

A PENALIZED TIME-FREQUENCY BAND FEATURE SELECTION AND CLASSIFICATION PROCEDURE FOR IMPROVED MOTOR INTENTION DECODING IN MULTICHANNEL EEG

Victoria Peterson^{*1,2,3}, Dominik Wyser², Olivier Lambercy², Ruben Spies⁴ and Roger Gassert²

¹*Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, UNL, CONICET, FICH, Santa Fe, Argentina.*

²*Rehabilitation Engineering Laboratory, Department of Health Sciences and Technology, ETH Zurich, Switzerland.*

³*Facultad de Ingeniería, Universidad Nacional de Entre Ríos (FI-UNER), Oro Verde - Entre Ríos - Argentina.*

⁴*Instituto de Matemática Aplicada del Litoral, UNL, CONICET, FIQ, Santa Fe, Argentina.*

Abstract: *Objective.* Motor imagery brain-computer interfaces (MI-BCIs) based on electroencephalography (EEG), a promising technology to provide assistance and support rehabilitation of neurological patients with sensorimotor impairments, require a reliable and adaptable subject-specific model to efficiently decode motor intention. The most popular EEG feature extraction algorithm for MI-BCIs is the common spatial patterns (CSP) method, but its performance strongly depends on the predefined frequency band and time segment length for analyzing the EEG signal. *Approach.* In this work, a novel method for efficiently decoding motor intention for EEG-based BCIs performing multiple frequency band analysis in multiple EEG segments is presented. This decoding algorithm uses raw multichannel EEG data which are decomposed into specific T temporal and F frequency bands. Features are extracted at each t - f band by using CSP. Feature selection and classification are simultaneously performed by means of a fast procedure, based on elastic-net regression, which allows for the inclusion of a priori discriminative information into the model. The effectiveness of the proposed method is tested off-line on two public EEG-based MI-BCI datasets and on a self-acquired dataset in two configurations: multiple temporal windows and single temporal window. *Main results.* The experimental results show that the proposed multiple time-frequency band method yields overall accuracy improvements of up to 9% (average accuracy of 84.8%) as compared to the best current state-of-the-art methods based on filter bank analysis and CSP for MI detection. Also, classification variability is reduced, making the proposed method more robust to intra-subject EEG fluctuations. *Significance.* This paper presents a novel approach for improving motor intention detection by automatically selecting subject-specific spatio-temporal-spectral features, especially when MI has to be detected against rest condition. This technique contributes to the further advancement and application of EEG-based MI-BCIs for assistance and neurorehabilitation therapy.

Keywords: *Motor Imagery, Brain-Computer Interfaces, Time-Frequency Bands, Sparse Feature Selection, Mixed-Norm Penalization.*

*vpeterson@sinc.unl.edu.ar

1 INTRODUCTION

Brain-computer interfaces (BCIs) are advanced communication systems which can restore, enhance or replace the normal pathways used by people to interact with and control their environment [1]. By means of a BCI, a person communicates with the outside world based only on recorded brain activity signals. The most common type of non-invasive BCI relies on electroencephalography (EEG) [2, 3]. EEG-based BCIs translate changes of electric potentials acquired at the surface of the head with cerebral origin into output commands aimed to control an external device, such as a computer (e.g. [4, 5]), a wheelchair (e.g. [6, 7]) or a prosthesis (e.g. [8, 9]). For rehabilitation purposes, BCIs based on motor imagery (MI), i.e., the mental simulation of real movements without any overt motor output, are a promising approach to support motor recovery of neurological patients with sensorimotor impairments [10]. Through an MI-BCI controlled robotic orthosis, it is hypothesized that the direct sensory feedback related to the motor intention could support the recovery of motor function and stimulate brain plasticity by closing the sensorimotor loop [11]. It is well-known that the brain activity associated with imagined movements produces changes in the sensorimotor rhythms similar to those resulting from the corresponding motor execution [12]. These changes reveal amplitude reductions or increments in certain frequency bands, which are known as event-related desynchronizations (ERD) and event-related synchronizations (ERS), respectively. These ERD/ERS patterns are most prominent in the mu (8-12 Hz) and beta (13-30 Hz) bands, and are also observable in gamma oscillations (> 30 Hz) close to 40 Hz [13].

Reliably decoding an MI task from the ongoing EEG signal is a challenging pattern recognition problem, mainly due to the low signal-to-noise ratio of the raw measurements, poor spatial resolution and the intrinsic non-stationary characteristic of the problem [14]. The imagery of certain movements (e.g., left vs. right hand or foot vs. hand) can be spatially distinguished in the sensorimotor area. Thus, MI detection involves recognizing the associated ERD/ERS patterns in the sensorimotor cortex, for which it is common to perform spatial filtering [15]. Common spatial patterns (CSP) maximizes the variance of a spatially filtered signal for one of the MI conditions while minimizing it for the other one [16]. Hence, since the variance of a band-pass filtered EEG signal is equal to its band power, CSP constitutes an appropriate framework for distinguishing between two MI conditions [17]. Although good classification performance ($> 75\%$) can be achieved by using CSP, its success strongly relies on the correct selection of the filter band (typically a wide band from 8-30 Hz) and the EEG time segment extracted after the MI instruction cue [17]. One way to address the issue of selecting the “best” subject-specific frequency band is by exhaustive search. Although this “manual” approach has been proven to be effective, it is impractical (due to the computational time required) and it has not been standardized [18]. For automatic subject-specific frequency bands selection, Ang et al. [19] proposed the filter bank CSP (FBCSP) method, a four-stage procedure in which CSP is applied at several fixed frequency bands, and where the most relevant sub-band CSP features are automatically pair-wise selected based upon mutual information criteria. A modified version of FBCSP, called discriminant FBCSP (DFBCSP) [20], applying subject-specific filter

bank based on Fisher's ratio before feature extraction, was introduced yielding better classification accuracies than those achieved by FBCSP. More recently, the sparse filter bank CSP (SFBCSP) [21] was proposed, in which a small number of sub-band CSP features are automatically selected based on LASSO (least absolute shrinkage and selection operator) regression [22]. The SFBCSP method has shown better overall classification performances over publicly available BCI competition datasets as compared to the traditional CSP, FBCSP and DFBCSP methods [21]. As a drawback, SFBCSP requires cross-validation estimates of a regularization parameter (being a time consuming process) and it does not take into account any discriminative information among features, neglecting important prior knowledge for feature selection.

Although all of the above filter bank methods have shown better performance than the traditional wide-band CSP, they all use a pre-defined time window in which the filtering procedure is performed (typically from 0.5 to 2.5 s after onset of the MI cue). For the selection of the optimal time window for MI classification, several groups have proposed approaches that automatically select the "optimal" time window segments for MI classification. In particular, Ang et al. [23] extended their previous FBCSP method by applying filter banks to three predefined time segments, in which the optimal time segment is selected after cross-validation as the one with highest mutual information. More recently, the sparse time-frequency segment common spatial pattern (STFSCSP) method was presented [24]. The STFSCSP method, after proper channel selection, improves the SFBCSP criterion by splitting up each frequency band into different time windows, showing the importance of optimizing both the frequency and temporal bands in an MI-BCI.

In this work we propose a method to overcome the limitations of fixed time segments and fixed frequency-bands analysis in MI-BCI, called penalized time-frequency band CSP (PTFBCSP), in which the "optimal" subject-specific time-frequency bands are automatically selected. The main characteristics of the proposed method are: i) it enables the use of multiple time windows for analyzing the EEG trials in different time segments; ii) filter bank analysis is performed at each time window and iii) the time-frequency band CSP feature selection and classification are simultaneously made in a fast procedure based upon an integrated model, which takes into account a priori discriminative information across features. Here the spatio-temporal-spectral features extracted with CSP are assumed to be independent. Unlike in [24], in which the time-frequency band feature selection is automatically made before classification by means of LASSO, in our work we use automatic time-frequency band feature selection simultaneously with classification by means of the generalized sparse linear discriminant analysis (GSDA) method [25]. We recall that GSDA was conceived as a regularized version of the traditional linear discriminant analysis (LDA), which simultaneously performs weighted feature selection, by means of a priori discriminative information between features, and classification. In [25], GSDA was successfully used to tackle the binary classification of event-related potentials in small training size scenarios, yielding the best overall classification performance as compared to other regularized state-of-the-art LDA methods.

In this paper we detail the methodological steps of the proposed method and present experimental classification results in three off-line scenarios using two publicly available EEG-based MI datasets and a self-acquired dataset using two configurations of PTFBCSP, namely when multiple temporal windows are used and when a single temporal window is used. The latter, which we call penalized frequency band CSP (PFBCSP), is a particular case of PTFBCSP when $T = 1$. We hypothesize that: i) the introduction of a priori discriminative information for feature selection in an integrated model could enhance class separability, and ii) the combination of time and frequency features (information) could lead to more accurate and robust classification. Comparison results with the wide-band CSP method, as well as with the state-of-the-art filter bank-CSP-based methods are presented.

The organization of this article is as follows. In Subsection 2.1 the PTFBCSP method is presented. In particular, both CSP and GSDA are briefly introduced. The two public MI-EEG databases and our self-acquired dataset are described in Subsection 2.2. The experimental results can be found in Subsection 2.3. Comparative classification results over the three datasets are presented in Section 3. Finally, discussions, concluding remarks and future works are presented in Sections 4 and 5, respectively.

2 MATERIALS AND METHODS

2.1 PENALIZED TIME-FREQUENCY BAND COMMON SPATIAL PATTERNS

The proposed method for MI classification is schematically depicted in Figure 1. PTFBCSP is a combination of signal processing and machine learning methods. As first step, each raw multichannel EEG signal is decomposed into T temporal windows ($tb_1, \dots, tb_t, \dots, tb_T$). Secondly, each t^{th} temporal segment is split up in F frequency bands ($fb_1, \dots, fb_f, \dots, fb_F$). Then, spatial features at each time-frequency band are extracted by means of CSP. The selection of the optimal spatio-temporal-spectral features together with classification is automatically made by GSDA.

2.1.1 Feature extraction based on CSP

The CSP method is one of the most popular feature extraction algorithms used in MI-BCIs. It was first used in multichannel EEG-based BCIs for left vs. right motor imagery [16]. Given two EEG single trials recorded under two MI conditions, CSP finds a linear transformation maximizing the variance of one condition while minimizing the variance of the other one.

Let $\{\mathbf{X}_c^i\}_{i=1}^{n_c}$ be a set of n_c band-pass filtered EEG trials, where $c = 1, 2$ represents each one of the two considered MI conditions. Here, each \mathbf{X}_c^i is a $p \times m$ matrix, where p and m denote the number of channels and the number of sample points per channel, respectively. In the following we shall assume that each single trial \mathbf{X}_c^i is standardized to zero mean and unit variance. The goal of CSP is to find p “appropriate” spatial filters,

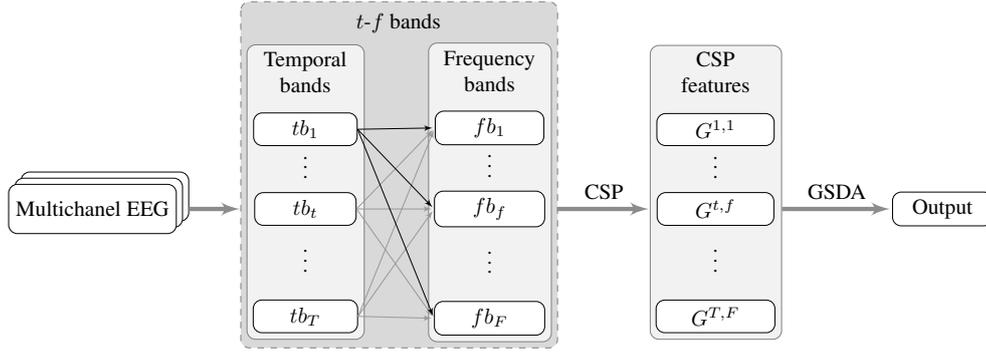


Figure 1: Schematic representation of the proposed PTFBCSP. The multichannel EEG trial is decomposed into T temporal windows. Afterwards, each t^{th} temporal segment is divided into F frequency bands. The CSP is applied at each time-frequency segment. Finally, the selection of the optimal time-frequency CSP features is automatically performed together with classification by GSDA.

$\mathbf{w}_1, \dots, \mathbf{w}_p \in \mathbb{R}^p$, for linearly transforming an input signal \mathbf{X} according to:

$$\mathbf{Z} = \mathbf{W}' \mathbf{X}, \quad (1)$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_p]$ and $'$ denotes transpose. The optimization criterion followed by CSP to find the p spatial filters relies on pool estimates of the covariance matrices for each of the two classes:

$$\Sigma_c = \frac{1}{n_c} \sum_{i=1}^{n_c} \mathbf{X}_c^i (\mathbf{X}_c^i)', \quad c = 1, 2. \quad (2)$$

The CSP analysis proceeds with the simultaneous diagonalization of the two covariance matrices Σ_1 and Σ_2 :

$$\mathbf{W}^T \Sigma_1 \mathbf{W} = \Delta_1, \quad (3)$$

$$\mathbf{W}^T \Sigma_2 \mathbf{W} = \Delta_2, \quad (4)$$

where each column of the matrix \mathbf{W} is appropriately scaled so that the two strictly positive diagonal matrices Δ_c , $c = 1, 2$, satisfy $\Delta_1 + \Delta_2 = \mathbf{I}$. By letting $\gamma_j = \frac{\delta_1^j}{\delta_2^j}$, where δ_c^j is the j^{th} diagonal element of Δ_c , solving Eqs. (3) and (4) is equivalent to solving the following generalized eigenvalue problem:

$$\Sigma_1 \mathbf{w}_j = \gamma_j \Sigma_2 \mathbf{w}_j, \quad j = 1, \dots, p. \quad (5)$$

The generalized eigenvalues γ_j quantify the variance ratio between classes at feature j . A large γ_j value indicates that the corresponding spatial filter \mathbf{w}_j yields high variance for class 1 and low variance for class 2, while the opposite happens for a small γ_j value. Since the goal is to linearly transform the input signals into signals having the largest possible band-power differences between both classes, the extracted spatial filters \mathbf{w}_j are rearranged according to the corresponding γ_j values in decreasing order. Note that \mathbf{W} is non-singular since its columns \mathbf{w}_j are linearly independent. Thus, from Eq. (1) we obtained $\mathbf{X} = (\mathbf{W}^{-1})' \mathbf{Z}$. Each column vector $\mathbf{a}_j \in \mathbb{R}^p$, $j = 1, \dots, p$, of the matrix $\mathbf{A} = (\mathbf{W}^{-1})' \in \mathbb{R}^{p \times p}$ is called a spatial pattern.

In general, only a small number K of spatial filters is used for feature computation. Let us define \mathbf{Z}_K^* as the $p \times 2K$ matrix formed by stacking side-by-side the first and last K rows of the matrix \mathbf{Z} , as follows $\mathbf{Z}_K^* = [\mathbf{z}_1, \dots, \mathbf{z}_K, \mathbf{z}_{p-(K+1)}, \dots, \mathbf{z}_p]'$. Then, the CSP feature vector is defined as $\mathbf{g} \doteq [g_1, \dots, g_{2K}]'$, where

$$g_j \doteq \log(S^2(\mathbf{z}_j^*)), \quad j = 1, \dots, 2K, \quad (6)$$

with $S^2(\mathbf{z}_j^*) \doteq \frac{1}{p} \sum_{\ell=1}^p (\mathbf{z}_j^{*\ell} - \mu_j)^2$, $\mu_j \doteq \frac{1}{p} \sum_{\ell=1}^p \mathbf{z}_j^{*\ell}$.

In the context of the PTFBCSP, the CSP algorithm is implemented at each frequency band f coming from a particular time segment t . This results in $2KF$ extracted features per time-band for each single EEG trial. By concatenating these time-frequency band CSP features and piling up the corresponding resulting vector of each one of the n training samples, we construct the following feature matrix \mathbf{G} of dimension $n \times q$, $q \doteq 2KFT$:

$$\mathbf{G} \doteq \begin{bmatrix} g_{1,1} & \dots & g_{1,2KF} & \dots & g_{1,2KFT} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g_{n,1} & \dots & g_{n,2KF} & \dots & g_{n,2KFT} \end{bmatrix}, \quad (7)$$

where $g_{i,j}$ denotes the j^{th} time-frequency feature extracted from the i^{th} EEG trial.

2.1.2 Feature selection and classification via GSDA

Selecting the most relevant features from the feature matrix \mathbf{G} is tantamount to detecting the most relevant frequency-bands along all time-segments. The interaction between those features must then be taken into account to properly analyze their discriminability contributions.

The GSDA method, just like traditional LDA, seeks to find a discriminant direction in which the linear separation between classes is maximized [25]. The feature selection and classification are simultaneously performed, aiming to obtain maximal class discriminability. A priori information about discrepancy between features is included into the model through appropriate ad-hoc anisotropic penalizers. Although complete details about GSDA can be found in [25], a brief summary of this method is presented below.

Let \mathbf{G} be, as before, the feature matrix, and \mathbf{Y} be a $n \times 2$ matrix of binary variables such that y_{ic} is an indicator of whether the i^{th} observation belongs to the c^{th} class ($c = 1, 2$). Let $\boldsymbol{\theta} \in \mathbb{R}^2$ be a score vector (which transforms binary variables into real ones by mapping the subspace $\mathbb{R}^{n \times 2}$ into \mathbb{R}^n , by $\mathbf{Y} \rightarrow \mathbf{Y}\boldsymbol{\theta}$), and $\boldsymbol{\beta} \in \mathbb{R}^q$ the discriminant vector. Then GSDA consists of solving the following regularized constrained least squares problem:

$$\begin{aligned} (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) &= \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^q, \boldsymbol{\theta} \in \mathbb{R}^2} \{ \|\mathbf{Y}\boldsymbol{\theta} - \mathbf{G}\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\mathbf{D}_1 \boldsymbol{\beta}\|_1 + \lambda_2 \|\mathbf{D}_2 \boldsymbol{\beta}\|_2^2 \}, \\ \text{s.t.} \quad & \frac{1}{n} \boldsymbol{\theta}^T \mathbf{Y}^T \mathbf{Y} \boldsymbol{\theta} = 1, \end{aligned} \quad (8)$$

where λ_1 and λ_2 are appropriately chosen positive regularization parameters, which balance the amount of sparsity and the number of correlated variables in the solution vector, respectively, and \mathbf{D}_1 and \mathbf{D}_2 are $q \times q$

positive definite matrices, appropriately constructed so as to point-wise weigh the solution as indicated by a priori discriminative information.

The solution to the optimization problem (8) can be approximated by iteratively alternating two steps (with an adequate initialization). Namely:

1. Keep $\boldsymbol{\theta}$ fixed and find $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^q} \{ \|\mathbf{Y}\boldsymbol{\theta} - \mathbf{G}\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\mathbf{D}_1\boldsymbol{\beta}\|_1 + \lambda_2 \|\mathbf{D}_2\boldsymbol{\beta}\|_2^2 \}. \quad (9)$$

2. Keep $\boldsymbol{\beta}$ fixed and find $\hat{\boldsymbol{\theta}}$:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^2} \|\mathbf{Y}\boldsymbol{\theta} - \mathbf{G}\boldsymbol{\beta}\|_2^2 \quad s.t. \quad \frac{1}{n} \boldsymbol{\theta}^T \mathbf{Y}^T \mathbf{Y} \boldsymbol{\theta} = 1. \quad (10)$$

The first step (Eq. (9)) is a generalized version of the well-known elastic-net (e-net) problem [26], which can be re-written by means of LASSO [22] if \mathbf{D}_1 is invertible. The GSDA is implemented with automatic regularization parameter selection [25]. This implementation makes use of the LARS-EN algorithm introduced in [27] in which a *stop* parameter, that sets an upper bound either for the ℓ_1 -penalizing term or for the number of non-zero values in the solution vector $\hat{\boldsymbol{\beta}}$, can be used for early stopping.

The GSDA method allows a priori discriminative information to be taken into account by the model. The discriminative power at each feature can be measured by means of a divergence measure. For that, we assume each one of the elements $g_{i,j}$ of the feature matrix \mathbf{G} is a realization of either one of two random variables whose distributions uniquely characterize each one the two MI conditions. For any fixed j^{th} feature, $j = 1, \dots, q$, let $f_1^j(n)$ and $f_2^j(n)$, $n \in \mathcal{N}$, be the probability functions (corresponding to an appropriate discretization \mathcal{N}) associated to MI classes 1 and 2, respectively. Then, the discrepancy d between both classes at the j^{th} feature can be measured by “distances” between these two probability functions. In particular, let us consider the Jensen-Shannon (JS) divergence [28], defined as:

$$\begin{aligned} d(j) &= d(f_1^j, f_2^j) \\ &\doteq H(\pi f_1^j + (1 - \pi) f_2^j) - \pi H(f_1^j) - (1 - \pi) H(f_2^j), \end{aligned} \quad (11)$$

where $H(\cdot)$ is the Shannon entropy $H(f) \doteq - \sum_{n \in \mathcal{N}} f(n) \log(f(n))$ [29] and π is a predefined parameter with $0 < \pi < 1$. For balanced binary classification problems it is natural to set $\pi = 1/2$, in that case the JS divergence defined by (11) is symmetric and its square root is a metric in the strict mathematical sense [28].

A large value of $d(j)$ given by Eq. (11) means that there is relevant discriminative information at feature j . Thus, since we wish to highlight those features containing a significant amount of discriminative information, matrices \mathbf{D}_1 and \mathbf{D}_2 in Eq. (8) can be constructed so as to strongly penalize those features which contains little or no discriminative information, while avoiding penalization in the remaining ones. Following these ideas, we

define the anisotropy matrices \mathbf{D}_1 and \mathbf{D}_2 as follows:

$$\mathbf{D}_1 \doteq \text{diag}(1 - \alpha_j + \alpha_j c_j), \quad (12)$$

$$\mathbf{D}_2 \doteq \text{diag}(c_j), \quad (13)$$

where: $c_j \doteq \frac{(\prod_{\ell=1}^q d(\ell))^{1/q}}{d(j)}$ and $\alpha_j \doteq \frac{\max_{1 \leq \ell \leq q} \{c_\ell\} - c_j}{\max_{1 \leq \ell \leq q} \{c_\ell\} - \min_{1 \leq \ell \leq q} \{c_\ell\}}$, for $j = 1, \dots, q$.

At the end of the procedure, the optimal discriminant vector $\hat{\beta}$, solution of (8), will contain a few non-zero entries (due to the ℓ_1 -norm penalization) which correspond to the most discriminant t - f -band CSP features. Finally, a simple linear classification rule can be constructed over the projected feature set $\mathbf{G}^* \doteq \mathbf{G}\hat{\beta} \in \mathbb{R}^n$.

2.2 DATASETS

In order to evaluate the applicability of our proposed method in the context of a BCI-based functional hand motor rehabilitation scenario, a self-acquired dataset of kinesthetic MI vs. rest condition was designed and recorded with healthy subjects. In addition, two publicly available datasets, widely used for the evaluation of new algorithms in the MI-BCI community ([19, 20, 21, 23, 24]), were used to test and compare our method in off-line cross-validation scenarios.

2.2.1 Dataset 1: MI vs. rest

This dataset was collected from eleven healthy subjects (3 females, 4 left-handed, mean age \pm SD = 25.45 \pm 2.50 years) without any previous BCI experience. The experiment was approved by the local ethics committee (BASEC-Nr. Req-2017-00631, Cantonal Ethics Commission, Zurich, Switzerland). For data acquisition, a transportable 64-channel EEG system (eegoTMrt Ant Neuro, Netherlands) was used. EEG data was collected using 64 surface electrodes placed in accordance with the international 10-20 system. All signals were recorded with CPz serving as reference and AFz as ground. The signals were sampled at 512 Hz and band-pass filtered between 0.5 Hz and 40 Hz and then downsampled to 128 Hz. In order to monitor that no actual hand movement was performed during MI tasks, electromyography (EMG) of two hand muscles and two forearm muscles (thenar eminence, first dorsal interosseous, flexor digitorum superficialis and extensor digitorum communis) were simultaneously measured with surface EMG electrodes (Noraxon TeleMyo DTS, Noraxon, USA).

The MI-BCI paradigm consisted of two conditions, namely the kinesthetic imagination of movement of the dominant hand (grasping movement) and rest/relax condition. The session was composed of four runs separated by short breaks. Each run consisted of 40 trials (20 for each of the two conditions), yielding a total of 160 trials at the end of the session. The maximum voluntary contraction (MVC), i.e. the maximum force which a person can produce in a specific isometric exercise, was recorded at the beginning of the session (for wrist extension, wrist flexion, thumb abduction and finger abduction). The EMG of each muscle was normalized with respect to the corresponding MVC to further analyze muscle activation during the EEG trials based on the maximum

EMG amplitude in each muscle. At the beginning of the session each subject was clearly instructed on the tasks, and asked to practice until they felt comfortable in performing motor imagery. To assess motor imagery ability, five kinesthetic items of the KVIQ-10 questionnaire [30] were administered. These items corresponded to trunk movement, as well as proximal and distal movements of the upper and lower limbs. The KVIQ-10 is a fast questionnaire which assesses the intensity of MI sensation on a five-point ordinal scale, ranging from 1 (“no kinesthetic sensation”) to 5 (“as clear as executing an action”). At the end of the questionnaire the ability of each subject in performing MI could be assessed according to his/her final score value (maximal value: KVIQ-10=25).

During the experiment, subjects were comfortably seated in front of a computer screen with both arms resting on a desk. The dominant hand with the attached EMG electrodes was placed inside a cardboard box (see Figure 2a) in order to prevent subjects from seeing their hand which could lead to visual rather than kinesthetic MI. The experimental protocol is schematically depicted in Figure 2b. Each trial began with a fixation cross ($t = -3$ s), followed by an audible beep cue two seconds later ($t = -1$ s) in order to get the subject’s attention. At the start of the task ($t = 0$ s), the subject was asked to imagine grasping movements of the dominant hand for a period of 4 s if a visual cue (red right arrow) appeared (i.e., MI condition). For the Rest condition, i.e. if no arrow was presented after the beep, the subject was asked to relax for the same duration. A randomly selected break-time (between 2.5 and 4.5 s) followed each trial in order to avoid subject routinization to the protocol timing. At the end of each run, a longer break (> 2 min) allowed the subject to relax.

2.2.2 Dataset 2: Dataset IVa of BCI competition III

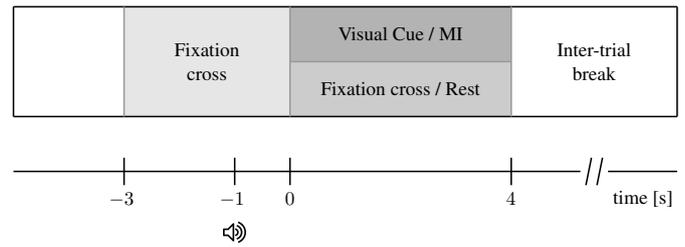
This dataset contains EEG recordings of five healthy subjects (aa, al, av, aw, ay) acquired with 118 electrodes at the extended international 10-20 system positions. Signals were band-pass filtered between 0.05 and 200 Hz. The signals were downsampled from 1000 Hz to 100 Hz. Each subject performed 280 trials of motor imagery of either their right hand or right foot (140 for each condition). The MI task was indicated by a visual cue and lasted 3.5 s. For further information about this dataset we refer the reader to [31].

2.2.3 Dataset 3: Dataset IIb of BCI competition IV

This dataset consists of EEG data from nine right-hand dominant healthy subjects of a study published in [32]. Three bipolar recordings (C3, Cz and C4) were recorded with a sampling frequency of 250 Hz. For each subject five sessions were performed. In each session, the subjects were asked to perform MI of either their left or right hand. In this work, similarly to [20, 21], the third session of the dataset (B0103T, B0203T, ..., B0903T) was used. A total of 160 EEG trials (80 per each MI condition) is available for each subject, in which the MI task lasted 4.5 s.



(a)



(b)

Figure 2: (a) experimental set-up showing both the EEG and EMG acquisitions systems, as well as the visual cue presented during a MI trial, for a left-handed subject, (b) schematic representation of one experimental trial with timing references in seconds.

2.3 DATA ANALYSIS

Our self-acquired dataset (dataset 1) contains EEG signals simultaneously recorded with EMG of four muscles involved in grasping movements. These EMG data were analyzed in order to verify whether the MI tasks were corrupted by actual movement. Besides visual inspection of the EMG signals, a standard approach for detecting muscle activation consists of determining the onset times at which the EMG signal exceeds a pre-defined threshold [33]. The envelope of the EMG signal was estimated by means of a lowpass 4th-order zero-phase Butterworth filter with 50Hz cut-off frequency applied to the zero-mean rectified EMG signals. Sliding windows of 500 ms along each EMG segment, relative to the corresponding MI trial, were taken. For each trial and each one of the four muscles considered, the trial was considered contaminated with EMG activity if the mean value of the sliding window was greater than 20% of the corresponding maximum MVC envelope value. For one of the subjects, more than 50% of the trials in one run were contaminated with EMG activations, reason for which this subject was not taken into account in the subsequent analysis. Thus, dataset 1 includes EEG recording of 10 subjects.

To prevent crosstalk from non-motor related regions, to reduce the dimensionality of the signals as well as the setup time for real-time applications, for datasets 1 and 2, only 28 electrodes covering the sensorimotor areas, in accordance to [34], were selected for further analysis.

The proposed PTFBCSP method in its two configurations, was tested with all three datasets. Since in literature it has been shown that SFBCSP provides the best overall classification accuracy as compared to both DFBCSP and to the original FBCSP methods [21], in this work comparative experiments were done against CSP and SFBCSP.

In order to separately analyze the impact of using, on the one hand, GSDA for feature selection and, on the other hand, multiple time windows, our proposed PTFBCSP method was first tested in its single time window version, PFBCSP. For the single time window analysis, for any EEG trial, EEG segments from 0.5 to 2.5 s after the onset of the visual cue were extracted. For fair comparison purposes, each EEG segment was band-pass filtered with a zero-phase Chebyshev type II filter between 4-40 Hz for the CSP algorithm, while for the filter bank methods (SFBCSP and PFBCSP), 17 sub-bands ($F = 17$) between 4 and 40 Hz with 4 Hz bandwidth and 2 Hz of overlap were applied ($fb_1 = 4 - 8$ Hz, $fb_2 = 6 - 10$ Hz, \dots , $fb_F = 36 - 40$ Hz). The proposed frequency band setting is based on the methodology proposed and validated in [20, 21]. The final number of CSP filter pairs was set to 1 ($K = 1$). Therefore, after applying CSP to each sub-band, a feature vector of dimension $q = 2KF = 34$ was constructed from each EEG segment. To avoid poor estimation of the covariance matrices (which generally occurs when too few training data are available), the CSP algorithm was implemented by automatic shrinkage covariance estimation as proposed in [35]. The dimension of the resulting feature matrix \mathbf{G} was $n \times 34$, where n is the number of available training samples.

The SFBCSP method was implemented as proposed in [21]. Thus, the LASSO optimization problem was solved by using the coordinate gradient descent algorithm [36] in which the regularization parameter λ is estimated by cross-validation on training data. A new matrix $\tilde{\mathbf{G}}$ of smaller dimension than \mathbf{G} was constructed from those columns of \mathbf{G} for which the sparse LASSO solution vector contained non-zero entries. The PFBCSP framework was implemented by automatic parameter selection with early stopping. In particular we chose the GSDA stop parameter equal to 8, i.e. $\|\hat{\beta}\|_0 \leq 8$. Note that while for SFBCSP the classification rule had to be learned from the optimized featured matrix $\tilde{\mathbf{G}}$, in the case of PFBCSP, the classification rule was applied using the new feature vector $\mathbf{G}^* = \mathbf{G}\hat{\beta} \in \mathbb{R}^n$, for which the classes are maximally separated by a linear function. As suggested in [19], the classification rule was implemented by means of the LDA classification algorithm.

The multiple time-windows analysis in the PTFBCSP method, is a straightforward extension of the single time-window frequency-band analysis described above. In this case, however, instead of using a single pre-defined EEG segment, a number T of EEG segment of 2 s with overlap of 1.5 s were extracted from each trial ($tb_1, \dots, tb_t, \dots, tb_T$) from 0 to T_{\max} s after the visual cue, where T_{\max} is the length of the MI trial. The window overlap of 1.5 s was selected in order to cover the entire length of an MI trial, optimize the number of temporal segments and include the specific 0.5-2.5 s time segment commonly used in the EEG-BCI literature. The multiple time-windows analysis together with the frequency band analysis resulted in a total of $q = 2KFT$ CSP features per trial ($q = 136$ for dataset 1, $q = 204$ for dataset 2 and $q = 170$ for dataset 3). In the case

of PTFBCSP the GSDA stop parameter was set to $4T$. The final classification performance of CSP, SFBCSP, PFBCSP and PTFBCSP was evaluated by using 10×10 -fold cross-validation. In order to statistically analyze the difference between the performances yielded by each method, one-way ANOVA and the Tukey-Kramer procedure with a level of significance $\alpha = 0.01$ were performed.

3 RESULTS

3.1 COMPARATIVE CLASSIFICATION RESULTS

Figures 3, 4 and 5 show the average classification results of the 10×10 cross-validation procedure achieved by CSP, SFBCSP, PFBCSP and PTFBCSP, for datasets 1, 2 and 3, respectively. Note that for all datasets both the PFBCSP and PTFBCSP methods show the highest accuracy. In particular, the use of multiple time segments together with filter bank analysis yields remarkable classification improvements for subjects 2 and 3 of dataset 1, subject 4 of dataset 2 and subject 1 of dataset 3. On the other hand, there are also some subjects (e.g. subject 3 of dataset 1, subject 3 of dataset 2 and subject 5 of dataset 3) for which this procedure does not improve classification performance.

The overall classification results across subjects for each dataset is summarized in Table 1. Best classification results are shown in bold. Note that both configurations of the proposed method outperform all the others, and all improvements were proved to be statistically significant at a level of $\alpha = 0.01$. For dataset 1, the single time window PFBCSP method yielded accuracy increments of 15.2% and 3.9%, as compared to CSP and SFBCSP, respectively. In the case of datasets 2 and 3, those increments were 11.6% and 2.3%, and 10% and 2.9%, respectively. In the case of the window-time varying PTFBCSP procedure, accuracy increments of 20.9% and 9.0%, were obtained as compared to CSP and SFBCSP, respectively. The corresponding values for dataset 2 were found to be 12.6% and 3.3%, while for dataset 3 those values were 11.4% and 4.3%, respectively.

In addition, since an improvement in the BCI performance due to subject-specific models usually implies higher computational cost, we have evaluated the average required computation time on a PC with Intel® Core™ i7-6700K and 64 GB RAM under Matlab 2017a for each method considered. In particular, for training the SFBCSP method, the average computational costs were found to be around 1, 0.6, and 1.5 s for datasets 1, 2 and 3, respectively, while for PFBCSP these values were found to be 0.7, 0.4 and 1.1 s. In the case of the proposed PTFBCSP method the training time was found to be 3.8, 2.5 and 4.2 s, for datasets 1, 2 and 3, respectively.

For our dataset, the ability of each subject in performing MI was measured by five kinesthetic items and ranged from “poor” (between 5 and 10) to “very good” (from 20 to 25) MI ability. In order to investigate the relation between the BCI classification accuracy and the reported MI ability, a pairwise correlation analysis was performed between the mean accuracy value reached by PTFBCSP for each subject (over the 10×10 cross-validation) and the corresponding KVIQ-10 score (sum of the 5 items considered). This analysis revealed no

statistically significant correlation ($R = -0.4, p\text{-value} = 0.25$).

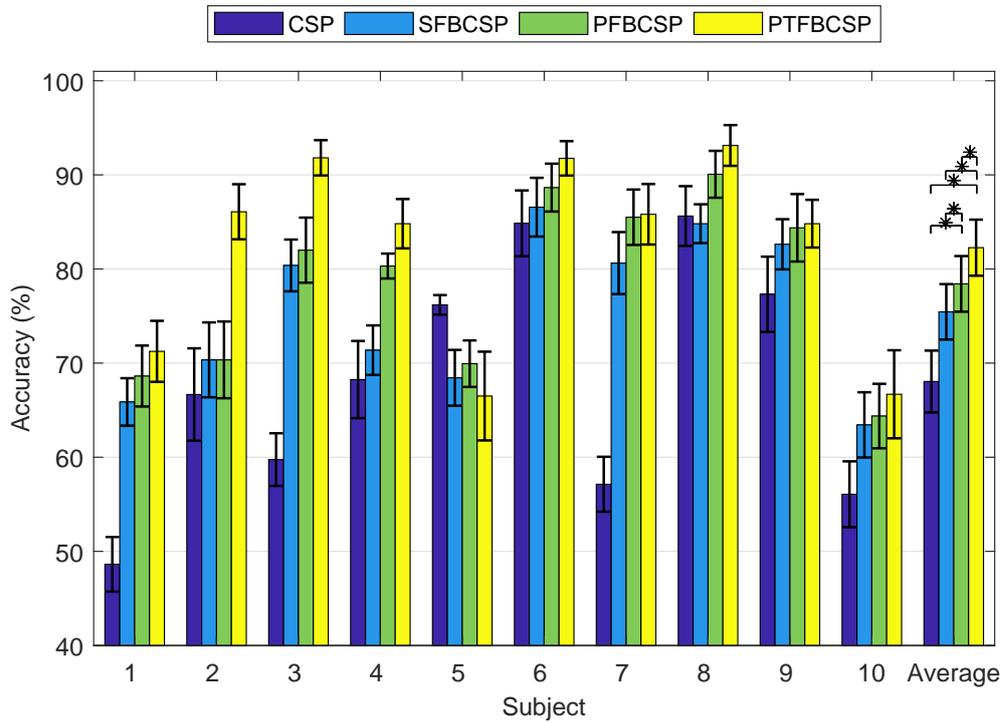


Figure 3: Average classification results on test data by 10×10 cross-validation achieved by CSP, SFBCSP, PFBCSP and PTFBCSP, for dataset 1. Here * denotes statistical significance difference between methods.

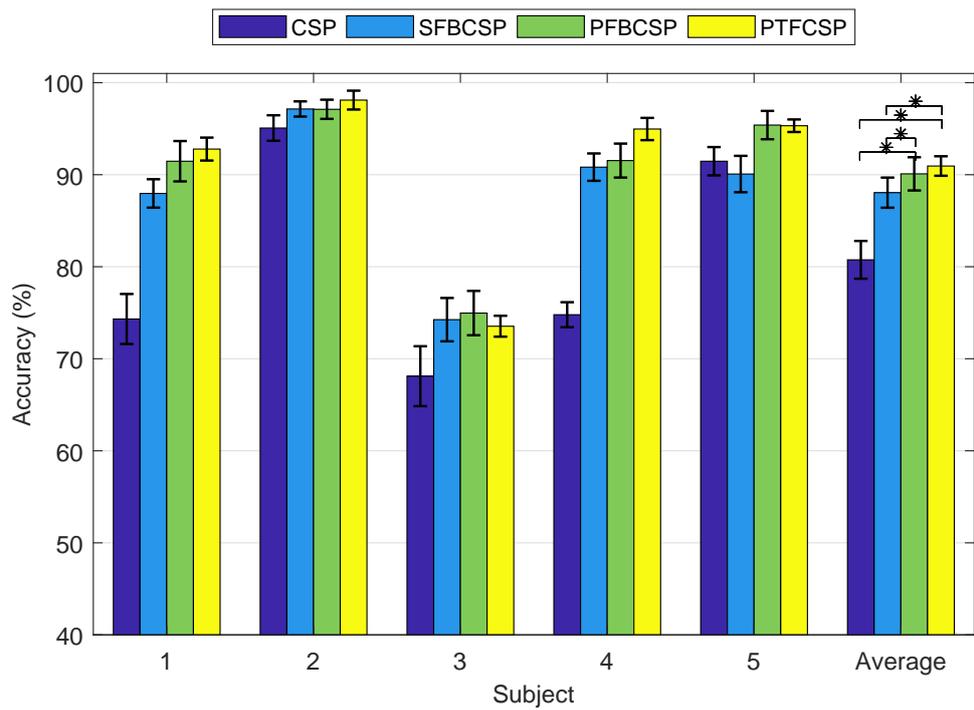


Figure 4: Average classification results on test data by 10×10 cross-validation achieved by CSP, SFBCSP, PFBCSP and PTFBCSP, for dataset 2. Here * denotes statistical significance difference between methods.

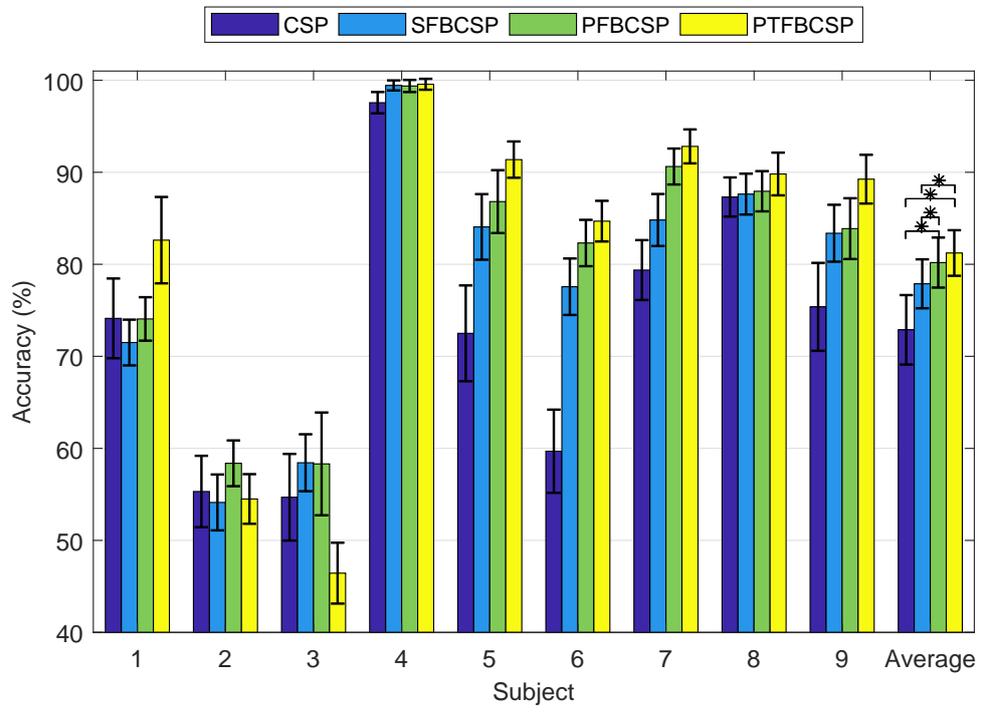


Figure 5: Average classification results on test data by 10×10 cross-validation achieved by CSP, SFBCSP, PFBCSP and PTFBCSP, for dataset 3. Here * denotes statistical significance difference between methods.

Dataset	CSP	SFBCSP	PFBCSP	PTFBCSP
1	68.05 (\pm 3.28)	75.45 (\pm 2.95)	78.42 (\pm 2.96)	82.26 (\pm 2.98)*
2	80.75 (\pm 2.05)	88.05 (\pm 1.64)	90.09 (\pm 1.81)	90.94 (\pm 1.06)*
3	72.88 (\pm 3.77)	77.89 (\pm 2.66)	80.19 (\pm 2.72)	81.23 (\pm 2.46)*

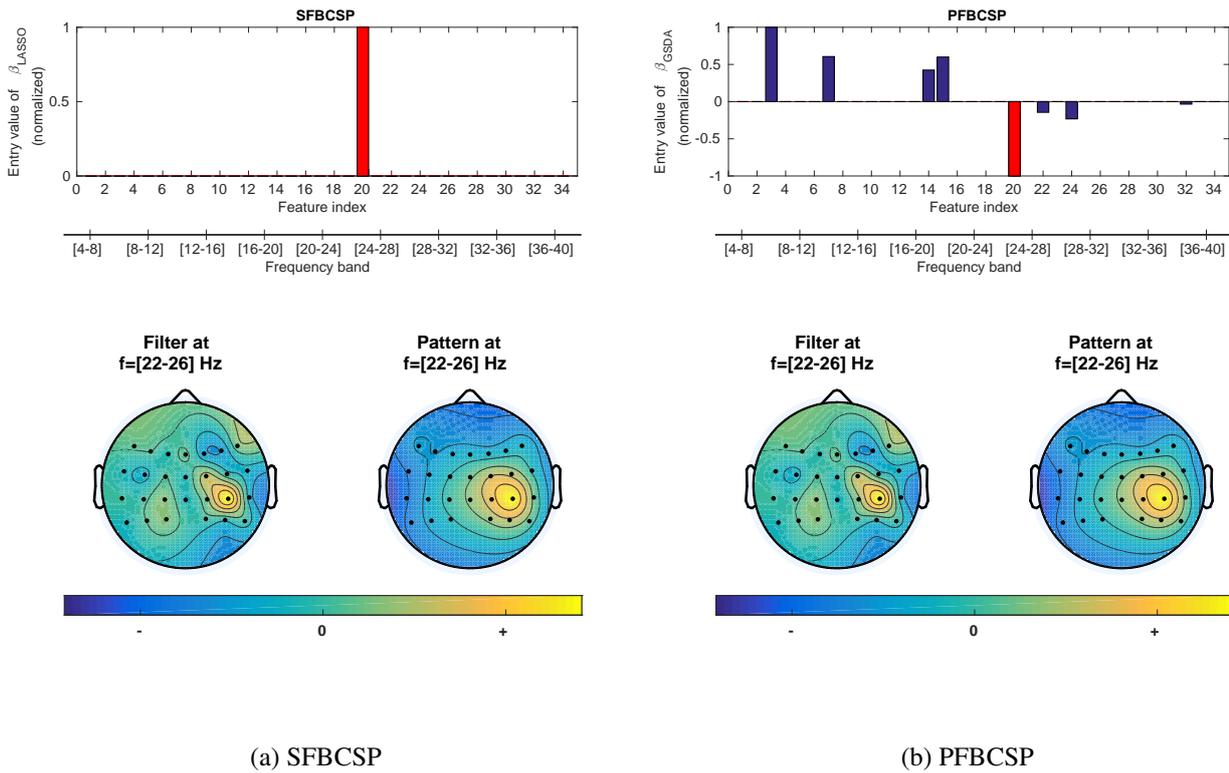
Table 1: Overview of the overall accuracy values and the corresponding standard deviation in parenthesis (in %) for each tested method and for each dataset. Best classification result are shown in bold. Here * denotes that PTFBCSP>SFBCSP with p -values < 0.01 .

3.2 TOPOGRAPHICAL MAPS OF THE SPATIAL TIME-FREQUENCY PATTERNS

The application of CSP allows for a neurophysiological interpretation of the resulting patterns [37]. The spatial patterns, i.e. the column vector of the matrix A introduced in Subsection 2.1.1, illustrate how the presumed sources are projected to the scalp, while (due to their weighting function) the spatial filters, used to project the original signals to optimally discriminate two classes with respect to variance, may resemble the patterns [17]. The proposed PTFBCSP method yields a solution vector with a small number of non-zero entries, which precisely represent the selected time-frequency band features. By plotting the topography maps of the most significant spatial filter and pattern associated to the most relevant feature, the neural correlates of the analyzed signals can be investigated. For a randomly selected trial of subject 2 of dataset 1, the selected features and the most significant spatial filter and pattern (associated to the most relevant frequency band) are represented in Figure 6, for the single time window methods (SFBCSP and PFBCSP). Similarly, for the multiple time windows method (PTFBCSP), the selected features at each time segment considered as well as the first two most significant spatial filters and patterns (associated to the first two most relevant features) are depicted in Figure 7.

4 DISCUSSION

In the present work a penalized time-frequency band common spatial pattern algorithm for efficiently decoding motor intention in EEG-based BCI systems was proposed. A standardized subject-specific method, which allows for multiple frequency band analysis in both single and multiple EEG time segments by considering a priori spatio-temporal-spectral information in a fast selection and classification procedure, was presented. Unlike most existing MI decoding methods, our approach provides an automatic way for choosing the “optimal” subject-specific time-frequency bands. The algorithm uses multichannel raw EEG data which are decomposed into T temporal segments and F frequency bands. By combining CSP and GSDA, the first spatial filter associated to each class is used for feature computation and then, feature selection and classification are simultaneously performed with automatic parameter selection. By subject-specifically optimizing both temporal and



(a) SFBCSP

(b) PFBCSP

Figure 6: Top: selected features for each of the single time window filter bank methods considered. Bottom: topographical maps of the most significant spatial filter and pattern corresponding to the most relevant feature (red bar) for a left-handed subject of dataset 1. a) SFBCSP, b) PFBCSP.

frequency bands, individual classification improvements of up to 20% as compared to current state-of-the-art MI decoding method, were achieved.

Two different configurations of the PTFBCSP method, for single and multiple time windows, were tested off-line. The overall classification results (Table 1) show, on the one hand, that the use of a priori discriminative information among features by GSDA (PFBCSP vs. SFBCSP) improves classification performance (up to 3.9%) and, on the other hand, that the use of multiple time windows (PTFBCSP vs. SFBCSP) further improves the efficiency for motor intention decoding (improvements above 3.3%), mainly when MI has to be detected in contrast to rest condition (dataset 1, improvements of 9%). Although statistically significant differences between PTFBCSP and PFBCSP were only found for dataset 1, the obtained results indicate a clear trend in the use of multiple time windows for improving MI detection, especially when the MI task has to be detected against a rest condition. Moreover, for some particular subjects (see figures 3, 4 and 5) the improvements obtained by PTFBCSP are above 10% (e.g. subject 2 of dataset 1, subject 1 of dataset 3). The variances along the 10×10 -fold cross-validation show that in some cases (dataset 2), PTFBCSP is more robust to changes in the training set than all the other single-time window methods considered (see standard deviation values in Table 1).

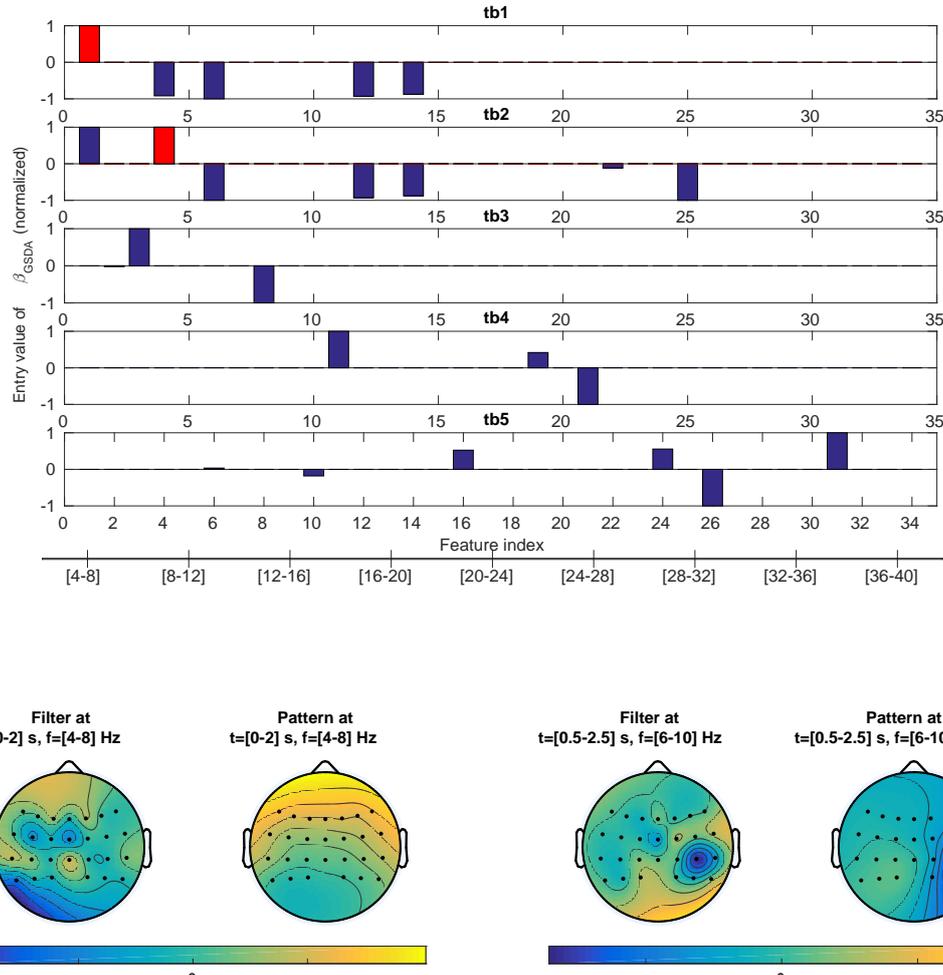


Figure 7: Top: selected features for the proposed PTFBCSP method at each time segment $tb_1 = 0 - 2s$, $tb_2 = 0.5 - 2.5s, \dots, tb_5 = 2 - 4s$ (arranged in rows). Bottom: topographical maps of the first two most significant spatial filters and patterns corresponding to the first two most relevant features (red bars) for a left-handed subject of dataset 1.

The sparse time-frequency segment common spatial pattern (STFSCSP) method [24] and the mutual information based selection of optimal spatial-temporal patterns (MIBS-OSTP) [23] are two state-of-the-art methods for MI classification based on multiple time windows and filter bank analysis. Unfortunately, we were unable to reproduce the results of these two methods (some signal processing steps used in the algorithms were not completely reported and the corresponding codes are not publicly available), being the reason for not having them included in our numerical experiments. However, the classification results obtained by PTFBCSP in the two publicly available datasets can be compared to those obtained by STFSCSP and MIBS-OSTP. In particular, in [23], by using MIBS-OSTP, the average classification result over 10×10 -fold cross-validation achieved with the BCI Competition IV Dataset IIb (dataset 3) was reported to be $78.25 \pm 1.12 \%$, while for the PTFBCSP

method the analogous result was 81.23 ± 2.46 %, corresponding to a classification improvement of about 3.8 %. A similar comparison can be made with respect to STFSCSP, which is applied after Fisher's linear discriminant criteria for automatic subject-specific channel selection. In fact, [24] reports an average classification result for the BCI competition III Dataset IVa (dataset 2) of 92.66 ± 4.44 %. Although the average accuracy value is slightly higher (presumably due to the channel selection procedure) than what is achieved by our PTFBCSP method (90.94 ± 1.06 %), the standard deviation is quite large, making PTFBCSP more robust.

The topography maps of the most relevant CSP pattern associated to the most relevant time-frequency feature allow for neurophysiological interpretations of the spatio-temporal-spectral patterns. For the single time window methods (Figure 6), the frequency band information is located between 22 and 26 Hz, while for the multiple time segments analysis (Figure 7) the most relevant frequency bands vary from time segment to time segment. It is remarkable how both single time filter bank methods select the same feature as the most relevant, which, by visual analysis of the spatial filter, clearly corresponds to the sensorimotor area involved in the MI task (the subject was left-handed). In the case of multiple-time windows, the first most informative frequency band is found between 4 and 8 Hz. This 4-8 Hz sub-band corresponds to the theta EEG rhythms, which are highly correlated with relaxation states [38]. For the second most informative frequency band, enclosing the mu and beta rhythms, the topography maps here again can be clearly associated with the MI condition. Therefore, the analysis of the topography maps of the most relevant features suggest that the multiple-time windows method can identify the most relevant information from both conditions.

Reliably detecting an MI task strongly depends on the subject's abilities to operate a BCI. It is known that between 15 and 30% of MI-BCI users fail to modulate their brain signals to effectively control a BCI [39]. This phenomenon, known as BCI-illiteracy, should be taken into account in order to avoid frustrating training procedures. To control for motor imagery performance, we applied the kinesthetic KVIQ-10 questionnaire. No statistically significant correlation with classification accuracy was found, which is in accordance with previous work [40], indicating that the reliability of such questionnaires to assess MI ability is low. In addition, and although for our self-acquired dataset the subjects were asked to perform cycling MI, the stimulation protocol did not present internal cues in order to ensure the beginning of a new MI task. This may cause beta-rebound effect within certain time segments which could perhaps negatively influence the proposed framework.

A limitation of this work is that the classifier was only tested off-line. Nevertheless, in light of the followed data processing steps and the low testing times (of only a few milliseconds), adapting the algorithm to an on-line implementation should be rather straightforward. Clearly, however, for these real-time scenarios, classification accuracy would need to be re-evaluated in the on-line context. In addition, for those datasets with a large number of channels, we selected 28 electrodes covering the sensorimotor areas. From the set-up time of the experiment and subject's comfort point of view, it is desired to find the minimal number of electrodes for efficient MI detection. An advantage of our method is that, independently from how the electrodes are selected,

it offers the flexibility to adapt (without any modification) to such changes in the data structure. This claim is supported by the classification results achieved with dataset 3 (above 81%), which relies on only 3 channels.

The current version of the method was evaluated by setting the same GSDA stopping parameter and the same electrode set for all the subjects and datasets. It is expected that the presented classification accuracies can be further improved by making these selection subject-dependent. Future works include the study of inference methods for subject-specific regularization parameters estimation such as Bayesian learning [41], the balancing principle [42] or aggregation [43]. In addition, regarding optimal channel selection, the approach known as “group LASSO” [44] can be included in our model for subject-specific selection of the best sub-set of EEG channels [45, 46]. Finally, real-time implementation of the proposed method must be studied, in order to conclude whether the trial duration is optimal, whether the multiple time segments is suitable for on-line feedback and whether the method is robust and accurate enough for establishing a good BCI communication.

5 CONCLUSION

A novel classification method combining multiple frequency bands and time windows to increase classification accuracy in MI-BCIs was proposed. Two configurations of the presented algorithm were tested with public datasets from BCI competitions as well as with our own dataset, showing better classification results when compared to other state-of-the-art time-frequency band CSP methods.

When MI has to be detected against a rest condition, the multiple-time windows framework on average improves MI detection as compared to the single time window methods. As a drawback it requires longer MI times after the onset of the cue. Hence, if faster detection speed is required, PFBCSP constitutes a good compromise between classification accuracy and time delay after the MI cue.

In addition to showing improvements with respect to state-of-the-art classification methods, several research directions for further increasing performance are highlighted. For instance, the subject-specific adaptation capabilities of the proposed method could be further improved by selecting the most relevant channels, time segments and hyperparameters for each subject. We consider the present work an important step towards improved motor intention detection, which could also find application in rehabilitation scenarios, e.g. in the form of an MI-BCI-triggered robotic hand orthosis.

6 ACKNOWLEDGEMENTS

This research was supported in part by the ETH Zurich Foundation in collaboration with Hocoma AG, the Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina, through PIP 2014-2016 No. 11220130100216-CO, the Air Force Office of Scientific Research, AFOSR/SOARD, through Grant FA9550-14-1-0130 and by the Universidad Nacional del Litoral, UNL, through CAI+D-UNL 2016 PIC No. 50420150100036LI.

The first author expresses her sincere gratitude to the Zeno Karl Schindler Foundation (ZKF, Geneva,

Switzerland) for financial support and for providing her with the opportunity to visit the Rehabilitation Engineering Laboratory at ETH Zurich in the context of a 6-months doctoral exchange program.

The authors have confirmed that any identifiable participants in this study have given their consent for publication. The authors also thank Christian Mancini for setting up part of the experimental protocol used to collect our self-acquired dataset.

REFERENCES

- [1] J. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: principles and practice*. Oxford University Press, USA, 2012.
- [2] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of neural engineering*, vol. 4, no. 2, 2007.
- [3] J. del R. Millán, P. W. Ferrez, F. Galán, E. Lew, and R. Chavarriaga, "Non-invasive brain-machine interaction," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 22, no. 05, pp. 959–972, 2008.
- [4] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalography and clinical neurophysiology*, vol. 78, no. 3, pp. 252–259, 1991.
- [5] A. Nijholt, D. P.-O. Bos, and B. Reuderink, "Turning shortcomings into challenges: Brain-computer interfaces for games," *Entertainment Computing*, vol. 1, no. 2, pp. 85–94, 2009.
- [6] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán, "A brain-actuated wheelchair: asynchronous and non-invasive brain-computer interfaces for continuous control of robots," *Clinical neurophysiology*, vol. 119, no. 9, pp. 2159–2169, 2008.
- [7] Y. Yu, Z. Zhou, Y. Liu, E. Yin, N. Zhang, Z. Wang, Y. Liu, X. Wu, D. Hu, *et al.*, "Self-paced operation of a wheelchair based on a hybrid brain-computer interface combining motor imagery and P300 potential," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2516–2526, 2017.
- [8] D. J. McFarland and J. R. Wolpaw, "Brain-computer interface operation of robotic and prosthetic devices," *Computer*, vol. 41, no. 10, 2008.
- [9] S. R. Soekadar, M. Witkowski, N. Vitiello, and N. Birbaumer, "An EEG/EOG-based hybrid brain-neural computer interaction (BNCI) system to control an exoskeleton for the paralyzed hand," *Biomedical Engineering/Biomedizinische Technik*, vol. 60, no. 3, pp. 199–205, 2015.
- [10] A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, Ö. Yilmaz, F. L. Brasil, G. Liberati, M. R. Curado, E. Garcia-Cossio, A. Vyziotis, *et al.*, "Brain-machine interface in chronic stroke rehabilitation: a controlled study," *Annals of neurology*, vol. 74, no. 1, pp. 100–108, 2013.
- [11] K. K. Ang, C. Guan, K. S. Phua, C. Wang, L. Zhou, K. Y. Tang, G. J. Ephraim Joseph, C. W. K. Kuah, and K. S. G. Chua, "Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke," *Frontiers in Neuroengineering*, vol. 7, p. 30, 2014.
- [12] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neurosc. letters*, vol. 239, no. 2, pp. 65–68, 1997.
- [13] R. P. Rao, *Brain-computer interfacing: an introduction*. Cambridge University Press, 2013.
- [14] W. Samek, M. Kawanabe, and K.-R. Muller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2014.
- [15] G. Pfurtscheller and F. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.
- [16] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [17] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [18] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (sbcs) for brain-computer interface," in *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on*, pp. 204–207, IEEE, 2007.

- [19] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390–2397, IEEE, 2008.
- [20] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain-computer interfaces," *IEEE transactions on biomedical engineering*, vol. 56, no. 11, pp. 2730–2733, 2009.
- [21] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface," *Journal of neurosc. methods*, vol. 255, pp. 85–91, 2015.
- [22] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [23] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition*, vol. 45, no. 6, pp. 2137–2144, 2012.
- [24] M. Miao, H. Zeng, A. Wang, C. Zhao, and F. Liu, "Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: An sparse regression and weighted naïve bayesian classifier-based approach," *Journal of neuroscience methods*, vol. 278, pp. 13–24, 2017.
- [25] V. Peterson, H. L. Rufiner, and R. D. Spies, "Generalized sparse discriminant analysis for event-related potential classification," *Biomedical Signal Processing and Control*, vol. 35, pp. 70–78, 2017.
- [26] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [27] K. Sjöstrand, L. H. Clemmensen, R. Larsen, G. Einarsson, and B. K. Ersbøll, "SpaSM: A matlab toolbox for sparse statistical modeling," *Journal of Statistical Software*, vol. 84, no. 10, 2018.
- [28] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991.
- [29] C. E. Shannon and W. Weaver, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 379-423, pp. 623–656, 1948.
- [30] F. Malouin, C. L. Richards, P. L. Jackson, M. F. Lafleur, A. Durand, and J. Doyon, "The kinesthetic and visual imagery questionnaire (kviq) for assessing motor imagery in persons with physical disabilities: a reliability and construct validity study," *Journal of Neurologic Physical Therapy*, vol. 31, no. 1, pp. 20–29, 2007.
- [31] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [32] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 4, pp. 473–482, 2007.
- [33] P. W. Hodges and B. H. Bui, "A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography," *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, vol. 101, no. 6, pp. 511–519, 1996.
- [34] S. Marchesotti, R. Martuzzi, A. Schurger, M. L. Blefari, J. R. del Millán, H. Bleuler, and O. Blanke, "Cortical and subcortical mechanisms of brain-machine interfaces," *Human brain mapping*, vol. 38, no. 6, pp. 2971–2989, 2017.
- [35] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces," *Proceedings of the IEEE, Institute of Electrical and Electronics Engineers*, vol. 103, no. 6, pp. 871–890, 2015.

-
- [36] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, p. 1, 2010.
- [37] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *Neuroimage*, vol. 87, pp. 96–110, 2014.
- [38] R. Hebert and D. Lehmann, "Theta bursts: an eeg pattern in normal subjects practising the transcendental meditation technique," *Electroencephalography and clinical neurophysiology*, vol. 42, no. 3, pp. 397–405, 1977.
- [39] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, "High theta and low alpha powers may be indicative of BCI-illiteracy in motor imagery," *PloS one*, vol. 8, no. 11, p. e80886, 2013.
- [40] M. L. Blefari, J. Sulzer, M.-C. Hepp-Reymond, S. Kollias, and R. Gassert, "Improvement in precision grip force control with self-modulation of primary motor cortex during motor imagery," *Frontiers in behavioral neuroscience*, vol. 9, p. 18, 2015.
- [41] Q. Li, N. Lin, *et al.*, "The bayesian elastic net," *Bayesian analysis*, vol. 5, no. 1, pp. 151–170, 2010.
- [42] K. Ito, B. Jin, and T. Takeuchi, "A regularization parameter for nonsmooth tikhonov regularization," *SIAM Journal on Scientific Computing*, vol. 33, no. 3, pp. 1415–1438, 2011.
- [43] J. Chen, S. Pereverzyev Jr, and Y. Xu, "Aggregation of regularized solutions from multiple observation models," *Inverse Problems*, vol. 31, no. 7, p. 075005, 2015.
- [44] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [45] R. Flamary, N. Jrad, R. Phlypo, M. Congedo, and A. Rakotomamonjy, "Mixed-norm regularization for brain decoding," *Computational and mathematical methods in medicine*, vol. 2014, 2014.
- [46] T. Yu, Z. Yu, Z. Gu, and Y. Li, "Grouped automatic relevance determination and its application in channel selection for P300 BCIs," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 1068–1077, 2015.