

## Accepted Manuscript

Towards correlation-based time window selection method for motor imagery BCIs

Jiankui Feng, Erwei Yin, Jing Jin, Rami Saab, Ian Daly, Xingyu Wang, Dewen Hu, Andrzej Cichocki



PII: S0893-6080(18)30055-8  
DOI: <https://doi.org/10.1016/j.neunet.2018.02.011>  
Reference: NN 3902

To appear in: *Neural Networks*

Received date: 12 September 2017  
Revised date: 9 January 2018  
Accepted date: 14 February 2018

Please cite this article as: Feng, J., Yin, E., Jin, J., Saab, R., Daly, I., Wang, X., et al., Towards correlation-based time window selection method for motor imagery BCIs. *Neural Networks* (2018), <https://doi.org/10.1016/j.neunet.2018.02.011>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Towards correlation-based time window selection method for motor imagery BCIs

Jiankui Feng<sup>#1</sup>, Erwei Yin<sup>#2</sup>, Jing Jin<sup>\*1</sup>, Rami Saab<sup>1</sup>, Ian Daly<sup>3</sup>, Xingyu Wang<sup>1</sup>, Dewen Hu<sup>4</sup>, Andrzej Cichocki<sup>5,6,7</sup>

1 Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai, P. R. China

2 National Key Laboratory of Human Factors Engineering China Astronaut Research and Training Center, Beijing 100094, P. R. China

3 Brain-Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK

4 College of Mechatronic Engineering and Automation, National University of Defense Technology Changsha, Hunan 410073, P. R. China

5 Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Wako-shi, Japan,

6 Systems Research Institute PAS, Warsaw

7 Nicolaus Copernicus University (UMK), Torun, Poland

\*Corresponding author: E-mail: jinjingat@gmail.com, #equal contribution

**Abstract:** The start of the cue is often used to initiate the feature window used to control motor imagery (MI)-based brain-computer interface (BCI) systems. However, the time latency during an MI period varies between trials for each participant. Fixing the starting time point of MI features can lead to decreased system performance in MI-based BCI systems. To address this issue, we propose a novel correlation-based time window selection (CTWS) algorithm for MI-based BCIs. Specifically, the optimized reference signals for each class were selected based on correlation analysis and performance evaluation. Furthermore, the starting points of time windows for both training and testing samples were adjusted using correlation analysis. Finally, the feature extraction and classification algorithms were used to calculate the classification accuracy. With two datasets, the results demonstrate that the CTWS algorithm significantly improved the system performance when compared to directly using feature extraction approaches. Importantly, the average improvement in accuracy of the CTWS algorithm on the datasets of healthy participants and stroke patients was 16.72% and 5.24%, respectively when compared to traditional common spatial pattern (CSP) algorithm. In addition, the average accuracy increased 7.36% and 9.29%, respectively when the CTWS was used in conjunction with Sub-Alpha-Beta Log-Det Divergences (Sub-ABLD) algorithm. These findings suggest that the proposed CTWS algorithm holds promise as a general feature extraction approach for MI-based BCIs.

**Keywords:** Brain-computer interface; Correlation; Feature extraction; Time window selection; Common spatial pattern

## 1. Introduction

The aim of a brain-computer interface (BCI) is to provide a communication channel for patients who have lost normal communication abilities due to severe motor impairments (Wolpaw,

McFarland, Neat, & Forneris, 1991). A BCI system can transform brain activities into control commands (Daly, Nasuto, & Warwick, 2011; Wolpaw, & Wolpaw, 2012, Jin, Zhang, Daly, Wang, & Cichock, 2017) and has gained interest in neuroscience and rehabilitation engineering (Wolpaw, & Wolpaw, 2012; Dornhege, 2007; Jin, Sellers, Zhou, Zhang, Wang, Cichocki, 2015). Among BCI systems, the motor imagery (MI)-based BCI stands out for its advanced approach without the need for body movement and with only minimal requirements for auxiliary equipment (Yuan, & He, 2014). MI-based BCI systems utilize the brain activity associated with imagined motor movements as control commands for external devices (He, Baxter, Edelman, Cline, & Ye, 2015; Onose, Grozea, et al., 2012). These systems operate without external stimulus and thus are more easily used than stimuli-based BCIs (Wolpaw, & Wolpaw, 2012; Pan, Li, Gu, & Yu, 2013). Upon imagining movement, the rhythmic activities of the brain observed in the mu and beta rhythms are suppressed contralaterally. These phenomena are termed event-related desynchronization (ERD) and event-related synchronization (ERS), respectively (Pfurtscheller, & Da Sliva, 1999).

EEG data is characterized by its high-dimensionality, low signal to noise ratio, and susceptibility to outliers (Pfurtscheller, & Da Sliva, 1999; Thiyam, Cruces, & Olias, 2017; Jin, Allison, Sellers, Brunner, Horki, Wang & Neuper.). The dimensionality can be reduced by extracting those subspaces where features have highest discriminative power (Pfurtscheller, & Da Sliva, 1999; Thiyam, Cruces, & Olias, 2017). Common spatial pattern (CSP) (Fukunaga, 2013) is a method for extracting features to reduce the dimensionality, and is one of the most efficient algorithms applied to MI-based BCIs (Ramoser, Müller-Gerking, & Pfurtscheller, 2000). The CSP algorithm was first used to detect abnormalities in the EEG data (Koles, 1991) and introduced into BCI applications in 2000 (Ramoser, Müller-Gerking, & Pfurtscheller, 2000). Currently, the majority of state-of-the-art MI-BCIs use a fixed time window of EEG data to extract the MI features. However, the time latency during an MI period varies between trials for each participant, and it cannot be known with certainty when participants start to perform the MI task (Ang, Chin, Zhang, & Guan, 2012; Rodriguez-Bermudez, Garcia-Laencina, & Roca-Dorda, 2013). Some studies have extracted original EEG data 4-7s and 0-3.5s after the trial begins (Qiu, Jin, Lam, Zhang, Wang, & Cichocki, 2016) or 2.5-4.5s after the cue appears (He, Wei, Wang, & Zou, 2012). Some other studies did not extract the original EEG data after the cue appeared (Solis-Escalante, Müller-Putz, Brunner, Kaiser, & Pfurtscheller, 2010; Reinhold, Faller, et al., 2015). This approach commonly leads to low classification accuracy because of interference from invalid data.

In this study, we propose a novel correlation-based time window selection (CTWS) algorithm for MI-based BCIs. First, all the training MI samples of each class were averaged as a reference signal and updated based on correlation analysis and performance evaluation until the optimized reference signals for each class were found. Second, the optimized reference signals were employed to adjust the starting points of MI time windows for both training and testing samples, based on correlation analysis. Third, the feature extraction and classification algorithms were used for target detection, and the cross-validation method was performed to evaluate the average classification accuracy.

The remainder of this paper is organized as follows: Section 2 describes applied datasets and proposed methods; Section 3 shows the result of classification accuracy and distribution of features extracted; Section 4 presents the discussion; and Section 5 concludes the study.

## 2. Methods

## 2.1 Description of the datasets

Dataset 1 (BCI Computation IV Dataset I): The dataset contains 59 EEG channels with a sampling rate of 100 Hz, recorded from seven participants, including four healthy individuals and three artificially generated ‘participants’ (Blankertz, Dornhege, et al., 2007). For the purpose of the present study, only the calibration data (consisting of two runs totaling 200 trials) for each participant were used. In the experiment, the participants performed two-class motor imagery selected from the three classes left hand, right hand, and feet. As shown in Fig. 1 (a), each trial started from a visual cue pointing left, right, or down. The cue was displayed for a period of 4s, during which the participant was instructed to perform the cued motor imagery task. These periods were interleaved with 2s of blank screen and 2s with a fixation cross shown in the center of the screen. The fixation cross was superimposed on the cues, i.e. it was shown for 6s. The entire time length of the single trial was 8 s. More details about the dataset can be found on the following website: [http://www.bci.de/competition/iv/desc\\_1.html](http://www.bci.de/competition/iv/desc_1.html). The best window length for classification of the BCI Computation IV dataset 1 were found to be among 1s, 1.5 or 2s (Gouy-Pailler, Mattout, Congedo, & Jutten). In this study, we set the window length to 2s.

Dataset 2: This dataset was collected by ourselves from seven stroke patients. We acquired the EEG signals via the g.USBamp (Guger Technologies, Graz, Austria), and sampled at 256 Hz. Sixteen electrodes over the motor cortex (FC3, FCZ, FC4, C5, C3, C1, CZ, C2, C4, C6, CP3, CP1, CPZ, CP2, CP4, and PZ) were placed according to the international 10–20 system standard, and referenced to FCz and grounded to TP10. During the experiments, the patients were instructed to imagine moving either their left or right hand for 60 trials in total. As shown in Fig. 1 (b), each trial lasted eight seconds and started with a warning “beep” sound used to prompt the patient to be prepared. Two seconds later, a cue of MI task was displayed during which the patient was instructed to perform the cued motor imagery task. Six second later, a “relax” command was played, informing patients that he or she could rest for 2s. Thus, the time length of a single trial was 10s.

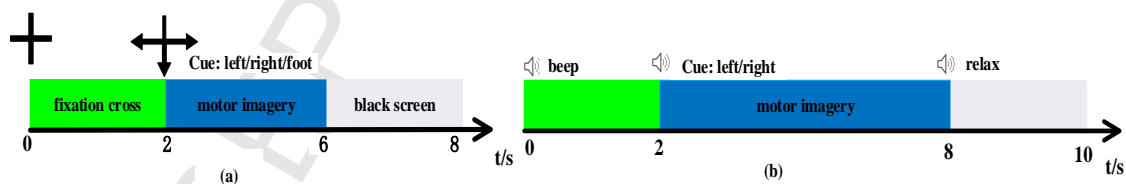


Fig. 1. Illustration of the experimental protocol for a trial in dataset 1 (a) and dataset 2 (b).

## 2.2 Common spatial pattern

The CSP algorithm is an efficient method used to extract discriminative features from the EEG that is commonly used in MI-based BCI systems (Nicolas-Alonso, Corralejo, Gomez-Pilar, Álvarez, & Hornero, 2015; Nasihatkon, Boostani, & Jahromi, 2009; Alvarez-Meza, Velasquez-Martinez, & Castellanos-Dominguez, 2015; Aghaei, Mahanta, & Plataniotis, 2016). The CSP algorithm learns a projection vector to maximize the variance of one class and minimize

the variance of the other class at the same time (Dornhege, Blankertz, Curio, & Müller, 2004; Lemm, Blankertz, Curio, & Müller, 2005). The CSP operation is as follows:

$$\arg \max_w \frac{w^T \Sigma_1 w}{w^T (\Sigma_1 + \Sigma_2) w}, \quad (1)$$

where  $w$  represents the projection vector,  $\Sigma_1$  and  $\Sigma_2$  represent the spatial covariance matrices of the two classes, respectively. It can be regarded as the problem of finding generalized eigenvalues:

$$\Sigma_1 w = (\Sigma_1 + \Sigma_2) w D. \quad (2)$$

$D$  is the diagonal matrix which contains the eigenvalues of  $\Sigma_1$ .

Selecting feature vectors corresponding to the maximum and minimum characteristic values from  $w$  as the projection matrix  $\bar{w} \in R^{N \times 2m}$ ,  $m = 2$ . Then, original EEG samples are projected as follows:

$$Z = \bar{w}^T X, X \in R^{N \times M}, \quad (3)$$

where  $N$  is the number of channels,  $M$  is the number of sample points for each channel,  $^T$  denotes the transpose operator.

$f_p$  can be obtained from  $Z_p$  ( $p = 1 \dots 2m$ ) as features of the original EEG data, expressed as follows:

$$f_p = \log \left( \frac{\text{var}(Z_p)}{\sum_{i=1}^{2m} \text{var}(Z_i)} \right). \quad (4)$$

### 2.3 Support vector machine

Support vector machines (SVMs) find a discrimination hyperplane by maximizing margins between two classes to identify classes (Burges, 1998; Bennett, & Campbell, 2000; Wang, Zhang, Zhong, & Zhang, 2013). The hyperplane can be represented as  $w^T X + b = 0$ ,  $w \in R^d$  is called the weight vector and  $b$  is a scalar (Qiu, Jin, Lam, Zhang, Wang, & Cichocki, 2016). The margins are the distance between the two separated hyperplanes, the training samples nearest the hyperplane are called support vectors. The aim of SVM is to find the optimal hyperplane, as follows:

$$\begin{aligned} \min \Phi(w, \xi) &= \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i, c \geq 0, \\ \text{s.t. } y_i (w^T A^{(i)} + b) &\geq 1 - \xi_i, \xi_i \geq 0, i = (1, \dots, n). \end{aligned} \quad (5)$$

where  $A^i$  denotes a feature vector of the  $i$ th training sample,  $y$  denotes the class label, and  $\xi$  denotes a slack variable. A linear kernel was used in this study. (Qiu, Jin, lam, Zhang, Wang, & Cichocki, 2016).

### 2.4 Correlation-based time window selection algorithm

The aim of the correlation-based time window selection (CTWS) algorithm is to extract the discriminative MI features in the time domain. To illustrate the proposed algorithm, we incorporated CSP and SVM into the structure of the CTWS algorithm for feature extraction and

classification, respectively. Note that, the feature extraction and classification algorithms (i.e., CSP and SVM) are substitutable.

As shown in Fig. 2, the main principle of CTWS algorithm is to constantly adjust the time window of the training data to find the optimized reference signals. The flow of the CTWS algorithm is listed as follows:

---

**Algorithm** Correlation-based time window selection algorithm (CTWS)

---

- 1 **Initialize** Divide the dataset of each participant into 10 blocks. Nine blocks of one dataset (RT-S) were used to obtain the optimized reference signal (OR  $n$ ,  $n=1$  or  $2$ ) and classifier mode, and the remaining one block was used as test data. Set the two classes in each dataset as class 1 and class 2. Set the original start point of the feature time window. Set the classification accuracy (CA) to zero.
  - 2 **For** Current  $h$  in 10-fold cross-validation
  - 3     **For** Current  $k$  in K-Run
  - 4         Calculate reference signal 1 (R1) and reference signal 2 (R2) by averaging the samples of class 1 (C1) and class 2 (C2) over trials, respectively.
  - 5         Generate  $2n + 1$  new time windows by sliding the starting time point forward and backward  $n$  time sampling points, respectively.
  - 6         Select the time window (TW) by finding the maximum correlation with R1 and R2, respectively, and acquire the new C1 (NC1) and C2 (NC2). Thus, we have new dataset (NT-S).
  - 7         Use 10-fold cross-validation to obtain CA based on CSP and SVM.
  - 8         **If** The new CA was higher than the previous CA.
  - 9             Replace T-S with NT-S, and go to 3.
  - 10         **Else**
  - 11             Go to 3.
  - 12         Obtain optimized reference signal OR1 and OR2 based on current NT-S.
  - 13         Calculate the correlation between RT-S and (OR  $n$ ,  $n=1$  or  $2$ ), respectively, to select the TW, thus obtain new R1 T-S (NR1 T-S) and new R2 T-S (NR2 T-S). This step is similar with steps 4 and 5.
  - 14         Employ CSP algorithm to extract the features  $f1$  from NR1 T-S and  $f2$  from NR2 T-S, respectively.
  - 15         Train the SVM classifier.
  - 16         Use the remaining one block of the dataset (Test-S) to calculate testing features  $f1$  and  $f2$ . This step is similar with steps 12 and 13.
  - 17         Calculate the CA of current Test-S using the SVM classifier.
  - 18         Calculate the average CA after 10-fold cross-validation.
  - 19 **End**
- 

To further explain the CTWS algorithm, we employed dataset 1 and dataset 2 as examples to describe the process of the algorithm. In this study, 10-fold cross-validation was used to evaluate the performance of the presented method from each participant. The dataset of each participant was divided into 10 blocks. There were 200 samples for two classes (100 samples for each class) in dataset 1, which was divided into 10 blocks. Therefore, each block contains 20 samples (10 samples for each class). Moreover, there were 60 samples for two classes (30 samples for each

class) in dataset 2, which was also divided into 10 blocks. Therefore, each block contains 6 samples (3 samples for each class). Nine blocks were used as training data and the remaining one block was used as test data.

The original start point of the feature time window (2s) was 2 seconds after cue time in dataset 1 and was 1 second after cue time in dataset 2. To reduce the computation time of the algorithm, only channels C3 and C4 were used to calculate the reference signals in this study. For each participant, the reference signals (R1 and R2) at channels C3 and C4 were acquired by averaging corresponding channel signals over trials of each class. Because there were two classes, four reference signals were obtained, and resulting in two reference signals for each class. Ten data points before and after the start points for each sample were selected as start points of 21 new slide time windows (only containing channels C3 and C4). The correlation between the time windows and the reference signals of the corresponding class were calculated, respectively. After that, the correlation value were calculated for each slide time window, which can be represented as

$$\text{cov}(R_{ij}, C_{ij_k}) = \frac{1}{N_t - 1} \sum_{t=1}^{N_t} (R_{ij}(t) - \bar{R}_{ij}(t))(C_{ij_k}(t) - \bar{C}_{ij_k}(t)), i=1,2, j=3,4, k=1,2,\dots,2n+1, \quad (6)$$

where  $i$  is the index of class,  $j$  is the index of channel,  $t$  is the index of current point in the time window with the length of  $N_t$ ,  $R$  is the reference signal, and  $C$  is the signal of current sample,  $\bar{R}(t)$  and  $\bar{C}(t)$  are the average value of  $R$  and  $C$  over  $t$ , respectively. For example,  $Ri3$  and  $Ri4$  represent the reference signals at channels C3 and C4 of class  $i$ , respectively.  $Ci3_k$  and  $Ci4_k$  represent the signal of channels C3 and C4 of class  $i$  at  $k$ th time window, respectively. In this way, the adjusted time window for the sample, which obtained the highest averaged correlation value, was selected

$$V = \arg \max_k (\text{cov}(Ri3, Ci3_k) + \text{cov}(Ri4, Ci4_k)), k=1,2,\dots,2n+1, \quad (7)$$

where  $V$  represents the time window with maximum average correlation value,  $n$  is the number of generated new time windows ( $n$  was set to 10 in this study). The selected time window was used to update the start point of the sample at all channels. The updated training samples NT-S consisted of new training samples NC1 and NC2, belonging to class1 and class2 respectively, were obtained by shifting the time window of T-S to  $V$ . After that, we replace T-S with NT-S, thus obtain the updated samples (the new T-S). As shown in Fig. 2, by doing this for several repetitions, the optimized reference signals (OR1 and OR2) at C3 and C4 for each class were obtained.

After obtaining the optimized reference signals, new time windows for each sample were generated by sliding the selected starting time point forward and backward one second, respectively. After that, the optimal time windows for each raw sample were selected based on the correlation with OR1 and OR2, respectively. Therefore, we acquired the new training samples (NR1 T-S and NR2 T-S) and testing samples (N1 Test-S and N2 Test-S). Then, CSP and SVM were employed for the model training and accuracy calculation. Finally, the process above was repeated ten times in a 10-fold cross validation scheme to evaluate the average classification accuracy.

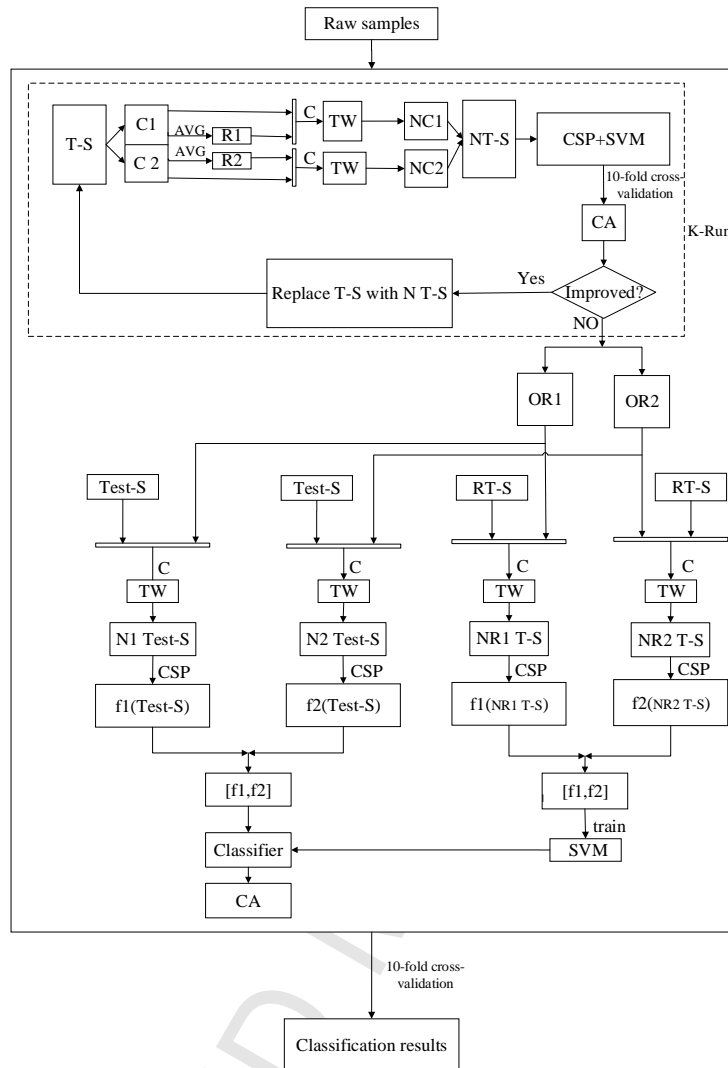


Fig. 2. Structure diagram of the algorithm. T-S: training samples. C1: training samples of class 1. C2: training samples of class 2. AVG: The C3 and C4 channel signal samples are averaged. R1: reference signal of class 1. R2: reference signal of class 2. C: correlation. TW: adjust time window of samples. NC1: new samples of class 1. NC2: new samples of class 2. NT-S: new training samples. CA: classification accuracy. OR1: optimized reference signal of class 1. OR2: optimized reference signal of class 2. RT-S: raw training samples. NR T-S: new raw training samples. f1, f2: features extracted by CSP. Test-S: test samples. N test-S: new test samples.

### 3. Results

To determine whether the starting time varied in motor imagery (MI) tasks, we calculated the starting time based on the correlation-based time window selection (CTWS) algorithm for dataset 1 and dataset 2 (see Fig. 3). The results indicate that the starting time of any one individual varies from one trial to the next during motor imagery.



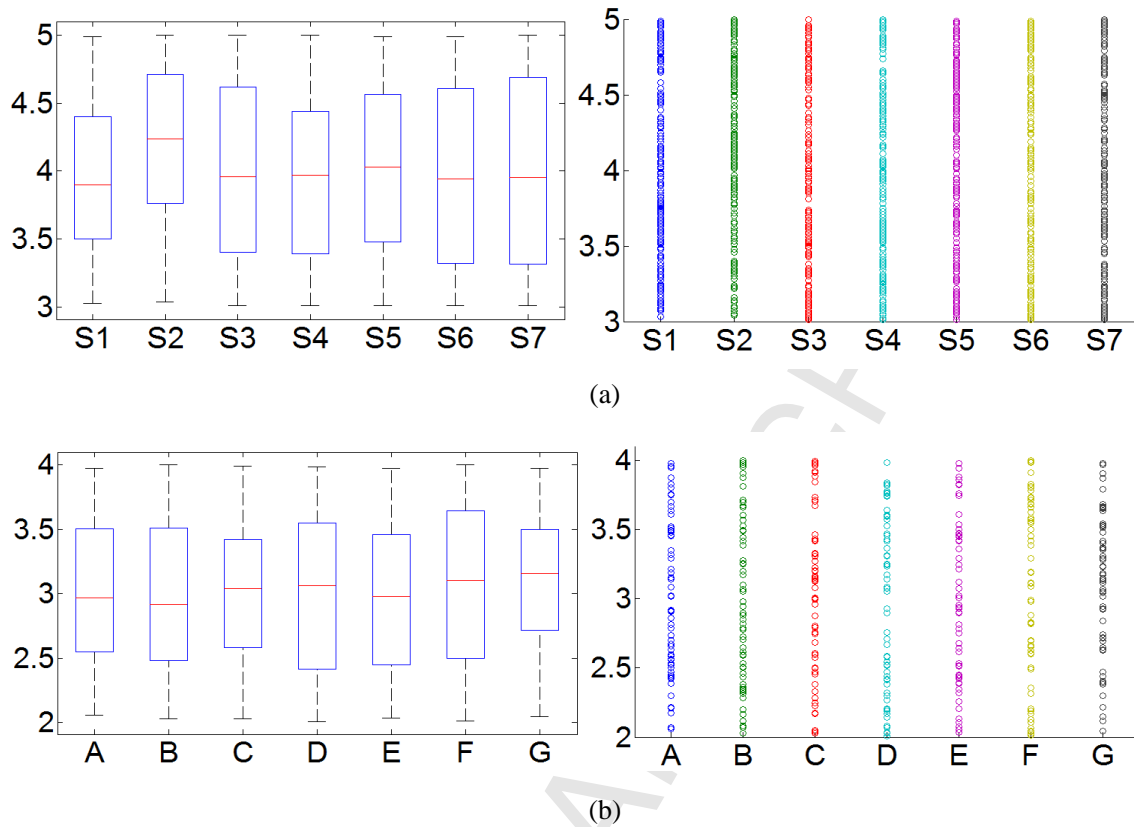


Fig. 3. Boxplots and Scatter plot represents the starting time of selected time windows for each participant in dataset 1 (a) and dataset 2 (b). (The time schedule can be found in Fig. 1, in which the cue time is 2 second).

We also compared the CSP feature distribution between analysis attempts made with and without the CTWS algorithm to illustrate the effectiveness of the proposed method. Fig. 4 depicts the feature distribution of each class in dataset 1 and dataset 2. The blue and red circles represent the two different feature classes, respectively. For each participant, the subfigure at the top was acquired by CSP directly, the subfigure below was acquired by combing CTWS with CSP (see Fig. 2). As shown in Fig. 4, the results demonstrate that the features extracted by CTWS+CSP were easier to classify compared to the CSP algorithm.

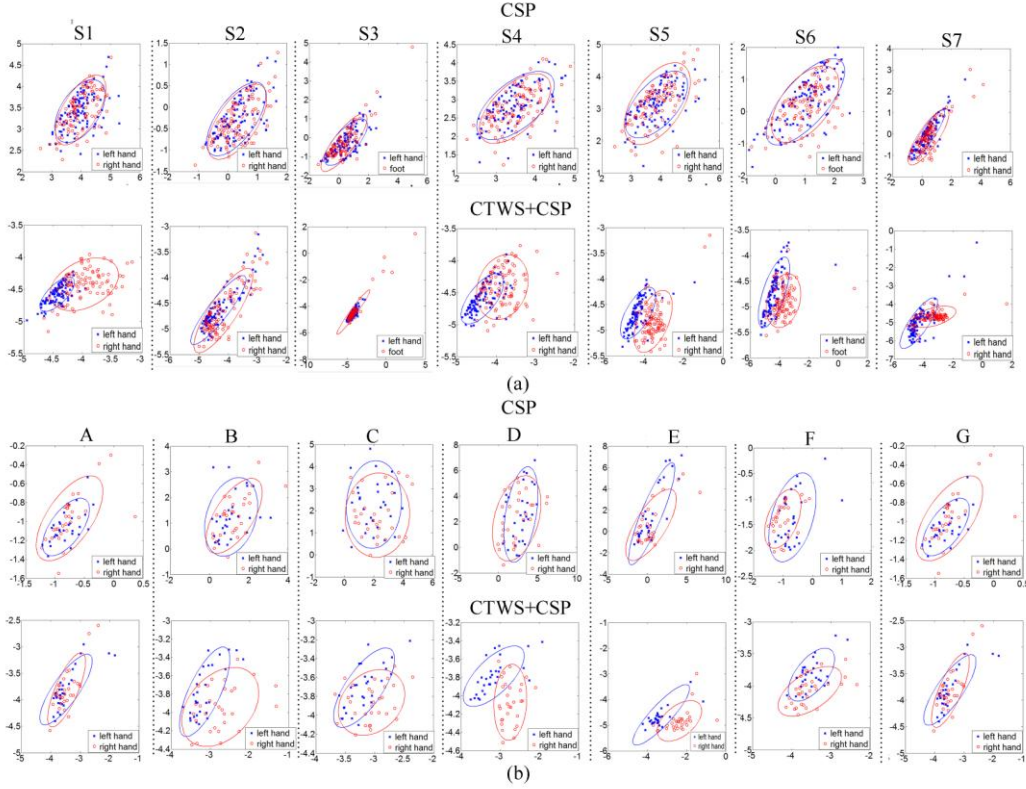


Fig. 4. Feature distribution of each class extracted by CSP and combining CTWS with CSP in dataset 1 (a) and dataset 2 (b).

To evaluate the effectiveness and universality of the proposed CTWS algorithm, we compared the classification accuracies of CSP with and without the CTWS algorithm using dataset 1 (healthy individuals) and dataset 2 (stroke patients). Paired-sample  $t$ -tests were performed for further statistical comparisons of system performance across participants. As shown in Table 1, the results indicated that CTWS+CSP achieved better performance compared to using CSP directly. More specifically, the average classification accuracy of CTWS+CSP was significantly improved 16.72% and 5.24% when compared to CSP alone (82.93% versus 66.21%,  $p < 0.005$ ; 71.67% versus 66.43%,  $p < 0.05$ ). Note that, the both accuracies of subjects G in dataset 2 did not exceed the chance level (50%). Therefore, we recalculated the performance of CTWS+CSP and CSP for dataset 2 by excluding subject G. The average classification accuracy of CTWS+CSP was improved 4.17% compared to that of CSP alone when subjects G was excluded (75.28% versus 71.11%,  $p = 0.0813$ ).

**Table 1**

Comparison of the classification accuracy between CSP and CTWS+CSP for dataset1 and dataset2. P-Value denotes the paired  $t$ -test of classification accuracy between CSP and CTWS+CSP. "L vs R" is left hand and right hand and "L vs F" is left hand and feet.

dataset 1 (BCI Computation IV Dataset I)			dataset 2 (stroke patients)		
Subject	CSP	CTWS+CSP	Subject	CSP	CTWS+CSP
S1(L vs R)	64.50	82.50	A (L vs R)	46.67	60.00
S2(L vs R)	51.50	76.00	B (L vs R)	71.67	75.00
S3(L vs F)	53.00	66.00	C (L vs R)	60.00	60.00
S4(L vs R)	89.00	96.00	D (L vs R)	85.00	88.33
S5(L vs R)	93.50	98.00	E (L vs R)	88.33	90.00

S6(L vs F)	46.00	80.00	F (L vs R)	75.00	78.33
S7(L vs R)	66.00	82.00	G (L vs R)	38.33	50.00
Mean	66.21±18.56	82.93±11.12	Mean	66.43±18.92	71.67±15.33
P-Value	0.0048		P-Value	0.0355	

As we mentioned in Subsection 2.4, the CSP algorithm is substitutable in the structure of the CTWS algorithm. Recently, the Sub-Alpha-Beta Log-Det Divergences (Sub-ABLD) algorithm, a modified version of CSP algorithm, was reported to outperform the other existing algorithms for MI feature extraction (Thiyam, Cruces, & Olias, 2017). In this study, we further incorporate the Sub-ABLD algorithm into the proposed CTWS algorithm to verify its universality. The feature distribution comparison was shown in Fig. 5. We could observe the features extracted by CTWS+Sub-ABLD were easier to classify compared to the Sub-ABLD algorithm using both dataset 1 and dataset 2. As shown in Table 2, the incorporation CTWS algorithm into Sub-ABLD (CTWS+Sub-ABLD) achieved significantly higher classification accuracy than that of Sub-ABLD. In particular, the average accuracy increased from 77.50% to 84.86% (improved 7.36%,  $p < 0.01$ ) and 59.52% to 68.81% (improved 9.29%,  $p < 0.05$ ).

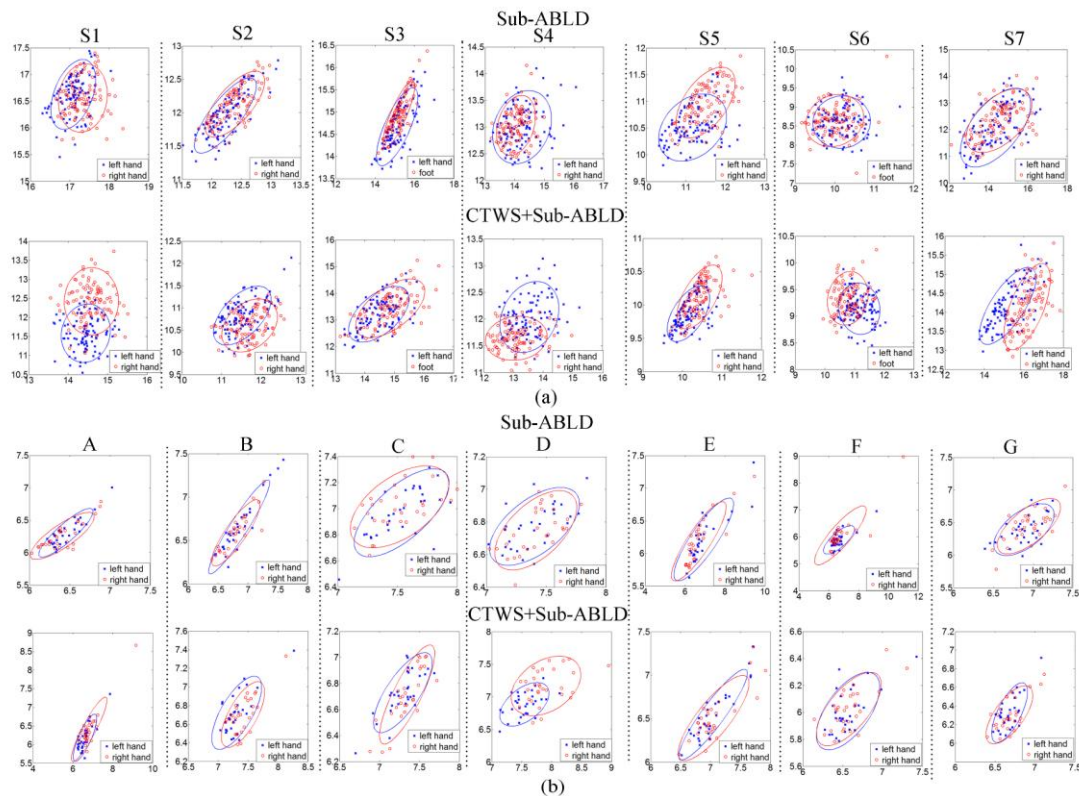


Fig. 5. Feature distribution of each class extracted by Sub-ABLD and combining CTWS with Sub-ABLD in dataset 1 (a) and dataset 2 (b).

**Table 2**

Comparison of the classification accuracy between Sub-ABLD and CTWS+Sub-ABLD for dataset 1 and dataset 2. P-Value denotes the paired  $t$ -test of classification accuracy between Sub-ABLD and CTWS+Sub-ABLD. “L vs R” is left hand and right hand and “L vs F” is left hand and feet.

dataset 1 (BCI Computation IV Dataset I)	dataset 2 (stroke patients)
--	-----------------------------

Subject	Sub-ABLD	CTWS+Sub-ABLD	Subject	Sub-ABLD	CTWS+Sub-ABLD
S1(L vs R)	69.50	83.00	A(L vs R)	41.67	53.33
S2(L vs R)	65.00	67.00	B(L vs R)	65.00	61.67
S3(L vs F)	78.00	85.50	C(L vs R)	41.67	63.33
S4(L vs R)	84.00	93.00	D(L vs R)	66.67	85.00
S5(L vs R)	95.50	99.00	E(L vs R)	86.67	86.67
S6(L vs F)	72.00	85.50	F(L vs R)	65.00	76.67
S7(L vs R)	78.50	81.00	G(L vs R)	50.00	55.00
Mean	77.50±10.15	84.86±10.04	Mean	59.52±16.28	68.81±12.04
P-Value		0.0075	P-Value		0.0373

#### 4. Discussion

Feature extraction is one of the most important steps in motor imagery (MI)-based BCI systems (Park, Hwang, et al., 2013; Boostani, Graimann, Moradi, & Pfurtscheller, 2007; Kevric, & Subasi, 2017). In particular, CSP, a spatial feature extraction algorithm, has become the most commonly used algorithm in the MI-based BCI research field. In recent years, several studies have extended the CSP algorithm to the frequency domain, proposing spatial-spectral feature extraction algorithms, such as RCSP (Lotte, & Guan, 2011), SSCSP (Shin, Lee, Lee, & Lee, 2012), FERCS (Su, Li, & Wang, 2015), FBCSP (Ang, Chin, Zhang, & Guan, 2008), and Wavelet-CSP (Robinson, Vinod, Ang, Tee, & Guan, 2013). Although these modified CSP algorithms compensate for the shortcomings of conventional CSP, none of them consider the variation in the time latency during the MI task. As shown in Fig. 3, the starting times significantly varied across trials for each participant. Therefore, fixing the start of the time window used for MI feature extraction would likely reduce the classification accuracies. Although a few studies have examined the effect of time window selection and selected time windows via a manual approach, this method requires extensive time for paradigm design and could not achieve consistent satisfactory performance (Dornhege, Blankertz, Curio, & Müller, 2004; Qiu, Jin, Lam, Zhang, Wang, & Cichocki, 2016).

In this study, the proposed CTWS algorithm considers the time variations among trials during an MI task, and uses the correlation analysis to automatically adjust the time window for each sample. The experimental results demonstrate that the classification accuracy with CTWS significantly increased compared to that of other more traditional approaches (see Tables 1 and 2). The superior results achieved by the CTWS algorithm may be mimicked by extending current feature extraction approaches to the time domain. We note that the feature extraction algorithm in the structure of the CTWS algorithm is substitutable. The non-stationary nature of EEG data is also a challenge for EEG signal analysis (Thiyam, Gruces, & Olias, 2017). Sub-ABLD exhibited a certain robustness to the presence of outlier trials in the dataset. This study combined CTWS with Sub-ABLD and achieved significantly improved classification accuracies, which further verified the universality of the CTWS algorithm. Moreover, we employed the CTWS on the MI dataset of stroke patients (dataset 2), which also lead to significant improvements of classification accuracy. However, we notice that the benefit of the CTWS algorithm on the stroke patient datasets (dataset 2) was reduced compared to the dataset of healthy individuals (dataset 1) (see the P-Value in Table 1 and 2). These results may be due to the relatively large disturbance of the stroke patients' MI features, which could not provide enough information for classification (see Fig. 4 and 5).

As shown in Fig. 2, the reference signal selection is the key aspect of the CTWS algorithm

and can severely affect the classification accuracy. It is impractical, due to the heavy load of computation, to employ all collected EEG channels in optimization of the reference signal. Therefore, in this study channels C3 and C4 were chosen to calculate the reference signal because they carry important characteristics of MI (Pfurtscheller, Brunner, Schlögl, & Da Sliva, 2006; Pregener, & Pfurtscheller, 1999). However, the spatial distribution of MI features varies among individuals (Qiu, Jin, Lam, Zhang, Wang, & Cichocki, 2016), thus channels C3 and C4 may not be the optimal channel set for some cases. In future work, to further improve the performance of the CTWS algorithm, we will choose the channel set for the reference signal optimization based on channel selection approaches.

## 5. Conclusion

In this study, we proposed a novel correlation-based time window selection (CTWS) algorithm for motor imagery (MI)-based BCIs. In our approach, the optimized reference signals for each class were selected based on correlation analysis and performance evaluation. After that, the starting points of the time windows for both training and testing samples were adjusted using correlation analysis again. Finally, the feature extraction and classification algorithms were employed to calculate the classification accuracy. Experimental results suggest that the CTWS algorithm can provide improved performance compared to directly using feature extraction approaches. More specifically, the average classification accuracy improved 16.72% ( $p < 0.005$ ) on the dataset of healthy participants (BCI Computation IV Datasets 1), and 5.24% on the dataset of stroke patients, when using the proposed approach compared to using CSP directly. Moreover, we evaluate the performance of CTWS used with Sub-ABLD (a recent proposed algorithm for MI feature extraction), and the average accuracy increased significantly by 7.36% ( $p < 0.01$ ) and 9.29% ( $p < 0.05$ ). The proposed CTWS algorithm paves the way for further MI feature extraction research.

## Acknowledgments

This work was supported in part by the Grant National Natural Science Foundation of China, under Grant Nos. 91420302, 61573142, 61703407. This work was also supported by the Programme of Introducing Talents of Discipline to Universities (the 111 project) under Grant B17017, the Fundamental Research Funds for the Central Universities (WH1516018), Shanghai Chenguang Program under Grant 14CG31, and the Foundation of Key Laboratory of Science and Technology for National Defense (No.6142222030301)

## References:

- 
- Aghaei, A. S., Mahanta, M. S., & Plataniotis K. N. (2016). Separable common spatio-spectral patterns for motor imagery BCI systems. *IEEE Transactions on Biomedical Engineering*, 63(1), 15-29.
- Alvarez-Meza, A. M., Velasquez-Martinez L. F., & Castellanos-Dominguez G. (2015). Time-series discrimination using feature relevance analysis in motor imagery classification. *Neurocomputing*, 151, 122-129.
- Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008). Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. In *Proceeding of the IEEE International Joint Conference on Neural Networks* (pp.2390-2397).

- Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2012). Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. *Pattern Recognition*, 45, 2137-2144.
- Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah?. *Acm Sigkdd Explorations Newsletter*, 2(2), 1-13.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K. R., & Curio, G. (2007). The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2), 539-550.
- Boostani, R., Graimann, B., Moradi, M. H., & Pfurtscheller, G. (2007). A comparison approach toward finding the best feature and classifier in cue-based BCI. *Medical & Biological Engineering & Computing*, 45(4), 403-412.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 121-167.
- Daly, I., Nasuto, S. J., & Warwick, K. (2011). Single tap identification for fast BCI control. *Cognitive Neurodynamics*, 5(1), 21-30.
- Dornhege, G., Blankertz, B., Curio, G., & Müller, K. R. (2004). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Transactions on Biomedical Engineering*, 51(6), 993-1002.
- Dornhege, G. (Eds). (2007). *Toward brain-computer interfacing*. MIT Press.
- Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Academic press.
- He, B., Baxter, B., Edelman, B. J., Cline, C., & Ye, W. W. (2015) Noninvasive brain-computer interfaces based on sensorimotor rhythms. *Proceedings of the IEEE*, 103(6), 907-925.
- He, W., Wei, P., Wang, L., & Zou, Y. (2012). A novel EMD-based common spatial pattern for motor imagery brain-computer interface. In *IEEE-EMBS International Conference on Biomedical and Health Informatics* (pp.216-219).
- Jin, J., Allison, B. Z., Sellers, E.W., Brunner C., Horki P., Wang, X., Neuper, C., (2011) Adaptive P300 based control system. *Journal of Neural Engineering*, 8(3):036006.
- Jin, J., Sellers, E. W., Zhou,S., Zhang, Y., Wang, X., Cichocki, A., (2015) A P300 Brain-Computer Interface Based on a Modification of the Mismatch Negativity Paradigm. *International Journal of Neural Systems*, 25(3): 1550011
- Jin, J., Zhang, H., Daly, I., Wang, X., Cichock, A., (2017) An improved P300 pattern in BCI to catch user's attention, *Journal of Neural Engineering*, 14(3):036001.
- Gouy-Pailler, C., Mattout, J., Congedo, M., & Jutten, C. BCI competition iv, dataset 1: motor imagery, uncued classifier application.
- Kevric, J., & Subasi, A. (2017). Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomedical Signal Processing and Control*, 31, 398-406.
- Koles, Z. J. (1991). The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 79(6), 440-447.
- Lemm, S., Blankertz, B., Curio, G., & Müller, K. R. (2005). Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Transactions on Biomedical Engineering*, 52(9), 1541-1548.
- Lotte, F., & Guan C. (2011). Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering*, 58(2), 355-362.
- Nasihatkou, B., Boostani R., & Jahromi M. Z. (2009). An efficient hybrid linear and kernel CSP approach for EEG feature extraction. *Neurocomputing*, 73(1-3), 432-437.
- Nicolas-Alonso, L. F., Corralejo, R., Gomez-Pilar, J., Álvarez, D., & Hornero, R. (2015). Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain-computer interfaces. *Neurocomputing*, 159(C), 186-196.
- Onose, G., Grozea, C., Angheliescu, A., Daia, C., Sinescu, C. J., Ciurea, A. V., Spircu, T., Mirea, A., And one, I., Spânu, A., Popescu, C., Danoczy, M., & Popescu, F. (2012). On the feasibility of using motor imagery EEG-based brain-computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up. *Spinal Cord*, 50, 599-608.
- Pan, J., Li, Y., Gu, Z., & Yu, Z. (2013). A comparison study of two P300 speller paradigms for brain-computer interface. *Cognitive Neurodynamics*, 7(6), 523-529.

- Park, S. A., Hwang, H. J., Lim, J. H., Choi, J. H., Jung, H. K., & Im, C. H. (2013). Evaluation of feature extraction methods for EEG-based brain-computer interfaces in terms of robustness to slight changes in electrode locations. *Medical & Biological Engineering & Computing*, 51(5), 571-579.
- Pfurtscheller, G., & Da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11), 1842-1857.
- Pfurtscheller, G., Brunner, C., Schlögl, A., & Da Silva, F. L. (2006). Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1), 153-159.
- Pregenzer, M., & Pfurtscheller, G. (1999). Frequency component selection for an EEG-based brain to computer interface. *IEEE Transactions on Rehabilitation Engineering*, 7(4), 413-419.
- Qiu, Z., Jin, J., Lam, H. K., Zhang, Y., Wang, X., & Cichocki, A. (2016). Improved SFFS method for channel selection in motor imagery based BCI. *Neurocomputing*, 207, 519-527.
- Ramoser, H., Müller-Gerking J., & Pfurtscheller G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4), 441-446.
- Reinhold, S., Faller, J., Friedrich, E., Opisso, E., Costa, U., Kübler, A., & Müller-Putz, G. (2015). Individually adapted imagery improves brain-computer interface performance in end-users with disability. *Plos One*, 10(5), e0123727.
- Robinson, N., Vinod, A. P., Ang, K. K., Tee, K. P., & Guan, C. T. (2013). EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm. *IEEE Transactions on Biomedical Engineering*, 60(8), 2123-2132.
- Rodriguez-Bermudez, G., Garcia-Laencina, P., J., & Roca-Dorda, J. (2013). Efficient automatic selection and combination of EEG features in least squares classifiers for motor imagery brain-computer interfaces. *International Journal of Neural Systems*, 23(4), 10.1142/S0129065713500159.
- Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P., & Müller, K. R. (2006). Towards adaptive classification for BCI. *Journal of Neural Engineering*, 3(1), R13.
- Shin, Y., Lee, S., Lee, J., & Lee, H. (2012). Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems. *Journal of Neural Engineering*, 9, 056002.
- Solis-Escalante, T., Müller-Putz, G., Brunner, C., Kaiser, V., & Pfurtscheller, G. (2010). Analysis of sensorimotor rhythms for the implementation of a brain switch for healthy subjects. *Biomedical Signal Processing and Control*, 5, 15-20.
- Su, Y., Li, Y., & Wang, S. (2015). Filter ensemble regularized common spatial pattern for EEG classification. *Proceedings of International Conference on Digital Image Processing (Vol.9631)*.
- Thiyam D. B., Cruces S., & Olias J. (2017). Optimization of alpha-beta log-det divergences and their application in the spatial filtering of two class motor imagery movements. *Entropy*, 19(3), 89.
- Wang, L., Zhang, X., Zhong, X., & Zhang, Y. (2013). Analysis and classification of speech imagery EEG for BCI. *Biomedical Signal Processing & Control*, 8(6), 901-908.
- Wolpaw, J. R., McFarland, D. J., Neat, G. W., & Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78, 252-259.
- Wolpaw, J. R., & Wolpaw, E. W. (Eds). (2012). *Brain-computer interfaces: principles and practice*. OUP USA.
- Yuan, H., & He, B. (2014). Brain-computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Transactions on Biomedical Engineering*, 61(5), 1425-1435