work.

## 2.2 Stage 1: Frequency Filtering Stage

The raw data,  $\mathbf{X} \in \mathbb{R}^{C \times T}$ , is first passed through a frequency filtering stage using a zero-phase Chebyshev Type II filter. The filter covers the frequency range between 3 and 30 Hz (most pertinent for SMRs). This stage yields a frequency filtered matrix, denoted  $\mathbf{E} \in \mathbb{R}^{C \times T}$  and centers the analysis about two frequency peaks relevant for SMRs: around 10 Hz and 20 Hz [8].

## 2.3 Stage 2: Spatial Filtering Stage

Due to volume conducting effects through the skull, scalp and brain layers, EEG signals undergo spatial smearing before acquisition [9]. This renders the task of separating classes more difficult. In this case, spatial filtering algorithms can be used to compute a projection matrix  $\mathbf{W}$ , which is  $C \times C$ , such that  $\mathbf{W}^T \mathbf{E} = \mathbf{Z}$ , which has reduced smearing [5][9]. The Common Spatial Patterns (CSP) algorithm has been successfully used for spatial filtering in motor imagery classification [3][5]. It designs spatial filters  $w_j$  (the columns of projection matrix  $\mathbf{W}$ ) which maximize the variances of the filtered time series and render the classes optimally separable. This is done by solving the following eigenvalue decomposition problem:

$$\Sigma_1 \mathbf{W} = (\Sigma_1 + \Sigma_2) \mathbf{W} \mathbf{D} \quad (1)$$

where  $\Sigma_1$  and  $\Sigma_2$  are estimates of the covariance matrices for each of the classes.  $\mathbf{D}$  is a diagonal matrix containing the eigenvalues of  $\Sigma_1$ .  $\mathbf{W}$  is sorted in descending order of the eigenvalues. The first 3 and last 3 filters are kept (i.e. the most important filter pair is the first, and last entry in the matrix) [11]. This results in a projection matrix that is  $C \times 6$ . The  $i^{th}$  trial in the spatially filtered signal  $\mathbf{V} \in \mathbb{R}^{N^{trials} \times 6}$  can then be computed using the new projection matrix :

$$V_i = \log \frac{\text{diag}(\mathbf{W}^T \mathbf{E}_i \mathbf{E}_i^T \mathbf{W})}{\text{tr}[\mathbf{W}^T \mathbf{E}_i \mathbf{E}_i^T \mathbf{W}]} \quad (2)$$

## 2.4 Stage 3: Classification Stage

For a binary classification task, a classifier, denoted by  $f(\mathbf{X}; \mathbf{W}, \mathbf{B})$ , is expected to output a real value, whose sign is interpreted as the predicted class (see eqn 1).

$$f(\mathbf{X}; \mathbf{W}, \beta) = \sum_{j=1}^J \beta_j \log(w_j^T \mathbf{X} \mathbf{X}^T w_j) + \beta_0 \quad (3)$$

The frequency-filtered raw data  $\mathbf{X}$  is first projected by  $j$  filters (columns of the projection matrix  $\mathbf{W}$ ). The log of the power of the projected signal is then taken and is combined with  $J$  dimensional features and a bias to produce a prediction [5].

In this work, five classification algorithms which have been relatively successful in motor imagery classification tasks are examined and compared to ensemble learning methods. Linear Discriminant Analysis (LDA) has been extensively used in BCI classification tasks. It maximizes between-class separation and minimizes within-class distances; it does so by finding a set of weights and a bias which make up the best discriminating projection [8]. The sign of the resulting output is taken as the class prediction. Support Vector Machine (SVM) classifiers calculate a decision hyperplane which maximizes the margins between the classes. The decision tree algorithm uses the extracted feature set to generate a decision-making scheme with branches and nodes. At each level of the tree, either one or several features are considered such that the information gain of class labels in the partition is maximized. A Multilayer perceptron is an assembly of several layers of neurons (input layers, one or many hidden layers and an output layer), which together produce a nonlinear decision boundary [8]. Finally, logistic regression computes the probability of a default class by linearly combining the inputs and passing the result through a sigmoid function to produce binary values.

#### 2.4.1 Ensemble Learning Methods

Ensemble learning consists of combining the decisions of many classifiers to improve the overall accuracy. Boosting is an ensemble learning method which aims to reduce bias by training classifiers in a cascade-like manner. In this sense, each new classifier focuses on the mistakes of its predecessor [8]. Voting is another scheme which simply consists of taking the majority vote between several classifiers trained on the original dataset. Bagging (Bootstrap-aggregation) is the third and final method employed in this work. It reduces overfitting by averaging predictions over subsets of the dataset that are randomly generated with replacement. Random Forests are based on the bagging algorithm. They use decision trees with splitting on random subsets of features to produce predictions [2].

### 3 Related Work

Extensive bodies of work have been devoted to improving motor imagery classification accuracies. In terms of feature extraction, popular methods include statistical spectrum estimation methods (both parametric and non-parametric)[7], Adaptive autoregressive (AAR) parameter estimation with the recursive least squares algorithm [12] and finally, and spatial filtering techniques [5]. Based on recent classification performances in the BCI competition IV, methods based on spatial filtering techniques seem to demonstrate superior classification accuracies [16].

Independent component analysis is a widely used unsupervised spatial filtering technique which separates multichannel EEG into statistically independent components. Naeem et al. compared three well-known ICA-based algorithms (FastICA, SOBI and Infomax) to Common Spatial Patterns (CSP) filtering. They found that CSP yielded better results [9]. This is unsurprising given that CSP is supervised algorithm. CSP also performed better in session-to-session testing.

Several variants on the CSP algorithm have also been proposed. Samek et al. provide a comparison of 7 variants and demonstrate that stationary-CSP marginally outperforms the rest [13]. Filter-bank CSP (FBCSP) is another variant on CSP and has been relatively successful during BCI competitions. Ang et al. applied FBCSP on a 4-class motor imagery dataset in BCI competition III. They compared the performance of several classifiers and concluded that FBCSP with Naïve Bayes Parzen windows yielded the highest test accuracies [1].

In terms of classification algorithms, many different procedures have been explored in literature. Linear Discriminant Analysis (LDA) and Support Vector Machines (SVMs) have been widely used with relative success [8]. Alois Schlogl et al. made use of adaptive autoregressive process features and SVMs to classify 4-motor imagery tasks as part of the 2005 BCI competition IIIa. They obtained better classification accuracy and high kappa values using SVM relative to LDA [14]. Others have examined the use of Hidden Markov Models (HMMs), Multilayer Perceptron (MLP) and K-Nearest Neighbours [8].

Methods based on Ensemble Learning such as Random Forest (RF) have been occasionally examined. Bentlemsan et al. used FBCSP with the RF technique. They ran the algorithm on the Graz data set B of the BCI competition (a two-class motor imagery dataset with significant EOG contamination). Their algorithm outperformed most previous classification approaches [2].

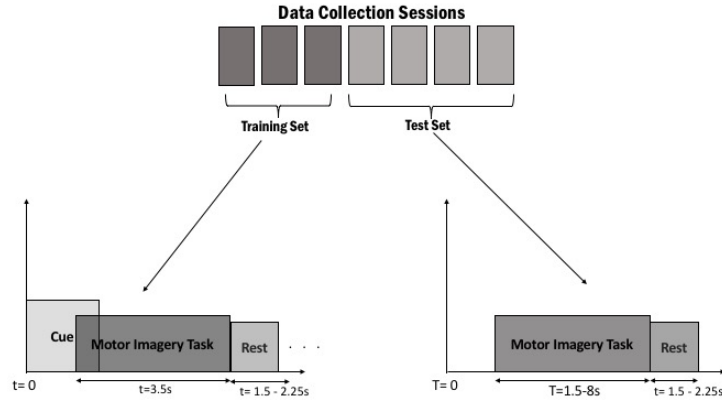


Figure 3: BCI Competition III Dataset IVb Experiment Setup

In our work, we aim to make further use of ensemble learning techniques such as voting, bagging and boosting to further improve classification accuracies.

## 4 Model Comparisons

### 4.1 Dataset Overview

The dataset examined in this work is provided as part of the BCI competition III in 2003, dataset IVb “Motor imagery, Uncued Classifier Application”. The particular problem which it attempts to address is that which involves classifying motor imagery in an asynchronous protocol design. In other words, no cues are provided along with the data to indicate that the subject has switched to a different mental target class [4]. As a byproduct of this paradigm, the dataset contains periods in which the user has no control intention. The algorithm is therefore expected to discern between left and right motor imagery as well as the resting state.

EEG data was recorded from one subject, sitting in a comfortable chair with arm rests. It is split into 7 sessions. The first 3 form the training set, and contain cues indicating for 3.5 seconds which of the 3 motor imageries the subject should perform. The remaining 4 sessions do not contain cues and have active periods of motor imagery varying between 1.5 to 8 seconds, intermitted by resting periods of 1.75 to 2.25 seconds. The data was recorded using 118 channels and consists 120 trials of motor imagery. Signals were band-pass filtered between 0.05 and 200 Hz and then downsampled at 100Hz before being publicly available for analysis. The experimental setup is depicted in Figure 3.

### 4.2 Experimental Procedure

As mentioned earlier, LDA is one of the most widely used classification algorithms applied in BCI research. Thus, an LDA classifier is used as the baseline model in our study and its classification performance on the test set is the metric used to assess the different models. The experimental procedure consists of training and testing a wide variety of classification algorithms on the BCI competition III dataset. More specifically, six standard classifiers (i.e. LDA, Linear SVM, RBF-Kernel SVM, Artificial Neural-Network, Decision Tree and Logistic Regression) and three ensemble classifiers (i.e. Majority voting, Random Forest, Logistic Regression Boosted) are trained and have their respective hyper-parameters tuned using 5-fold cross validation. This cross validation scheme is adopted over 10-fold cross validation to reduce training time as the number of classifiers being used in the experiment are high. As mentioned earlier, the classifiers are trained using a CSP feature extraction algorithm (mentioned in Section 2), which results in a six-dimensional feature vector for each data point. In addition, the training dataset is comprised of 210 labeled instances that is split evenly between the two classes.

The test dataset contains 34,594 data instances, with 18,488 belonging to one class and 16,106 to the other. As mentioned earlier, a six-dimensional feature vector is extracted for each test instance.

Table 1: Classification test results

Classifier	Test Error	Cohen's Kappa
LDA	13.02%	0.737
Linear SVM	11.64%	0.764
RBF-Kernel SVM	11.13%	0.775
Neural Network	11.45%	0.769
Decision Tree	16.53%	0.671
Logistic Regression	12.98%	0.739
Majority Vote	10.81%	0.782
Random Forest	14.05%	0.713
Logit Boost	16.38%	0.681

Since, the number of data instances belonging to each class is different (roughly 2,000 more positive examples than negative examples), a second performance metric (other than test error) must be used; a metric that is insensitive to randomness caused by different number of examples in each class. Hence, Cohen's kappa value is also used. Using these metrics, three performance comparisons are made: 1) Baseline vs. Other Standard Classifiers, 2) Ensemble Classifiers vs. Standard Classifiers, 3) Ensemble Classifier vs. Respective Standard Classifier (i.e. Random Forest vs. Decision Tree, Logistic Regression Boost vs. Logistic Regression).

### 4.3 Results and Discussion

The results of the experiments are summarized in Table 1. Three classifiers (Linear SVM, RBF-Kernel SVM and Neural Network) performed better than LDA, one performed similarly (Logistic Regression) and one performed worse (Decision Tree). The application of a linear maximum margin classifier (i.e. Linear SVM) improved the classification performance by roughly 1.5%. Also, given that the training dataset has low feature dimension (6 dimensions) but a much greater amount of training instances (210); it is no surprise that the non-linear and more complex models, the RBF-Kernel based SVM and artificial neural network performed better than the LDA. In fact, the RBF-Kernel based SVM was the best performing standard classifier with a test error of 11.13% and kappa value of 0.775. In contrast, the Decision Tree classifier performed substantially worse with a test error of 16.53%, which may be a result of overfitting due to substantial tree size.

In terms of ensemble classifiers, three different models (i.e. Majority Vote, Random Forest and Logit Boost) were used for producing predictions on the test dataset. The Majority Vote classifier produced a test error of 10.81% and kappa value of 0.782, making it the best performing model of the study. In this classifier, predictions are made by aggregating the predictions made by the five standard classifiers (i.e. Linear SVM, RBF SVM, Neural Network, Decision Tree and Logistic Regression) and then choosing the class that was most frequently predicted. The other two ensemble classifiers, Random Forest and Logit Boost, performed worse than the baseline. The poor performance of Logit Boost is hypothesized to be a result of the increase in variance (leading to overfitting) that can occur when using boosting ensembles. The overfitting which has led to reduced performance is also evident when comparing the performance of Logit Boost with its underlying model, logistic regression (3.5% increase in test error). Unlike Logit Boost, the Random Forest model performed better than its underlying model producing about a 2.5% increase in performance compared to the Decision Tree. It is noted that the poor performance of the Decision Tree may be a hint of overfitting. Thus, it was expected that its "Bagged" ensemble version, which is known for reducing overfitting, has resulted in an increase in performance. Therefore, it seems that inadequate performance of the Random Forest model when compared to the baseline is due to the below-par performance of the underlying Decision Tree model.

## 5 Model Limitations

The limitations discussed will pertain to the best performing classification model in the study. This model consisted of CSP feature extraction followed by classification using the Majority Vote ensemble model. Two types of limitations exist: one as a result of the non-stationary and noisy characteristics of the human brain signals, and the other due to the inherit characteristics of an ensemble model.

As a consequence of head and ocular movement, the EEG signals collected typically suffer from poor signal-to-noise ratio. The EEG dataset provided in this study did not have electro-oculogram(EOG) readings nor did it have expert artifact labeling of the signals. This limited our capacity for data de-noising prior to classification. In more practical BCI settings, this will not be the case. As such, we expect our algorithm to generalize poorly to other datasets. Many different autonomous artifact removal algorithms have been proposed in the literature [6]. Thus, to address this limitation a pre-processing step before feature extraction is required to remove the potential artifacts.

The other BCI related limitation exists due to the inherit non-stationary nature of the EEG signals. More specifically, an individual's EEG signals rapidly vary over time and more so across different sessions. Thus, a classification pipeline that was trained using EEG signals from an earlier session may no longer be applicable for BCI classification in a future session. Several solutions have been proposed as a remedy to this issue, including adaptive classifiers and pre-session tuning of classifiers [15].

The final limitation arises due to the structure of the Majority Vote ensemble classifier. As mentioned earlier, this model is comprised of five different underlying classifiers, thus, training this model involves training the five underlying models. In addition, making predictions using the ensemble involves running the various classifiers and aggregating their predictions. In a BCI system, time constraints exist in both classifier training and prediction. This is especially the case if additional training is required as a remedy to the non-stationary limitations mentioned earlier. In addition, the EEG signals are being sampled at 1000 Hz, which corresponds to a new data point coming in to be classified every millisecond. Thus, the time it takes to train and test the model must be minimal. This can be an issue for the Majority Vote ensemble model as it is comprised of multiple classifiers. To overcome this limitation, the use of parallel computing and map-reduce have been proposed and shown to give adequate train and test time [18].

## 6 Conclusions

In this work, we examine the use of 5 classifiers and 3 ensemble learning schemes to predict outputs for a 2-class motor imagery classification problem. We demonstrate that ensemble models such as Majority Voting are able to outperform traditional methods such as Linear Discriminant Analysis. Other ensemble learning techniques such as boosting and bagging performed worse and may require further tuning to improve their classification accuracies. In the future, we hope to employ more robust feature extraction methods such as FB-CSP and to handle data de-noising in our data processing pipeline. Additionally, we aim to test our model on different data sessions to assess its ability to generalize to other data.

## References

- [1] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience*, 6(MAR):1–9, 2012.
- [2] Maouia Bentlemsan, Et Tahir Zemouri, Djamel Bouchaffra, Bahia Yahya-Zoubir, and Karim Ferroudji. Random forest and filter bank common spatial patterns for EEG-based motor imagery classification. *Proceedings - International Conference on Intelligent Systems, Modelling and Simulation, ISMS*, 2015-September:235–238, 2015.
- [3] Benjamin Blankertz, Motoaki Kawanabe, Ryota Tomioka, Friederike Hohlefeld, Vadim Nikulin, and Klaus-Robert Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. *Advances in neural information processing systems*, pages 1–8, 2007.
- [4] Benjamin Blankertz, Klaus Robert Müller, Dean J. Krusienski, Gerwin Schalk, Jonathan R. Wolpaw, Alois Schlögl, Gert Pfurtscheller, José Del R. Millán, Michael Schröder, and Niels Birbaumer. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):153–159, 2006.

- [5] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus Robert Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.
- [6] Mehrdad Fatourech, Ali Bashashati, Rabab K Ward, and Gary E Birch. Emg and eog artifacts in brain computer interface systems: A survey. *Clinical neurophysiology*, 118(3):480–494, 2007.
- [7] Pawel Herman, Girijesh Prasad, Thomas Martin McGinnity, and Damien Coyle. Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(4):317–326, 2008.
- [8] Fabien Lotte, Marco Congedo, A Lécuyer, F Lamarche, and B Arnaldi. A review of classification algorithms for EEG-based brain – computer interfaces. *Journal of Neural Engineering*, 4:24, 2007.
- [9] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller. Seperability of four-class motor imagery data using independent components analysis. *Journal of Neural Engineering*, 3(3):208–216, 2006.
- [10] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *Sensors*, 12(2):1211–1279, 2012.
- [11] Boris Reuderink and Mannes Poel. Robustness of the Common Spatial Patterns algorithm in the BCI-pipeline. *CTIT Technical Report Series*, pages 3–7, 2008.
- [12] German Rodriguez-Bermudez, Pedro Garcia-Laencina, and Joaquin Roca. Efficient Automatic Selection and Combination of EEG Features in Least Squares Classifiers for Motor Imagery Brain–Computer Interfaces. *International Journal of Neural Systems*, 23(04):1350015, 2013.
- [13] Wojciech Samek, Carmen Vidaurre, Klaus-Robert Müller, and Motoaki Kawanabe. Stationary common spatial patterns for brain–computer interfacing. *Journal of Neural Engineering*, 9(2):026013, 2012.
- [14] Alois Schlogl, Lee Felix, Horst Bischof, and Gert Pfurtscheller. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *Journal of Neural Engineering*, 2:1–8, 2005.
- [15] Pradeep Shenoy, Matthias Krauledat, Benjamin Blankertz, Rajesh PN Rao, and Klaus-Robert Müller. Towards adaptive classification for bci. *Journal of neural engineering*, 3(1):R13, 2006.
- [16] Michael Tangermann, Klaus Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J. Miller, Gernot R. Müller-Putz, Guido Nolte, Gert Pfurtscheller, Hubert Preissl, Gerwin Schalk, Alois Schlögl, Carmen Vidaurre, Stephan Waldert, and Benjamin Blankertz. Review of the BCI competition IV. *Frontiers in Neuroscience*, 6(JULY):1–31, 2012.
- [17] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller. EEG-based communication: Improved accuracy by response verification. *IEEE Transactions on Rehabilitation Engineering*, 6(3):326–333, 1998.
- [18] Gongqing Wu, Haiguang Li, Xuegang Hu, Yuanjun Bi, Jing Zhang, and Xindong Wu. Mrec4. 5: C4. 5 ensemble classification with mapreduce. In *ChinaGrid Annual Conference, 2009. ChinaGrid'09. Fourth*, pages 249–255. IEEE, 2009.
- [19] Han Yuan and Bin He. Brain-Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives. *IEEE Trans Biomed Eng.*, 61(5):1425–1435, 2015.