Neurocomputing 151 (2015) 122–129

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Time-series discrimination using feature relevance analysis in motor imagery classification

A.M. Alvarez-Meza *, L.F. Velasquez-Martinez, G. Castellanos-Dominguez

Universidad Nacional de Colombia, Signal Processing and Recognition Group, Campus La Nubia, km 7 via al Magdalena, Manizales, Colombia

ARTICLE INFO

Article history:
Received 1 February 2014
Received in revised form
28 June 2014
Accepted 10 July 2014
Available online 18 October 2014

Keywords:
Motor imagery
EEG data preprocessing
Short-time parameters
Feature relevance analysis

ABSTRACT

The use of motor imagery discrimination using feature relevance analysis (MIDFR) is investigated for classification tasks based on electroencephalography (EEG) signals. The method addresses the problem of a direct and automatic finding of the time-varying features influencing the most on distinguishing motor imagery tasks. The method introduces a stochastic relevance stage that is primarily used for properly handling the set of short-time features, which are extracted as to make prominent the nonstationary behavior of the EEG data. Furthermore, since it is widely accepted that the motor imagery information is concentrated in the μ and β neural activity bands, we make use of the empirical mode decomposition together with the common-spatial-patterns mapping. We test two different motor imagery databases, using a soft-margin support vector machine classifier that is validated by a 10-fold cross-validation methodology. Since the proposed MIDFR algorithm better encodes neural activity dynamics, experimental results carried out, which are also contrasted with other state-of-the art approaches, show that the proposed approach allows improving detection of MI classification tasks. Besides, the computed relevance on the EEG channels is in accordance with other clinical findings reported in the literature.

1. Introduction

Based on electrophysiological measures of brain functions, brain computer interfaces (BCI) are systems that use inferred information to provide new non-muscular channel for sending messages and commands to external devices [1]. The most commonly employed method for monitoring brain activity is the electroencephalography (EEG) implemented in several applications: neuropathological detection, analysis of cognitive behaviors, game controlling, among others. Also, the analysis of the human sensorimotor functions using EEG can help people with physical disability or degenerative neuropathologies (like locked-in syndromes or amyotrophic lateral sclerosis), where BCI systems are based on the cognitive neuroscience paradigm termed as motor imagery (MI) [2]. The MI relies on brain activity patterns of the imagination of a motor action, but without its physical implementation, e.g., the imagination of hand movements, whole body activities, relaxation, etc. [3–5].

In this sense, neural monitoring requires the identification of frequency bands (rhythms) mainly concentrating discriminant information about neural brain states. For instance, δ (0.5–4 Hz) rhythm is predominant in EEG recorded during deep sleep [6], whereas activity of θ (4–8 Hz) rhythm may occur in emotional or cognitive states [7]. For the concrete MI activities, both the μ (8–13 Hz) and β (13–30 Hz) waveforms are widely known for contributing the most to the motor activity classification since they are associated with those cortical areas directly connected to the brain's normal motor output channels [8,9]. Furthermore, movement or preparation for movement typically leads to decreasing activity of μ and β rhythms on the brain side contralateral to the movement [10]. Nevertheless, designing an MI-based BCI system requires an appropriate EEG data analysis to reach the needed performance for real-world BCI applications. Particularly, feature representation, feature selection and/or extraction, and classification methodologies must be suitably constructed to reveal the main MI patterns from EEG data.

About the preprocessing stage, common spatial pattern (CSP) is commonly used for extracting discriminative spatial patterns that contrast the power features of spatial patterns in different MI classes [9]. In fact, the CSP method constructs spatial filters that maximize the variance of a specific class, but it simultaneously minimizes the variance of the remaining classes. In order to achieve high classification accuracy, a pre-filtered broadband or subject-specific frequency bands are fixed to highlight the dynamics of interest. For finding the optimal spectra bands, several algorithms have been proposed, such as common spatio-spectral pattern, sub-band CSP and filter bank common spatial [11]. Nonetheless, the CSP preprocessing stage

* Corresponding author.
E-mail addresses: amalvarezme@unal.edu.co (A.M. Alvarez-Meza), lfvelasquezma@unal.edu.co (L.F. Velasquez-Martinez), cgcastellanosd@unal.edu.co (G. Castellanos-Dominguez).

http://dx.doi.org/10.1016/j.neucom.2014.07.077
0925-2312/

combined with the empirical mode decomposition (EMD) results in a powerful preprocessing tool for selecting informative frequency bands from EEG [12]. However, a direct and an automatic framework that allows finding those bands influencing the most on the MI performance remains still like an open issue.

Devoted to feature representation for BCI systems, different methods have been proposed in [13,8,14,15,4,16], among others: adaptive autoregressive (AAR) coefficients, Hjorth parameters, power spectral density (PSD), spatial filters like common spatial patterns (CSP), common spatial time–frequency patterns (CSTFP), and continuous and discrete wavelet transforms (CWT and DWT). Although many features may be extracted from the aforementioned methods, their main restriction is how to extract discriminative features as much as possible since several features may not contain relevant information introducing redundancy [17]. Consequently, extracted feature set may decrease classification accuracy. To cope with this drawback, there is a need to find attribute subset preserving, as much as possible, input data variability to allow identifying the most discriminant information, in terms of identifying different EEG data classes. On one hand, several approaches have been used to identify the relevance of the computed features in BCI systems, as given in [13,12,8,18]. Nevertheless, most of these feature selection methods are computationally expensive (mainly heuristics methods). In addition, most of them are based on supervised measures that can lead to getting overfitted results. Other approaches measure contribution of the rhythms in terms of their stochastic variability extracted from every considered short-time feature set to make prominent the nonstationary behavior of the EEG data [19]. However, there is a need for identifying the most discriminating features in MI discriminating task by finding a tradeoff between system complexity and accuracy [20,21].

This work discusses a time-series discrimination method based on feature relevance analysis within the MI classification framework, termed motor imagery discrimination using feature relevance analysis. This method initially includes CSP as a preprocessing stage to carry out the introduced feature relevance analysis approach that is based on an eigen-decomposition method to select a set of features best representing the MI process. Moreover, to deal with the non-stationary EEG data dynamics , the CSP filtering is carried out together with the empirical mode decomposition (EMD) aiming to highlight the μ and the β rhythms. Afterward, a set of short-time parameters are computed, from which the feature set is extracted. Short-time parameters are estimated using both, time-based and frequency-based, representations. We carry out the variability-based relevance analysis to address the problem of direct and automatic selection of the time-varying features influencing the most on distinguishing motor imagery tasks. This dimension reduction stage handles the multivariate EEG short-time rhythm representation within a subspace framework that searches for a projection maximally bearing input information when preserving only those data that contribute most to the brain activity classification, as discussed in [22]. Additionally, a soft-margin support vector machine (SVM) based classifier is trained, and the BCI system is validated by the commonly used 10-fold cross validation methodology. The remainder of this work is organized as follows: in Section 2, the theoretical background of proposed MIDFR is described. Experiments and results are presented in Section 3. Lastly, Section 5 outlines the work discussion and conclusions.

2. Methods

2.1. EEG data preprocessing

Let $\Psi=\{\boldsymbol{Y}_r : r=1,\ldots,R\in\mathbb{N}\}$ be a set of R raw EEG data trials, where $\boldsymbol{Y}_r\in\mathbb{R}^{C\times T}$ is the rth observed trial with $C\in\mathbb{N}$ being channels and $T\in\mathbb{N}$ being time samples. Let $\Upsilon=\{l_r\}$ be the class label set of Ψ, termed the MI paradigm condition, where $l_r\in\{-1,+1\}$. In order to achieve high classification accuracy, a pre-filtered broad band or subject-specific frequency bands are fixed to highlight the dynamics of interest. Here, we carry out a data-driven supervised decomposition of the EEG multi-channel data, termed *common spatial patterns* (CSP) [12]. The CSP calculates a spatial filter matrix $\boldsymbol{W}\in\mathbb{R}^{C\times C}$ by linearly projecting every EEG channel row vector $\boldsymbol{y}_r^c\in\mathbb{R}^T$ $(c=1,\ldots,C)$, located on the original sensor matrix $\boldsymbol{Y}_r$, to the projected surrogate sensor matrix $\widehat{\boldsymbol{Y}}_r\in\mathbb{R}^{C\times T}$. Thus, the CSP-based estimation is as follows:

$$\widehat{\boldsymbol{Y}}_r=\boldsymbol{W}^\top\boldsymbol{Y}_r,$$

Here the filter matrix, $\boldsymbol{W}$, has column vectors, $\boldsymbol{w}_c\in\mathbb{R}^C$, that are extracted from the MI condition covariance matrix set $\{\boldsymbol{\Sigma}^l\in\mathbb{R}^{C\times C} : l\in\{-1,+1\}\}$, where each lth matrix is calculated over all trails of the label class $l_r=l$, as follows:

$$\boldsymbol{\Sigma}^l=\mathbb{E}(\boldsymbol{y}_r^c\top\boldsymbol{y}_r^c : \forall c),\quad \forall l_r=l, \tag{1}$$

where notation $\mathbb{E}(\cdot)$ stands for the expectation operator. Herein, we denote each lth covariance as $\boldsymbol{\Sigma}^-$ and $\boldsymbol{\Sigma}^+$.

Then, CSP mapping can be obtained as the simultaneous diagonalization of the covariance matrices in Eq. (1), that is, $\boldsymbol{W}^\top\boldsymbol{\Sigma}^l\boldsymbol{W}=\boldsymbol{\Delta}^l$. In fact, $\boldsymbol{\Delta}^l\in\mathbb{R}^{C\times C}$ is a diagonal matrix with diagonal elements $\lambda_c^l\in\mathbb{R}^+$. Since $\boldsymbol{W}$ must be orthogonally rotated, the following constraint is also imposed: $\boldsymbol{\Delta}^+ + \boldsymbol{\Delta}^- = \boldsymbol{I}$, with $\boldsymbol{I}\in\mathbb{R}^{C\times C}$ being the identity matrix. As a result, the filter matrix can be computed by solving the generalized eigenvalue problem in the form: $\boldsymbol{\Sigma}^+\boldsymbol{w}_c=\lambda_c\boldsymbol{\Sigma}^-\boldsymbol{w}_c$, with $\lambda_c=\lambda_c^+/\lambda_c^-$ and $\lambda_c^+ + \lambda_c^- = 1$, where $\boldsymbol{\omega}_c$ holds each cth basis decomposition set in $\boldsymbol{W}$. Here, λ_c^l is the variance of the cth surrogate channel for the lth condition. Therefore, the larger the λ_c^l value – the higher the variance that is achieved by the corresponding spatial filter $\boldsymbol{w}_c$ for the lth condition [23].

However, authors in [12] state that the use of empirical mode decomposition (EMD) together with the CSP mapping enhances the classification accuracy since EMD is able to adaptively extract components carrying MI information that better fits for the frequency band selection needed in the CSP method. Therefore, we use the EMD as a filter-bank strategy to extract brain activity bands of interest from the CSP preprocessed EEG trial matrix $\widehat{\boldsymbol{Y}}$. Particularly, the EMD algorithm is performed over each row vector $\widehat{\boldsymbol{y}}^c\in\mathbb{R}^T$ in $\widehat{\boldsymbol{Y}}$, yielding the residual and intrinsic modes of $\widehat{\boldsymbol{y}}^c$ computed as

$$\widehat{\boldsymbol{y}}^c=\sum_{n=1}^{N}\boldsymbol{\zeta}_n+\boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{\zeta}_n\in\mathbb{R}^T$ stands for every intrinsic mode function (IMF), $\boldsymbol{\epsilon}\in\mathbb{R}^T$ is the residual, and $N\in\mathbb{N}$ is the number of computed IMFs. In Eq. (2), EMD iteratively estimates each zero-mean IMF amplitude based on the first-order derivative criterion until the residual value asymptotically becomes a small constant, when no more IMF terms can be further decomposed. Therefore, the main frequency bands of $\widehat{\boldsymbol{Y}}$ are extracted as the first $N_l<N$ IMFs of the vector $\widehat{\boldsymbol{y}}^c$ as

$$\boldsymbol{z}^c=\sum_{n=1}^{N_l}\boldsymbol{\zeta}_n, \tag{3}$$

where $\boldsymbol{z}^c\in\mathbb{R}^T$ is a row vector representing the cth EEG channel in the preprocessed data matrix $\boldsymbol{Z}\in\mathbb{R}^{C\times T}$.

2.2. Extraction of short-time feature set

For discriminating MI paradigms, we extract a long-term feature set from the preprocessed EEG data matrix $\boldsymbol{Z}$, where each data trial is quantified as a single vector. Nonetheless, used feature extraction methods must accentuate the spectral band of interest in terms of MI classification. Particularly, the imagination and the

execution of tracking movements are associated with task-related power changes, which are frequently concentrated in the μ (8–13 Hz) and β (13–30 Hz) rhythms and measured over the sensorimotor cortex [9,10,24].

Throughout this study, we consider three representative feature extraction methods that according to [25] have been widely applied in many motor imagery tasks.

Power spectral density parameters: We initially estimate the power spectral density (PSD) of the preprocessed channel $\boldsymbol{z}$, noted as $\boldsymbol{s} \in \mathbb{R}^{N_B}$, where $N_B \in \mathbb{N}$ is the number of frequency bins that is fixed according to the spectral band of interest, where the most discriminative information for MI is concentrated. Provided the EEG sample frequency $F_s \in \mathbb{R}^+$, the PSD vector $\boldsymbol{s} = \{s_f : f = 1, \ldots, N_B\}$, with $s_f \in \mathbb{R}$ and $N_B = \lfloor F_s/2 \rfloor$, is estimated by means of the nonparametric Welch's method that calculates the widely known fast Fourier transform algorithm of a set of $M \in \mathbb{N}$ overlapping segments, which are split from the preprocessed EEG data vector. Due to the non-stationary nature of EEG data, the piecewise stationary analysis is carried out over the set of extracted overlapping segments that are further windowed by a smooth-time weighting window $\boldsymbol{\alpha} \in \mathbb{R}^L$ that lasts $L \in \mathbb{N}$ $(L < T)$, yielding a set of the windowed segments $\{\boldsymbol{v}^m \in \mathbb{R}^L : m = 1, \ldots, M\}$, where $v_i^m \in \mathbb{R}$ $(i = 1, \ldots, L)$ is the ith element of $\boldsymbol{v}^m$. Consequently, the following modified periodogram vector $\boldsymbol{u} = \{u_f \in \mathbb{R}^+ : f = 1, \ldots, N_B\}$, $\boldsymbol{u} \in \mathbb{R}^{N_B}$, is computed based on the discrete Fourier transform as follows:

$$u_f = \sum_{m=1}^{M} \left| \sum_{i=1}^{L} v_i^m \exp(-j2\pi i f) \right|^2.$$

Then, each PSD element is computed as $s_f = u_f/(M\nu)$, being $\nu = \mathbb{E}(|\alpha_i|^2 : \forall i \in L)$.

Wavelet-based parameters: Wavelet-based methods have been heavily exploited in MI research, e.g., [26,13] to capture the spectral dynamics of EEG trials that usually holds non-stationary spectral components. In this work, both the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT) are carried out.

As regards the former decomposition, such inner-product-based transformation quantifies similarity between a given equally sampled time series at time spacing $\delta_t \in \mathbb{R}$ and a previously fixed base function $\gamma(\eta)$, termed *mother wavelet* ruled by a dimensionless parameter vector $\eta \in \mathbb{R}$. Namely, each time element of the CWT vector $\boldsymbol{\sigma}^g \in \mathbb{C}^T$ is extracted from the preprocessed EEG time-series $\boldsymbol{z}$ at scale $g \in \mathbb{R}$ by accomplishing their convolution with the scaled and shifted mother wavelet in the form

$$\boldsymbol{\sigma}_t^g = \sum_{\tau=1}^{T} z_\tau \gamma^*((\tau - t)\delta_t/g), \tag{4}$$

where * notes the complex conjugate. To build a picture showing amplitude variations through time in Eq. (4), both procedures of the Wavelet scaling g and translating through the localized time index $t \in T$ are used. As a result, the extracted wavelet coefficients provide a compact representation pinpointing energy distribution of the EEG data in time and frequency domains.

In turn, the DWT effectively addresses the trade-off between time and frequency resolution in nonstationary signal analysis. DWT also provides multi-resolution and non-redundant representation by decomposing the considered time-series into a number of sub-bands at different scales, yielding more precise time–frequency information about $\boldsymbol{z}$ [14]. Aiming to extract suitable time–frequency information from the DWT, the following detail vector $\boldsymbol{b}^j \in \mathbb{C}$ at level j is defined [27]:

$$b_t^j = \sum_{k \in \mathbb{Z}} a_{j,k} \psi_{j,k}(t), \tag{5}$$

where $a_{j,k} = \sum_{t \in T} z_t h_{j,k}(t)$, with $a_{j,k} \in \mathbb{C}$, $h_{j,k}(t) \in \mathbb{C}$ is the impulse response of a given wavelet filter. Then, provided the wavelet $\psi(\cdot)$, the DWT-based decomposition of $\boldsymbol{z}$ is computed as $z_t = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} a_{j,k} \psi_{j,k}(t)$.

Hjorth parameters: For each windowed segment $\boldsymbol{v}^m$ a set of the following three parameters describe the EEG signal on the time domain:

- *Activity*, $\boldsymbol{\sigma}_{\boldsymbol{v}}^2 \in \mathbb{R}^M$, where each mth element is directly described by the signal power variance:

$$\sigma_m^2 = \mathrm{var}(\boldsymbol{v}^m). \tag{6}$$

- *Mobility*, $\boldsymbol{\phi}_{\boldsymbol{v}} \in \mathbb{R}^M$, with elements measuring the signal mean frequency

$$\phi_m = \sqrt{\mathrm{var}(\boldsymbol{v}^{m'})/\mathrm{var}(\boldsymbol{v}^m)}, \tag{7}$$

being $\boldsymbol{v}'$ the derivative of $\boldsymbol{v}$.

- *Complexity*, $\boldsymbol{\vartheta}_{\boldsymbol{v}} \in \mathbb{R}^M$, with elements measuring frequency variations as the deviation of the signal from the sine shape, that is

$$\vartheta_m = \phi_m'/\phi_m. \tag{8}$$

Lastly, once all the aforementioned short-time parameters are computed, several of their statistical measures are considered to extract the input feature space matrix $\boldsymbol{X} = \{\boldsymbol{x}_r : r = 1, \ldots, R\}$, with $\boldsymbol{X} \in \mathbb{R}^{R \times D}$. Hence, the row vector $\boldsymbol{x}_r \in \mathbb{R}^D$ holds $D = C \times Q$ concatenated features for the rth MI trial, being $Q \in \mathbb{N}$ the number of provided features of the rth trial for a given channel.

2.3. Stochastic relevance analysis of extracted feature set

In practice, the provided feature space matrix $\boldsymbol{X}$ reaches a very huge dimension being necessary to identify the most discriminating features by finding a tradeoff between system complexity and accuracy [20,21]. To this end, we measure contribution of the rhythms in terms of their stochastic variability extracted from every considered short-time feature set to make prominent the nonstationary behavior of the EEG data. Moreover, we carry out the variability-based relevance analysis that handles the multivariate EEG short-time representation within a subspace framework that searches for a projection maximally bearing input information when preserving only those data that contribute most to the brain activity classification, as developed in [22]. Particularly, the stochastic feature set can be written as a linear combination of $D' < D$ independent basis functions where the minimum mean squared-based error is assumed as the evaluation measure of the subspace-based linear transformation. Thus, the set of orthogonal vectors is estimated so that the D' resulting components can approximate each input feature in $\boldsymbol{X}$ in such a way that EEG data information is maximally preserved.

Therefore, provided a set of features $\Xi = \{\boldsymbol{\xi}_d : d = 1, \ldots, D\}$, where $\boldsymbol{\xi}_d \in \mathbb{R}^R$ corresponds to each column of the input data matrix $\boldsymbol{X}$, the relevance of $\boldsymbol{\xi}_d$ can be measured by computing the following variability vector $\boldsymbol{\rho} \in \mathbb{R}^D$ [28,29]:

$$\boldsymbol{\rho} = \mathbb{E}(|\rho_d \boldsymbol{\alpha}_d| : \forall d \in D' \leq D), \tag{9}$$

where $\rho_d \in \mathbb{R}^+$ and $\boldsymbol{\alpha}_d \in \mathbb{R}^D$ are respectively the eigenvalues and eigenvectors of the covariance matrix estimated as $\boldsymbol{X}^\top \boldsymbol{X}/D$.

The main assumption behind the relevance measure introduced in Eq. (9) is that the largest values of ρ_d should point out to the better input attributes since they exhibit higher overall correlations to the estimated principal components. The D' value is fixed as the number of dimensions needed to preserve some percentage of the input data variability. As a result, the calculated relevance vector $\boldsymbol{\rho}$ is employed to rank the original features.

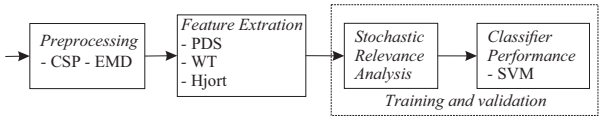

Fig. 1. Sketch of the proposed motor imagery discrimination using feature relevance analysis.

3. Experimental set-up

Fig. 1 shows the main sketch of the proposed motor imagery discrimination using feature relevance (MIDFR) analysis. Owing to highlight the latent patterns in $\boldsymbol{Y}_r$, we propose a time-series discrimination methodology appraising the following stages: (i) EEG preprocessing, (ii) feature extraction, and (iii) feature relevance analysis and classification.

3.1. BCI databases

In order to assess the proposed MIDFR methodology as a tool for supporting the development of BCI systems, we carry out experimental testing using two well-known MI databases that were used in the BCI competition IV (2008) and collected under the cognitive neuroscience paradigm in the concrete case of imagination of the hand movements. Specifically, two MI classes were measured for each subject: left hand and right hand (side chosen by the subject) [3].

Dataset `D1`: This EEG data collection (labeled as dataset I in the contest), provided by the Berlin brain–computer interface group,[1] contains signals obtained from seven subjects. For each one, the signals at 59 EEG positions are measured, being the sensorimotor area the most densely covered with the electrodes. Signals are bandpass filtered between 0.05 and 200 Hz and then digitized at 1000 Hz. Further, the database is down-sampled at $F_s = 100$ Hz, but previously an order 10 low-pass Chebyshev II filter is employed, having stop-band ripple 50 dB down and stop-band edge frequency 49 Hz. The whole MI session is performed without feedback. So, the database holds 100 repetitions of each MI class per person. Particularly, the EEG segments are extracted while a cue (indicating a side) is presented, i.e., an arrow pointing left or right is shown on a screen. The duration of each extracted segment is 4 s during which the observed subject is instructed to perform the cued MI task. These periods are interleaved with 2 s of blank screen and 2 s with a fixation cross shown in the center of the screen. All EEG channels per subject of the above-mentioned MI dataset are employed in our experiments.

Dataset `D2`: This EEG data collection (Dataset 2B) is provided by the Institute for Knowledge Discovery at the Graz University of Technology and contains two-choice trials, that is, MI of both the left and right hands [10]. The EEG signals are measured in nine subjects with three bipolar recordings (C3, Cz, and C4) during five performed sessions. The signals were sampled at $F_s = 250$ Hz and bandpass-filtered between 0.5 and 100 Hz. In the first two sessions, data were measured without feedback while the last three sessions were recorded including feedback. Each of the first two sessions were made up of six runs with 10 trials per class, resulting in 120 repetitions per each MI class. The remaining three sessions had four runs with 20 trials per class, i.e., 160 repetitions.

In this work, training and validation data of the first three sessions are employed to get 400 trials (240 without feedback and 160 with feedback). From the original EEG recordings, just the labeled segments (MI of right and left hands) are extracted to be used in this work. Each extracted segment duration is 5.5 s (feedback) and 5.25 s (without feedback).

[1] http://bbci.de/competition/iv/desc_1.html.

3.2. Extraction of the MIDFR feature set

In accordance with the EEG preprocessing stage described in Section 2.1, we perform the EMD-based CSP filtering procedure where we fix the value $N_I = 3$ since the first three IMFs accentuate MI information concentrated in both considered brain activity bands, as suggested in [12]. Based on the same band of interest, the segment length value L needed during calculation of the PSD and Hjort parameters is adjusted as $L > F_r/F_s$, being $F_r = 8$ Hz the smallest considered frequency [30]. Accordingly, the calculated PSD features are the norm and the two first statistical moments of $\boldsymbol{s}$ computed in both brain activity bands (noted as $\boldsymbol{s}_\mu$ and $\boldsymbol{s}_\beta$, respectively).

For computation of the WT-based feature subset, a suitable wavelet function must be used to optimize the classifier performance. We select the Morlet wavelet for the CWT analysis because its wave shape and EEG signals are alike and it allows extracting features better localized in the frequency domain [31,32]. Thus, we extract the short-time instantaneous CWT amplitudes using a couple of Morlet wavelets: one centered at 10 Hz (to extract the μ band) and another at 22 Hz (β band) as given in [33]. Provided the sample frequency F_s, we compute each WT feature vector in Eq. (4) to include both brain activity rhythms, resulting in the values $g_\mu = 60$ and $g_\beta = 27$ for the database `D1`, while $g_\mu = 100$ and $g_\beta = 70$ for `D2`.

Likewise, although there is a large selection of mother functions available, we test the Symlet wavelet (Sym-7) that is closely associated with the electrical brain activity and proved to be appropriate in similar applications [14,34]. Provided again the sample frequency F_s, we compute the detail coefficient vector $\boldsymbol{b}^j$ in Eq. (5) as to include both μ and β rhythms, resulting in the second and third levels for the database `D1`, while the third and fourth levels are selected for the database `D2`.

Among many amplitude estimators suggested for neurological activities [35,13,8], we extract the input feature set from all computed short-time parameters as shown in Table 1. As a result, the input feature space matrix $\boldsymbol{X}$ holds dimension $R = 200$ and $D = 1593$ for the database `D1`, while $R = 400$ and $D = 81$ for the database `D2`.

3.3. Classifier training and validation

Prior to classification, the feature relevance analysis is performed as stated in Section 2.3, where the number of dimensions D' in PCA is calculated as to have 95% of the explained variance. Afterward, the estimated relevance vector $\boldsymbol{\rho}$ is employed to rank the original features. Further, as the classification stage, the soft-margin support vector machine (SVM) classifier is trained. Here, a Gaussian kernel with a priori fixed bandwidth, $\theta \in \mathbb{R}^+$, is employed to estimate the similarity between two row vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ (both vectors immersed in $\boldsymbol{X}$) as follows:

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2}{2\theta^2}\right), \tag{10}$$

Then, we calculate the performed accuracy curve of the MI classification through the 10-fold cross validation scheme, adding one by one the features ranked by amplitude eigen-values of $\boldsymbol{\rho}$. On this calculated curve, we search for the optimal working point to fix the SVM regularization parameter $\pi \in \mathbb{R}^+$ and the kernel bandwidth θ value. Consequently, the π value is selected from the set {1, 10, 100, 1000}, while the θ value within the set $\{\theta_s, 10\theta_s, 100\theta_s, 1000\theta_s\}$. Here, the value $\theta_s \in \mathbb{R}^+$ is computed in accordance with Sylverman's rule

$$\theta_s = 0.9 \min(\mathbb{E}(\mathrm{std}(\Xi)), (1/1.34)\mathbb{E}(\mathrm{iqr}(\Xi)))$$

where notations std(·) and iqr(·) stand for the standard deviation and the interquartile range of a provided set of features, respectively.

Table 1
Amplitude estimators used as extracted features of EEG signals. Notation $\|\cdot\|_2$ stands for the 2-norm. # Feat/ch – the number of features per channel.

Parameter	*Features*			# Feat/ch
PSD	$\|\boldsymbol{s}_\mu\|_2^2$	$\mathbb{E}(\boldsymbol{s}_\mu)$	$\mathrm{var}(\boldsymbol{s}_\mu)$	6
	$\|\boldsymbol{s}_\beta\|_2^2$	$\mathbb{E}(\boldsymbol{s}_\beta)$	$\mathrm{var}(\boldsymbol{s}_\beta)$	
Hjort	$\max(\boldsymbol{\sigma}_{\boldsymbol{v}}^2)$	$\mathbb{E}(\boldsymbol{\sigma}_{\boldsymbol{v}}^2)$	$\mathrm{var}(\boldsymbol{\sigma}_{\boldsymbol{v}}^2)$	9
	$\max(\boldsymbol{\phi}_{\boldsymbol{v}})$	$\mathbb{E}(\boldsymbol{\phi}_{\boldsymbol{v}})$	$\mathrm{var}(\boldsymbol{\phi}_{\boldsymbol{v}})$	
	$\max(\boldsymbol{\vartheta}_{\boldsymbol{v}})$	$\mathbb{E}(\boldsymbol{\vartheta}_{\boldsymbol{v}})$	$\mathrm{var}(\boldsymbol{\vartheta}_{\boldsymbol{v}})$	
CWT	$\max(\lvert\boldsymbol{\sigma}^{g_\mu}\rvert)$	$\mathbb{E}(\lvert\boldsymbol{\sigma}^{g_\mu}\rvert)$	$\mathrm{var}(\lvert\boldsymbol{\sigma}^{g_\mu}\rvert)$	6
	$\max(\lvert\boldsymbol{\sigma}^{g_\beta}\rvert)$	$\mathbb{E}(\lvert\boldsymbol{\sigma}^{g_\beta}\rvert)$	$\mathrm{var}(\lvert\boldsymbol{\sigma}^{g_\beta}\rvert)$	
DWT	$\max(\lvert\boldsymbol{b}^{j_\mu}\rvert)$	$\mathbb{E}\left(\lvert\boldsymbol{b}^{j_\mu}\rvert\right)$	$\mathrm{var}(\lvert\boldsymbol{b}^{j_\mu}\rvert)$	6
	$\max(\lvert\boldsymbol{b}^{j_\beta}\rvert)$	$\mathbb{E}\left(\lvert\boldsymbol{b}^{j_\beta}\rvert\right)$	$\mathrm{var}(\lvert\boldsymbol{b}^{j_\beta}\rvert)$	
Total features per channel				27

4. Results

Fig. 2(a) and (c) shows the normalized average feature relevance vector estimated over all subjects for both considered databases, D1 and D2, respectively. As seen, features based on PSD, CWT, and DWT parameters supply higher relevance values than the Hjorth ones do. This fact may be explained since these representation methods allow extracting features more accurately from the rhythm bands with relevant information in terms of MI classification. In contrast, the Hjorth feature set reaches lower relevance values because they only consider second-order statistics that are not enough to encode the non-stationarity EEG dynamics within the MI discrimination framework.

The proposed MIDFR also appraises the channel distribution of the average relevance information as seen in Fig. 2(b) where several channels in the cortex area (*F2*), *F3*, *FCz* and *FC6*) concentrate discriminant information for dataset D1. This finding is in accordance to some clinical studies showing that the prefrontal cortex plays a role in holding sensory [36]. Specifically, anatomical studies in nonhuman primates have shown that the dorsal premotor area (PMdc), supplementary motor area (SMAc), and some regions in the anterior cingulate cortex directly project to the primary motor area (M1) and to the spinal cord. Therefore, it is likely that these caudal premotor areas are related to the movement generation. Besides, high relevance channel distribution values are obtained for *CFC* channels around the primary motor cortex (PMC). In this sense, individual subjects may show PMC activity during MI depending on their thinking strategy. Namely, activity associated to the task performance for the imagery mode is localized in the precentral sulcus, pointing out on the significance of this region in MI tasks. In addition, high relevance values are obtained from the parietal cortex, however, activity in the anterior parts of the parietal cortex (i.e., *CP* electrodes) most likely reflects somatosensory-motor association and sensory feedback from the movement mode. Moreover, detailed neuropsychological examination supports the role of the parietal cortex in generating mental movement representations. To illustrate, the posterior part of the parietal cortex, including the precuneus (*P* and *PO* electrodes), has been reported to be active during tasks involving motor imagery [36]; this finding matches the relevance channel configuration estimated by the proposed MIDFR method. Nonetheless, there are also some electrodes (*AF*, *F*, and *FC*) that get mostly low relevance channel distribution values.

In the case of the dataset D2, channel *C4* captures the salient variability from the MI paradigm, as seen in Fig. 2(d). Nonetheless, since only three channels are employed for discriminating the MI problem, further analysis about the channel relevance distribution is not feasible.

The results of the classifier validation on both tested datasets are shown in Fig. 3, where the MI discrimination accuracy is provided as a function of the number of chosen features that are according to the proposed relevance analysis approach. As seen, when adding features ranked in accordance with the proposed stochastic relevance analysis, discrimination performance, on average, raises till certain value. Moreover, performed accuracy tends asymptotically to a limit value (100 and 20 relevant features for D1 and D2, respectively), meaning that summing more features is not useful anymore in terms of the achieved classifier performance. However, in some cases, the MI performance has several local minima. This behavior may be explained by the fact that some features may hold highly redundant information.

For the sake of comparison, the MIDFR classifier performance is contrasted against several benchmark methods that test both databases, D1 and D2, in similar conditions. In the case of D1, Table 2 shows the best MI classification accuracy obtained per subject. As seen, the MIDFR approach reaches the highest accuracy along with the method given in [16] that involves common spatio-time–frequency patterns to design the time windows that are adopted for the MI task. It is worth noting that the MI discrimination procedure described in [12] also includes the EMD-based CSP preprocessing. However, the used adaptive frequency band selection together with the developed strategy of feature extraction are not enough, resulting in a low classification performance with a high standard deviation value. Another salient approach given in [9] is based on a robust learning mechanism that extracts and selects spatio-spectral features for differentiating multiple EEG classes. The algorithm also employs a non-linear regression and post-processing technique for predicting the time-series of class labels from the spatio-spectral features. However, the achieved MI discrimination turns out to be the lowest among the comparative approaches. Apparently, two main factors may affect the achieved accuracy: the extracted power-based short-time parameters that are not able to encode the stochastic EEG behavior; and the dimension reduction that uses the plain principal component analysis approach, but without taking into account the stochastic variability of the short-time features during multivariate projection. As a result, the introduction of the MIDFR algorithm performs better accuracy performance of MI discrimination patterns with high confidence, in comparison with benchmark methods for D1 dataset. Besides, it should be quoted that, unlike most of the-state-of-the-art approaches, MIDFR is tested over all the provided D1 subjects.

Regarding testing on the database D2 (see Table 3), the following remarks arise: the challenging conditions of the contest significantly increase because of the reduced number of measuring EEG electrodes (only three). In fact, all contrasted methods perform modest

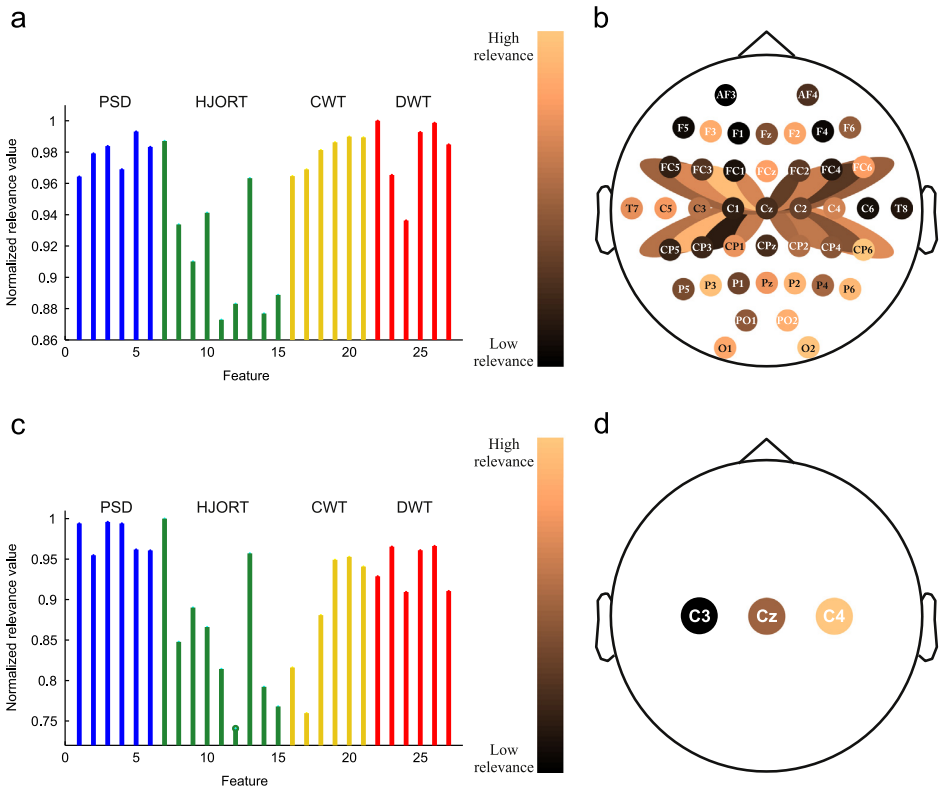

Fig. 2. Performed MIDFR relevance analysis: D1 – top row, D2 – bottom row. (a) and (c) Average feature relevance values. (b) and (d) Feature relevance channel distribution.

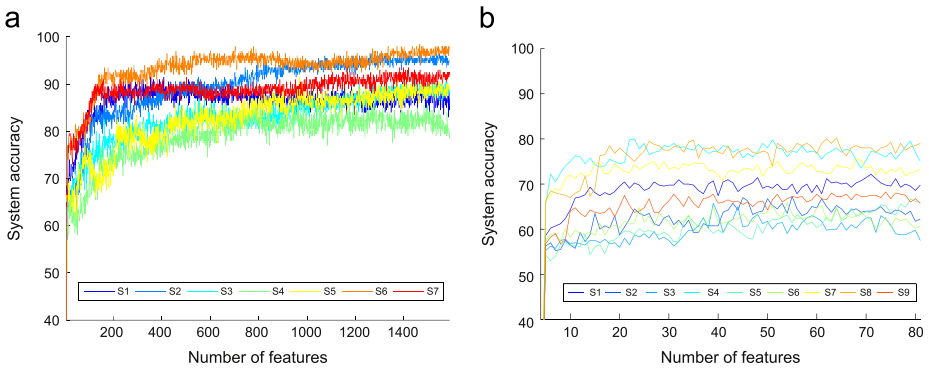

Fig. 3. MIDFR Performance curves. (a) D1 MIDFR results. (b) D2 MIDFR results.

Table 2
Accomplished classification results for D1 (average accuracy ± standard deviation [%]). (–) *Not provided.*

Subject	*He* [12]	*Zhang* [9]	*Higashi* [16]	*MIDFR*
S1	67.70 ± 02.20	77.20 ± 00.03	92.30 ± 02.50	91.50 ± 05.29
S2	70.70 ± 01.20	70.80 ± 0.02	90.60 ± 07.20	96.50 ± 03.37
S3	83.90 ± 01.30	–	–	91.50 ± 04.74
S4	93.00 ± 01.20	–	–	87.00 ± 06.32
S5	93.20 ± 1.20	–	–	91.50 ± 07.47
S6	–	76.80 ± 0.03	93.30 ± 03.60	98.50 ± 02.42
S7	–	80.00 ± 0.03	94.10 ± 04.10	93.50 ± 07.09
Mean	81.70 ± 12.06	76.20 ± 03.87	**92.58 ± 01.51**	**92.86 ± 03.77**

Table 3
Obtained classification results for D2 database.

Subject	*Delgado*[15]	*Darvishi*[18]	*Guo*[37]	*MIDFR*
S1	83	69	78.75	72.14 ± 6.83
S2	68	62	60.62	67.25 ± 6.71
S3	54	55	58.75	65.75 ± 6.67
S4	100	94	97.50	80.00 ± 4.86
S5	96	61	83.75	66.64 ± 5.82
S6	85	76	80.00	65.90 ± 8.37
S7	90	89	77.50	75.50 ± 10.33
S8	89	90	91.87	80.25 ± 4.48
S9	90	81	86.25	68.25 ± 4.26
Mean	**83.89** ± 14.38	75.22 ± 14.22	79.44 ± 12.90	72.30 ± **5.93**

values of accuracy on average. The best mean value of accuracy is reached by the approach given in [15] that involves spatial filtering of the EEG signals and estimation of power spectra based on autoregressive modeling of temporal segments of the EEG signals. Here, extracted short-time features feed a classifier based on hidden conditional random fields. Yet, inter-subject dispersion of accomplished accuracy values is so high that their confidence is not enough to be considered statistically significant. The same situation happens to other competitive approaches: the MI discrimination techniques presented in [18], where authors use a continuous wavelet transform for feature extraction together with an SVM classifier, and the method given in [37], where authors apply the independent component analysis based on the sliding window Infomax algorithm. Lastly, even though our proposed approach gets the lowest average accuracy, the inter-subject dispersion of the performed accuracy decreases to the half so that our results become more confident. The same figure remains for the computed intra-subject dispersion (6.48), meaning also that the MIDFR algorithm may imply a lower number of EEG measurements per subject.

5. Discussion and concluding remarks

An EEG data discrimination method is proposed as a tool for supporting MI classification tasks. The method includes the stochastic relevance analysis of the considered short-time features, which are extracted as to make prominent the nonstationary behavior of the EEG data. Furthermore, since it is widely accepted that the MI information is concentrated in the μ and β neural activity bands, we make use of the empirical mode decomposition together with the common-spatial-patterns mapping. We test two different BCI databases, using a soft-margin SVM based classifier that is validated by a 10-fold cross validation methodology. Since the proposed MIDFR algorithm better encodes neural activity dynamics, experimental results carried out on two different BCI databases show that the proposed approach allows improving detection of MI classification tasks.

However, during training and validation of the MIDFR algorithm, the following considerations are to be considered:

- The inclusion of the preprocessing stage is a cue procedure due to the high variability of the EEG measurements, but also to benefit from the fact that the MI information is concentrated within the μ and β neural activity rhythms. For this purpose, the pre-filtered broad-band procedure, termed EMD-based CSP, performs better accuracy than other preprocessing strategies using only CSP.
- Due to the imagination and execution of tracking movements are associated with neural rhythm power changes, we consider three representative feature extraction methods that have been widely applied in many motor imagery tasks (power spectral density, wavelet, and Hjorth methods). However, the ability of wavelet-based parameters to deal with non-stationary biosignals allows capturing more accurately spectral EEG dynamics with relevant information in terms of MI classification.
- From the obtained results for both databases, those testing methods carrying out the extraction of short-time feature sets get better accuracy than the ones using static parameters. Therefore, the stochastic feature extraction is a suitable strategy for tackling the EGG non-stationarity in MI tasks. However, the provided feature space often reaches enormous dimensions, making necessary the feature selection analysis. To this end, we measure the contribution of every considered short-time feature set, in terms of their stochastic variability, to make prominent the information time-varying behavior of the EEG data. Particularly, the proposed MIDFR algorithm outperforms those compared approaches that carry out multivariate dimension reduction without properly handling the stochastic variability of the considered short-time features.

As future work, the authors plan to improve the MIDFR algorithm by introducing more elaborate decomposition methodologies (other WT, improved versions of EMD, etc.) for a description of non-stationary signals. Moreover, it is worth noting that the used orthogonal eigen-based projection may have some limitations, being not enough to represent certain higher-order phenomena known to be present in EEG signals. Therefore, to encode complex data structures in BCI, big efforts should be done for extracting relevant EEG dynamics (kernel and/or information measures [38]). Additionally, hidden inter-channel relationships should be estimated to enhance extraction of main MI patterns. Finally, testing of the proposed methodology should be provided over other kind of EEG applications as Epilepsy detection and monitoring.

Acknowledgments

This work is carried out under grants provided by a M.Sc. scholarship funded by Universidad Nacional de Colombia, by the Ph.D. scholarship Programa Nacional de Formación de Investigadores ``Generación del Bicentenario'' 2011/2012 funded by Colciencias, and by the project 111045426008 funded by Colciencias.

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791.
[2] L. Yao, J. Meng, D. Zhang, X. Sheng, X. Zhu, Combining motor imagery with selective sensation towards a hybrid-modality BCI, IEEE Trans. Biomed. Eng. 99 (2013) 1–1.
[3] B.Z. Allison, E.W. Wolpaw, J.R. Wolpaw, Brain–computer interface systems: progress and prospects, Exp. Rev. Med. Dev. 4 (4) (2007) 463–474.
[4] A. Alvarez-Meza, L. Velasquez-Martinez, G. Castellanos-Dominguez, Feature relevance analysis supporting automatic motor imagery discrimination in EEG based BCI systems, in: EMBC, IEEE, 2013, pp. 7068–7071, http://dx.doi.org/10.1109/EMBC.2013.6611186.
[5] L.F. Velásquez-Martínez, A.M. Álvarez-Meza, C.G. Castellanos-Domínguez, Motor imagery classification for bci using common spatial patterns and feature relevance analysis, in: Natural and Artificial Computation in Engineering and Medical Applications, Springer, 2013, pp. 365–374, http://dx.doi.org/10.1007/978-3-642-38622-0_38.
[6] K. Blinowska, P. Durka, Electroencephalography (EEG), in: Wiley Encyclopedia of Biomedical Engineering, 2006.
[7] T. Tsoneva, D. Baldo, V. Lema, G. Garcia-Molina, EEG-rhythm dynamics during a 2-back working memory task and performance, in: EMBC, IEEE, 2011, pp. 3828–3831.
[8] G. Rodríguez, P.J. García, Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces, Med. Syst. 36 (1) (2012) 51–63.
[9] H. Zhang, C. Guan, K.K. Ang, C. Wang, BCI competition IV—data set I: learning discriminative patterns for self-paced eeg-based motor imagery detection, Front. Neurosci. 6 (2012).
[10] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, G. Pfurtscheller, Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment, IEEE Trans. Neural Syst. Rehab. Eng. 15 (4) (2007) 473–482.
[11] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (fbcsp) in brain–computer interface, in: IJCNN, IEEE, 2008, pp. 2390–2397, http://dx.doi.org/10.1109/IJCNN.2008.4634130.
[12] W. He, P. Wei, L. Wang, Y. Zou, A novel EMD-based common spatial pattern for motor imagery brain–computer interface, in: Biomedical and Health Informatics (BHI), IEEE, 2012, pp. 216–219, http://dx.doi.org/10.1109/BHI.2012.6211549.
[13] R. Corralejo, R. Hornero, D. Álvarez, Feature selection using a genetic algorithm in a motor imagery based brain computer interface, in: IEEE EMBC, 2011.
[14] O. Carrera-Leon, J.M. Ramirez, V. Alarcon-Aquino, M. Baker, D. D'Croz-Baron, P. Gomez-Gil, A motor imagery BCI experiment using wavelet analysis and spatial patterns feature extraction, in: 2012 Workshop on Engineering Applications (WEA), IEEE, 2012, pp. 1–6.
[15] J.D. Saa, M. Çetin, A latent discriminative model-based approach for classification of imaginary motor tasks from eeg data, J. Neural Eng. 9 (2) (2012) 026020.

[16] H. Higashi, T. Tanaka, Common spatio-time–frequency patterns for motor imagery-based brain machine interfaces, in: Computational Intelligence and Neuroscience, vol. 2013, 2013.
[17] M. Verleysen, D. François, The curse of dimensionality in data mining and time series prediction, in: Computational Intelligence and Bioinspired Systems, Springer, 2005, pp. 758–770, http://dx.doi.org/10.1007/11494669_93.
[18] S. Darvishi, M. C. Ridding, D. Abbott, M. Baumert, Investigation of the trade-off between time window length, classifier update rate and classification accuracy for restorative brain–computer interfaces, in: EMBC, IEEE, 2013, pp. 1567–1570, http://dx.doi.org/10.1109/EMBC.2013.6609813.
[19] J.D. Martínez-Vargas, J.I. Godino-Llorente, G. Castellanos-Dominguez, Time–frequency based feature selection for discrimination of non-stationary biosignals, EURASIP J. Adv. Signal Process. 2012 (1) (2012) 1–18.
[20] X. Wang, K.K. Paliwal, Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition, Pattern Recognit. 36 (10) (2003) 2429–2439.
[21] S. Bhattacharyya, A. Sengupta, T. Chakraborti, A. Konar, D. Tibarewala, Automatic feature selection of motor imagery eeg signals using differential evolution and learning automata, Med. Biol. Eng. Comput. 52 (2) (2014) 131–139.
[22] L. Duque-Muñoz, C. Aguirre-Echeverry, G. Castellanos-Domínguez, EEG rhythm analysis using stochastic relevance, in: XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, IFMBE Proceedings, vol. 41, 2014, pp. 658–661.
[23] B. Bankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Process. Mag. 08 (2008) 41–56.
[24] G. Pfurtscheller, C. Brunner, A. Schlögl, F.L. da Silvab, Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks, Neuroimage 31 (11) (2006) 153–159.
[25] G. Rodríguez-Bermúdez, P.J. García-Laencina, J. Roca-González, J. Roca-Dorda, Efficient feature selection and linear discrimination of EEG signals, Neurocomputing 115 (2013) 161–165.
[26] P. Herman, G. Prasad, T.M. McGinnity, D. Coyle, Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (4) (2008) 317–326.
[27] A. Subasi, Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction, Comput. Biol. Med. 37 (2) (2007) 227–244.
[28] G. Daza-Santacoloma, J. D. A.-L. no, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruíz, G. Castellanos-Domínguez, Dynamic feature extraction: an application to voice pathology detection, Intel. Aut. Soft Comput., http://dx.doi.org/10.1080/10798587.2009.10643056.
[29] J. Orozco, S. Murillo, A. Álvarez, J. Arias, E. Trejos, J. Vargas, G. Castellanos, Automatic selection of acoustic and non-linear dynamic features in voice, in: INTERSPEECH, 2011.
[30] A. Teixeira, A.M. Tomé, M. Boehm, C. Puntonet, E. Lang, How to apply nonlinear subspace techniques to univariate biomedical time series, IEEE Trans. Instr. Meas. 58 (8) (2009) 2433–2443.
[31] V. Bostanov, Bci competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the *t*-value scalogram, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1057–1061.
[32] O. Aydemir, T. Kayikcioglu, Wavelet transform based classification of invasive brain computer interface data, Radioengineering 20 (1) (2011) 31–38.
[33] S. Lemm, C. Schafer, G. Curio, Bci competition 2003-data set III: probabilistic modeling of sensorimotor μ rhythms for classification of imaginary hand movements, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1077–1080.
[34] M.H. Alomari, E.A. Awada, A. Samaha, K. Alkamha, Wavelet-based feature extraction for the analysis of EEG signals associated with imagined fists and feet movements, Comput. Inf. Sci. 7 (2) (2014) p17.
[35] A. Phinyomark, F. Quaine, Y. Laurillau, S. Thongpanja, C. Limsakul, P. Phukpattaranont. EMG amplitude estimators based on probability distribution for muscle–computer interface, Fluct. Noise Lett. 12 (03) (2013).
[36] T. Hanakama, I. Immisch, K. Toma, M.A. Dimyan, P.V. Gelderen, M. Hallett, Functional properties of brain areas associated with motor execution and imagery, J. Neurophysiol. 89 (2003) 989–1002.
[37] X. Guo, X. Wu, X. Gong, L. Zhang, Envelope detection based on online ica algorithm and its application to motor imagery classification, in: 2013 Sixth International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2013, pp. 1058–1061, http://dx.doi.org/10.1109/NER.2013.6696119.
[38] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, et al., Bci signal classification using a Riemannian-based kernel, in: Proceeding of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2012, pp. 97–102.

Andrés Marino Álvarez-Meza received his undergraduate degree in electronic engineering (2009) and his M.Sc. degree in engineering-industrial automation (2011) from the Universidad Nacional de Colombia. Actually, he is pursuing a Ph.D. in engineering-automatics at the same university. His research interests include machine learning and signal processing methods applied to image and video data analysis as well as bioengineering tasks.

Luisa F. Velasquez-Martinez received the B.S. degree in electronic engineering (2010) from the Universidad Nacional de Colombia. Currently, she is a Master student in engineering-industrial automation at Universidad Nacional de Colombia. Moreover, she is researching at Signal Processing and Recognition Group from the same university. Her research interests are feature extraction/selection for training pattern recognition systems, bio-engineering, and machine learning.

Germán Castellanos-Dominguez received his undergraduate degree in radiotechnical systems and his Ph.D. in processing devices and systems from the Moscow Technical University of communications and Informatics, in 1985 and 1990 respectively. Currently, he is a Professor in the Department of Electrical, Electronic and Computer Engineering at the Universidad Nacional de Colombia at Manizales. In addition, he is Chairman of the GCPDS at the same university. His teaching and research interests include information and signal theory, digital signal processing and bioengineering.