Expert Systems with Applications 41 (2014) 5285–5295

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Exploring dimensionality reduction of EEG features in motor imagery task classification

Pedro J. García-Laencina *, Germán Rodríguez-Bermudez, Joaquín Roca-Dorda

Centro Universitario de la Defensa de San Javier (University Centre of Defence at the Spanish Air Force Academy), MDE-UPCT, Calle Coronel Lopez Peña, s/n, 30720 Santiago de la Ribera, Murcia, Spain

ARTICLE INFO

Keywords:
Brain-computer interfaces
Electroencephalogram signals
Motor imagery
Dimensionality reduction
Feature transformation
Linear discriminants
Local Fisher Discriminant Analysis

ABSTRACT

A Brain-Computer Interface (BCI) system based on motor imagery (MI) identifies patterns of electrical brain activity to predict the user intention while certain movement imagination tasks are performed. Currently, one of the most important challenges is the adaptive design of a BCI system. For solving it, this work explores dimensionality reduction techniques: once features have been extracted from Electroencephalogram (EEG) signals, the high-dimensional EEG data has to be mapped onto a new reduced feature space to make easier the classification stage. Besides the standard sequential feature selection methods, this paper analyzes two unsupervised transformation-based approaches – Principal Component Analysis and Locality Preserving Projections – and the Local Fisher Discriminant Analysis (LFDA), which works in a supervised manner. The dimensionality in the projected space is chosen following a wrapper-based approach by an efficient leave-one-out estimation. Experiments have been conducted on five novice subjects during their first sessions with MI-based BCI systems in order to show that the appropriate use of dimensionality reduction methods allows increasing the performance. In particular, obtained results show that LFDA gives a significant enhancement in classification terms without increasing the computational complexity and, then, it is a promising technique for designing MI-based BCI system.

1. Introduction

Brain Computer Interface systems (BCIs) based on Electroencephalogram (EEG) signals allow to translate the user's brain activities into commands (Bashashati, Fatourechi, Ward, & Birch, 2007). The research of BCI technology is an interdisciplinary topic which has attracted an increasing interest during the past few years. There are several control signals types in BCI and, among them, motor imagery (MI) is one of the most studied (Nicolas-Alonso & Gomez-Gil, 2012). MI is a mental process by which an individual rehearses or simulates a given action. MI has been successfully used in several BCI applications, such as robotic devices (arms, legs, wheelchairs, vehicles, etc.), communication systems and environmental control. The neuro-feedback provided by a BCI system may improve cognitive performance, speech skills, affection, and pain management; and it has been also used in the treatment of mental disorders (Nicolas-Alonso & Gomez-Gil, 2012).

In general, a BCI can be defined as a pattern recognition system (Duda, Hart, & Stork, 2000; Lotte, Congedo, Lécuyer, Lamarche, & Arnaldi, 2007) composed of several consecutive modules aimed at identifying 'patterns' of electrical brain activity (see Fig. 1). The first module, Acquisition, is an interface to peripheral equipment that scans and transmits EEG signals. However, the raw EEG data transmitted from acquisition module present a signal–noise-rate (SNR) very low and, thus, it is needed the Signal Preprocessing module whose essential task is to denoise and to remove artifacts. Additionally, this module usually performs signal amplification. The third one is the Feature Extraction (FE) module that collects all preprocessed signals from former module and extracts features to get discriminative information and, then, it transmits these features to the Pattern Classification module. Finally, the Pattern Classification block determines the classes to which the particular brain signals belong. Depending on the BCI application, these predicted classes are translated to the corresponding feedback signals or control commands for external devices (Huster, Mokom, Enriquez-Geppert, & Herrmann, 2014).

One of the most popular tasks for MI-based BCI system is the paradigm of classifying motor imagination of left and right hand movements (Ahn, Cho, Ahn, & Jun, 2013; Bhattacharyya, Sengupta, Chakraborti, Konar, & Tibarewala, 2014; Blankertz et al., 2010; Fruitet, McFarland, & Wolpaw, 2010; Guger et al., 2001; Hsu, 2012; Pfurtscheller, Brunner, Schlgl, & da Silva, 2006; Sanei &

* Corresponding author.
E-mail addresses: pedroj.garcia@cud.upct.es (P.J. García-Laencina), german.rodriguez@cud.upct.es (G. Rodríguez-Bermudez).

http://dx.doi.org/10.1016/j.eswa.2014.02.043
0957-4174/

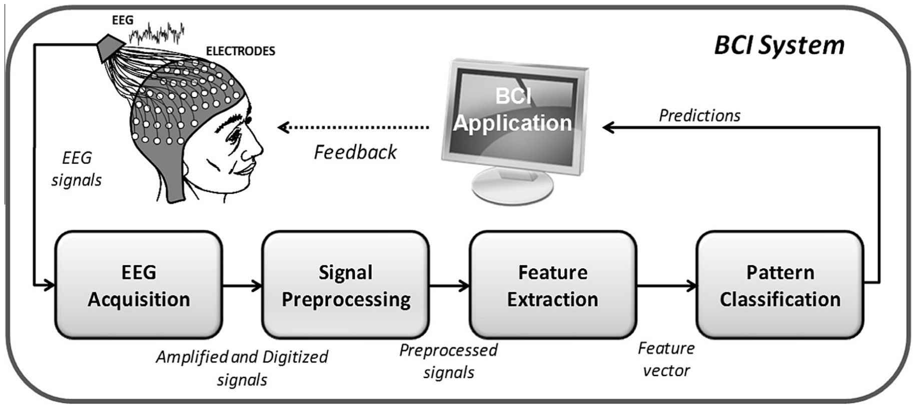

Fig. 1. General block diagram of a EEG-based BCI system.

Chambers, 2008; Siuly, Li, & Wen, 2014; Vidaurre, Sannelli, Müller, & Blankertz, 2011; Zhang et al., 2013). Nevertheless, and according to the state-of-the-art (Ahn et al., 2013; Blankertz, Tomioka, Lemm, Kawanabe, & Muller, 2008; Krusienski et al., 2011; Vidaurre et al., 2011), certain subjects may face difficulties to use MI-based BCI systems and, in these cases, the classification performances are quite poor even using multiple training sessions. On the other hand, it is also known that there are subjects that can properly use the BCI system from their first sessions. In addition to this, the design of BCI system has to take into account that every user is unique with respect to his/her cognitive abilities and, also, the high variability of the EEG signals from session to session or even within a session. Therefore, it is important to implement adaptive BCI systems for each subject aimed at providing a good enough classification performance from the beginning and, thus, the feelings of frustration/rejection can be avoided in those subjects with initial difficulties for handling BCI systems. Besides, since EEG signals vary over time, the BCI system should be able to adapt its internal structure/operation for each new trial as far as possible (McFarland, Sarnacki, & Wolpaw, 2011). As we will see throughout this paper, the proposed methodology and the conducted off-line experiments seek to achieve these objectives.

Many research efforts have done in the last two stages of feature extraction and classification for achieving adaptive BCI systems. This paper is focused on an intermediate stage between both of them: once input features have been extracted from EEG signals, the aim is to map the high dimensional input data onto a new reduced feature space which makes easier the classification stage. It entails to perform Dimensionality Reduction (DR) (van der Maaten, Postma, & van den Herik, 2009), which can be done by means of Feature Selection (FS) or Feature Transformation (FT) methods. FS approaches choose a subset of variables from the original extracted features. Note that FS does not change the original representation of the input variables; whereas FT entails a modification of the extracted features to a new low-dimensional feature space.

In general, there are two main kinds of FS techniques (Guyon & Elisseeff, 2003): filter and wrapper. In filter procedures, variables are chosen without running the classifier and the selection is performed by considering the statistical properties of the input data. On the other hand, wrapper methods select the most relevant features by evaluating different feature subsets on the classification algorithm. Note that wrapper procedures can provide better performance, but they generally entail a higher computational cost than filter methods due to the multiple trainings of the classification algorithm. Within the BCI literature (Bashashati et al., 2007; Nicolas-Alonso & Gomez-Gil, 2012), FS has been previously analyzed in several works. Koprinska (2009) evaluated five well-known filter FS methods in a EEG dataset from the BCI competition III and, according to its experimental results, classification performance was improved in comparison to the case without FS. Another alternative is to perform FS by means of exhaustive searches (forward and backward procedures) in a wrapper manner: it has been also done in BCI (Cabrera, Farina, & Dremstrup, 2010; Guerrero-Mosquera, Verleysen, & Navia Vazquez, 2010; McCann, Thompson, Syed, & Huggins, 2014; Sabeti, Boostani, Katebi, & Price, 2007Şiuly et al., 2014), despite it may be computationally expensive. In order to solve this inconvenient, some recent BCI studies (Fruitet et al., 2010; Rodríguez-Bermúdez, García-Laencina, & Roca-Dorda, 2013a, Rodríguez-Bermúdez, García-Laencina, Roca-González, & Roca-Dorda, 2013b) are based on a two-stage wrapper procedure, where EEG features are firstly ordered in fast way by means of regression-based procedures and, then, those relevant features that provide better classification results are chosen. Another FS wrapper procedure in BCI systems is the application of an Evolutionary Algorithm (EA) (Bhattacharyya et al., 2014), where Genetic Algorithms (GA) are the most popular type of EA (Atyabi, Luerssen, & Powers, 2013; Corralejo, Hornero, & Alvarez, 2011; Garrett, Peterson, Anderson, & Thaut, 2003; He, Hu, Li, & Li, 2013; Hsu, 2013; Rejer & Lorenz, 2013; Sabeti et al., 2007; Sardouie & Shamsollahi, xxxx). Although GA-based methods may provide accurate FS, its main disadvantages are its excessive running time for designing BCI systems and the fact that each run of the GA creates a different subset of features: GA requires several runs to determine which features are chosen most often.

With respect to transformation-based methods, they are not very common during the practical design of BCI systems, although these approaches have recently aroused great interest in other research fields (Chen, Liu, Yang, Liu, & Wang, 2011; Oh, Yoo, & Pedrycz, 2013; van der Maaten et al., 2009; Yu, 2011). From the different FT methods, Principal Component Analysis (PCA) is one of the most traditionally used in BCI systems (Bashashati et al., 2007; Lee et al., 2004; Nicolas-Alonso & Gomez-Gil, 2012; Yu, Chum, & Sim, 2014), although it has been also used as an EEG preprocessing method for noise reduction and channel selection (Naeem, Brunner, & Pfurtscheller, 2009). According to the BCI literature (Bashashati et al., 2007; Nicolas-Alonso & Gomez-Gil, 2012), there other signal processing techniques -such as Independent Component Analysis (ICA) and Common Spatial Patterns (CSP)- which are related with transformation-based methods. However, these procedures directly deal with the EEG signals and not with the obtained variables from the FE stage. In particular, the CSP method projects the multichannel EEG signals into a subspace (the obtained CSP features) with an optimal variance for the subsequent classification (Blankertz et al., 2008). With respect to ICA, this method has mostly used as a preprocessing tool before the FE step in order to remove artifacts, but it has also been used for obtaining a reduced set of linear transformed features from the EEG signals (Erfanian & Erfani, 2004).

It is important to remark that, in this work, the EEG features have been previously extracted before applying DR (FS or FT) methods. Given the original extracted EEG features, this paper explores, analyzes and experimentally compares different representative DR methods in BCI datasets. In particular, two widely-used FS procedures: Sequential Backward Selection (SBS) and Sequential Forward Selection (SFS); and three representative FT methods: PCA, Locality Preserving Projections (LPP) and Local Fisher Discriminant Analysis (LFDA). According to the current literature, the LFDA method has not been previously applied for BCI data and, in the case of LPP, there is a previous reference work (Gan, 2006). Another important contribution of this paper is the choice of the dimensionality in the projected space by an efficient Leave-One-Out Cross-Validation (LOO-CV) criterion, which makes suitable the adaptation of the BCI system for each trial. Finally, it should be remarked that the experiments have been conducted on five novice subjects during their first sessions with MI-based BCI systems and, moreover, the acquired EEG signals have not been previously analyzed. By exploring DR methods in these datasets, we pretend to make easier the design of BCI systems from first sessions with novice users.

The rest of the paper is structured as follows. In Section 2, the EEG data under study and the different modules of the implemented BCI system are described. Section 3 introduces the DR methods. Section 4 shows and discusses the experimental results, and finally, conclusions and future works are given in Section 5.

2. Material and methods

2.1. EEG acquisition and preprocessing

The EEG signals used in the current study were taken from University Centre of Defence at the Spanish Air Force Academy, Spain. The complete dataset consisted of five sets of EEG data obtained from five healthy male volunteers, denoted as Subjects A–E, during motor imagery for left and right hand movements. It is important to remark that all the subjects have never worked with BCI systems before and, thus, it were their first experiments and there was not a previous selection of the users.

For the BCI experiments, electrical brain signals were captured with an EEG amplifier system g.USBamp (g.tec Medical Engineering GmbH, Austria) using two bipolar channels, which have been mounted on the subject's head using Ag/AgCl passive electrodes. As it is usual for MI, these electrodes were placed over FC3, CP3, FC4 and CP4 positions. The EEG signal was amplified using g.USBamp that samples the data with 8 bits and 256 Hz. Additionally, the signals were preprocessed to avoid artifacts and noise using a digital band pass filter between 0.5 and 30 Hz and a notch filter to suppress the 50 Hz power line interference. To develop the experiments, all the subjects were seated and relaxed about 1 meter in front of the computer monitor and they were instructed not to move and to keep both arms and feet relaxed and to maintain throughout the experiment the fixation at the center of the monitor. The experiment started with the display of a fixation cross in the center of a screen. After 2 s, a warning stimulus was given in the form of a 'beep'. After 3 s, an arrow (cue stimulus) pointing to the left or right was shown for 1.25 s. The subject was instructed to imagine a right-hand movement or left-hand movement until the end of the trial, depending on the direction of the arrow. This MI period was 4 s. Then, one complete trial lasted 8 s and the time between two consecutive trials was randomized between 0.5 and 2.5 s to avoid adaptation. The subjects performed two complete sessions consisting each of 40 rounds (20 'left' and 20 'right' trials) without feedback and, thus, there are 80 trials per subject.

2.2. Feature extraction (FE)

Using different FE methods, it is expected that the BCI system combines and exploits the corresponding information from each procedure and this could improve the classifications results. In this work, the software of the FE module, which has been specifically implemented for these experiments, is able to compute in parallel and in real time three well-known FE methods for MI tasks: the band power in two different frequency bands of the EEG data, the Hjorth parameters and the adaptive autoregressive coefficients. This section introduces the basic notions of these representative FE methods that, according to Bashashati et al. (2007), have been widely applied in many BCI systems for MI tasks and their software implementations are available in the BioSig toolbox,[1] which is an open-source reference software for biomedical signal processing. Note that detailed descriptions of these FE methods are available in Tong and Thankor (2009).

- Band Power (BP) features (Pfurtscheller et al., 2006). There are four frequency bands that are identified for interpretation of the EEG signals (Sabeti et al., 2007; Sanei & Chambers, 2008) but, as it is usual in BCI systems, only the most reactive frequency bands for MI were computed (Pfurtscheller et al., 2006): the alpha (8–12 Hz) and beta range (16–24 Hz). Then, for each EEG signal, two BP features are computed as the energies of the alpha and beta bands.
- Hjorth (HJ) parameters (Hjorth, 1970). In 1970, B. Horjth introduced a set of three parameters to describe characteristics of the EEG signal in the time domain: *Activity*, the signal power (which is wide band filtered); *Mobility*, the mean frequency; and *Complexity*, the change in frequency. According to Hjorth (1970), the computation of the HJ parameters is only based on the EEG signal variance.
- Adaptive Auto-Regressive (AAR) coefficients (Schlögl, 2000). The AAR modeling is an evolution of the AR modeling. In AR modeling (Anderson, Stolz, & Shamsunder, 1998), each input sample is predicted by a weighted linear combination of the previous p samples, where p denotes the model order. These coefficients are computed using all the samples and get the coefficient vector to predict the next sample. However, in AAR modeling, the coefficient values are continuously updated to get a more accurate estimation. In this work, the AAR coefficients have been estimated using the Kalman filter algorithm. After consulting the literature and based on our previous experimental works (Billinger, Brunner, & Neuper, 2010; Rodríguez-Bermúdez et al., 2013a), p have been set to six.

2.3. Time course of EEG data

As we have mentioned above, the MI period is four seconds long in each trial. In this period, the BCI system has to extract informative features about the classification task but the EEG signal is nonstationary and, due to this, the features vary over time. Therefore, the most usual approach is to divide the MI period into different temporal windows and to consider statistically stationary each one (Lotte et al., 2007). After that, the FE methods are applied to each individual window and, finally, the obtained features are combined in a unique vector per trial by averaging, i.e., the feature vector for each trial is defined by the average of the feature values computed in each temporal window. In particular, we have used a

[1] http://biosig.sourceforge.net/.

1-s window of EEG data, as in other previous research works (Cabrera et al., 2010; Delgado-Saa & Cetin, 2011; Rodríguez-Bermúdez et al., 2013a), and we have applied the three FE methods described in the previous section. In the dataset under study, each single trial is composed by two EEG signals (from bipolar channels C3 and C4) and, thus, the FE stage computes eleven input features for each signal: the two features from BP method, the three HJ parameters and the six coefficients from AAR modeling. Then, a total of twenty-two input features are extracted for each single trial of EEG data. In this article,

$$\mathbf{x}_n = [x_{0n}, x_{1n}, \ldots, x_{Dn}]^\top \tag{1}$$

denotes the nth feature vector or pattern associated to the nth trial: a D + 1 column vector with components $x_{0n} = 1$ (which plays the role of the bias parameter) and x_{dn} a certain feature value (with $d = 1, 2, \ldots, D$ and $n = 1, 2, \ldots, N$; being $D = 22$ and $N = 80$ in each dataset). t_n denotes the label of $\mathbf{x}_n$, which may belong to two categories: $\mathcal{C}_1$ (left hand) or $\mathcal{C}_2$ (right hand).

2.4. Classification

From the great variety of classification methods applied for implementing BCI systems (Fruitet et al., 2010; Garrett et al., 2003; Liang, Saratchandran, Huang, & Sundararajan, 2006; Lotte et al., 2007; Muller, Anderson, & Birch, 2003; Müller, Krauledat, Dornhege, Curio, & Blankertz, 2004), *linear discriminants* are the most popular because they provide accurate, stable and robust performance with very low computational complexity (Lotte et al., 2007). Given a D-dimensional input vector $\mathbf{x}$, a linear discriminant is defined by a linear combination of the input features of $\mathbf{x}$ by a weight vector $\mathbf{w} = [w_0, w_1, \ldots, w_D] \in \mathbb{R}^{D+1}$, i.e., $y = \mathbf{w}^\top \mathbf{x}$. Then, for a binary classification problem, the sign of the output y can be used as decision rule: if $y \geqslant 0$, $\mathbf{x}$ belongs to class $\mathcal{C}_1$; and to class $\mathcal{C}_2$ otherwise.

There are many well-known procedures to compute the weight vector of the linear discriminant (Alpaydin, 2010; Bishop, 2006; Duda et al., 2000), such as the perceptron learning, Support Vector Machines (SVM), the Least Squares (LS) approach and the Fisher Discriminant Analysis (FDA). In this work, we consider the LS approach to realize linear discriminants, which solution coincides with that found from the FDA method if the targets are coded to N/N_1 and $-N/N_2$ for categories $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively (Duda et al., 2000). Following the LS approach, $\mathbf{w}$ is obtained by minimizing the Mean Square Error (MSE):

$$\min_{\mathbf{w}} \sum_{n=1}^{N} (y_n - t_n)^2 = \min_{\mathbf{w}} \sum_{n=1}^{N} \left(\mathbf{w}^\top \mathbf{x}_n - t_n\right)^2, \tag{2}$$

which is solved using the pseudoinverse of the input data matrix, $\mathbf{X}^\dagger$, and, then, the solution for the weight vector is

$$\hat{\mathbf{w}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{t} = \mathbf{X}^\dagger \mathbf{t}. \tag{3}$$

When $\mathbf{X}$ is not full rank (Serre, 2002), $\mathbf{X}^\dagger$ can be efficiently computed using Singular Value Decomposition (SVD).

3. Dimensionality reduction of EEG features by transformation-based methods

As it has been mentioned in Section 1, and once features have been extracted from the EEG signals, Dimensionality Reduction (DR) can be done either by Feature Selection (FS) or Feature Transformation (FT) to a low dimensional data space (Alpaydin, 2010; Fukunaga, 1990; Guyon & Elisseeff, 2003; van der Maaten et al., 2009). Up to now, FS approaches are typically preferred over transformation-based methods because the original representation/meaning of the extracted features is not modified. Besides, once the most representative features are chosen by the FS method, only them have to be computed; whereas, in FT approaches, all input variables are required to obtain the reduced feature space. However, these two aspects do not justify the advantages of FS over transformation-based methods for designing BCI systems. Although FS offers good interpretation, mapping high dimensional EEG data onto a new low-dimensional feature space can be interpreted as high level of feature extraction or feature fusion and, then, it can increase the classification accuracy if the feature mapping is properly performed. Additionally, due to the high variability of the EEG features and the continuous adaptation of the BCI system (Krusienski et al., 2011; Lotte et al., 2007; Müller et al., 2004; Shenoy, Krauledat, Blankertz, Rao, & Müller, 2006), discarding a subset of features may not be beneficial in classification terms for later stages of the design. Then, during the adaptive design of a BCI system (McFarland et al., 2011; Shenoy et al., 2006), all features have to be continuously computed and selected/transformed to increase its prediction capabilities for all time. Finally, BCI researchers are also interested on the 2D or 3D representation of EEG features without loss of discriminative information, because it allows to visually analyze the structure of the EEG data and detect possible outliers.

We now introduce the notions of DR by FT methods. Let the original data be $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^\top$ with N input vectors and each vector $\mathbf{x}_n \in \mathbb{R}^D$ ($n = 1, 2, \ldots, N$), where $\mathbf{X}$ is a $N \times D$ matrix. Assume that this dataset has intrinsic dimensionality r (where $1 \leqslant r \leqslant D$), which means that the samples in $\mathbf{X}$ are lying on or near a projected space with dimensionality r that is embedded in the original D-dimensional space. Transformation-based techniques map $\mathbf{X}$ with dimensionality D into a new dataset $\mathbf{Z}$ with dimensionality r, being $\mathbf{z}_n \in \mathbb{R}^r$ the low-dimensional representation of $\mathbf{x}_n$. This paper is focused on linear mapping techniques (Alpaydin, 2010; van der Maaten et al., 2009), where the projected samples are obtained by

$$\mathbf{z}_n = \mathbf{F}^\top \mathbf{x}_n, \tag{4}$$

being $\mathbf{F}$ a $D \times r$ transformation matrix. Linear techniques are simpler/easier to be implemented than non-linear transformations (Alpaydin, 2010; Lee & Verleysen, 2007; van der Maaten et al., 2009): they are fast and suitable for practical applications.

The different FT techniques can be categorized as either unsupervised or supervised depending on whether or not they use class-membership information, while computing the r-dimensional space (van der Maaten et al., 2009). Unsupervised approaches may be not necessarily useful in classification problems since they do not take the target information into account. The best known and most widely used feature mapping methods are Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA)[2], which are both linear projection approaches, unsupervised and supervised respectively. As an alternative to PCA (Duda et al., 2000), the Locality Preserving Projections (LPP) method has been proposed to achieve a linear transformation that optimally preserves local neighborhood information of the data (He & Niyogi, 2004). Recently, Sugiyama has developed the Local Fisher Discriminant Analysis (LFDA), which effectively combines the ideas of FDA and LPP (Sugiyama, 2007). Now, we discuss these four approaches for feature mapping.

3.1. Principal Component Analysis (PCA)

PCA is an unsupervised method that computes a linear mapping $\mathbf{F}$ in order to achieve a low-dimensional representation of the

[2] Note that FDA is also known as Linear Discriminant Analysis (LDA) and it may refer to the linear classification method which first projects samples onto a one-dimensional subspace and, then, classifies the projected samples using thresholding.

original data in which the amount of variance is maximal (Alpaydin, 2010; Duda et al., 2000). In general, PCA has been succesfully applied in many scientific fields and it also receives much attention in BCI experiments (Bashashati et al., 2007; Gan, 2006; Lee et al., 2004; Naeem et al., 2009; Rodríguez-Bermúdez & García-Laencina et al., 2013b; Yu et al., 2014).

In mathematical terms, PCA finds $\mathbf{F}$ that maximizes the cost function

$$\mathbf{F}_{PCA} = \underset{\mathbf{F}\in\mathbb{R}^{D\times r}}{\arg\max}\left[\mathrm{trace}\left(\mathbf{F}^{\top}\mathbf{S}\mathbf{F}\right)\right], \tag{5}$$

where $\mathbf{S}$ is the sample covariance matrix of $\mathbf{X}$. It is widely known that this linear transformation is formed by an orthogonal basis from the top eigenvectors (i.e. principal components) of the data covariance matrix, i.e., the eigenvectors corresponding to the r largest eigenvalues are retained. Note that the input data has to be normalized before applying PCA to mitigate the effects of scale and, also, it is recommended to avoid computing $\mathbf{S}$ explicitly because it may be hard when D is very large.

3.2. Locality linear projection (LPP)

Different from PCA which seeks a mapping in the directions of maximal variance, the Locality linear projection (LPP) algorithm (He & Niyogi, 2004) projects to preserve the neighborhood structure of the data and, then, it makes LPP relatively insensitive to noise and outliers, which PCA is sensitive to. In contrast to PCA, the LPP algorithm has hardly been explored to reduce the dimensionality of EEG data during the design of BCI systems (Gan, 2006).

Before describing the LPP method (He & Niyogi, 2004), we need to define that $A_{n,m}$ is the affinity between $\mathbf{x}_n$ and $\mathbf{x}_n$. Assuming that $A_{n,m}\in[0,1]$; $A_{n,m}$ tends to 1 if $\mathbf{x}_i$ and $\mathbf{x}_j$ are 'close' and it tends to 0 if $\mathbf{x}_n$ and $\mathbf{x}_m$ are 'far apart'. From He and Niyogi (2004) and Sugiyama (2007), there are several ways of defining $A_{n,m}$, being a practical and extended choice:

$$A_{n,m} = \exp\left(-\frac{\|\mathbf{x}_n-\mathbf{x}_m\|^2}{\sigma_n\sigma_m}\right). \tag{6}$$

σ_i is the local scaling around $\mathbf{x}_n$ defined by $\sigma_n = \|\mathbf{x}_n - \mathbf{x}_n^{(k)}\|$ where $\mathbf{x}_n^{(k)}$ is the kth nearest neighbor of $\mathbf{x}_n$. In Zelnik-Manor and Perona (2005), it is demonstrated that $k = 7$ works well.

Once the affinity matrix has been defined, and according to He and Niyogi (2004), the LPP transformation matrix is given by:

$$\mathbf{F}_{LPP} = \underset{\mathbf{F}\in\mathbb{R}^{D\times r}}{\arg\min}\left[\frac{1}{2}\sum_{n,m=1}^{N}A_{n,m}\|\mathbf{F}^{\top}\mathbf{x}_n-\mathbf{F}^{\top}\mathbf{x}_m\|^2\right]$$
$$\text{subject to } \mathbf{F}^{\top}\mathbf{X}\mathbf{D}\mathbf{X}^{\top}\mathbf{T} = \mathbf{I}_r, \tag{7}$$

where $\mathbf{I}_r$ is the identity matrix in $\mathbb{R}^r$ and $\mathbf{D}$ is the N-dimensional diagonal matrix with $D_{n,n} = \sum_{m=1}^{N}A_{n,m}$. The constraint of (7) is included to remove an arbitrary scaling factor in the transformation. The minimization of (7) tries to ensure that $\mathbf{x}_n$ and $\mathbf{x}_m$ are 'close' then its corresponding projected samples are kept closed.

It has been proved that the solution of (7) can be obtained by solving the generalized eigenvalues problem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^{\top}\varphi = \gamma\mathbf{X}\mathbf{D}\mathbf{X}^{\top}\varphi \tag{8}$$

where $\mathbf{L}$ is the graph-Laplacian matrix (Chung, 1997). The vectors that minimize the objective function are given by minimum eigenvalues solutions to (8). Let the column vectors $\{\varphi_d\}_{d=1}^{D}$ be the generalized eigenvectors ordered according to their eigenvalues $\gamma_1 \geqslant \gamma_2 \geqslant \cdots \geqslant \gamma_D$ of the problem defined in (8). According to He and Niyogi (2004), the transformation in LPP is written as follows:

$$\mathbf{F}_{LPP} = (\varphi_D|\varphi_{D-1}|\ldots|\varphi_{D-r+1}). \tag{9}$$

3.3. Fisher Discriminant Analysis (FDA)

Before explaining the notions of LFDA, we introduce Fisher Discriminant Analysis (FDA), which is the most popular supervised method for feature transformation (Alpaydin, 2010; Bishop, 2006; Duda et al., 2000; Fukunaga, 1990). It seeks the linear transformation, $\mathbf{F}_{FDA}$, which maximizes the between-class distance and minimizes the within-class distance simultaneously. Although FDA is not strictly a discriminant, and as it has been already mentioned in Section 2.4, FDA is often used to design linear classifiers in many scientific fields, such as BCI (Bashashati et al., 2007; Lotte et al., 2007; Muller et al., 2003).

For obtaining $\mathbf{F}_{FDA}$, we need to consider that $\mathbf{S}_B$ and $\mathbf{S}_W$ respectively denote the between-class covariance matrix and the within-class covariance matrix. According to Sugiyama (2007) and Gao, Kwan, and Huang (2009), $\mathbf{S}_B$ and $\mathbf{S}_W$ are given in the graph embedding framework by:

$$\mathbf{S}_B = \frac{1}{2}\sum_{n,m=1}^{N}V_{n,m}^{B}(\mathbf{x}_n-\mathbf{x}_m)(\mathbf{x}_n-\mathbf{x}_m)^{\top}, \tag{10}$$

$$\mathbf{S}_W = \frac{1}{2}\sum_{n,m=1}^{N}V_{n,m}^{W}(\mathbf{x}_n-\mathbf{x}_m)(\mathbf{x}_n-\mathbf{x}_m)^{\top}, \tag{11}$$

where

$$V_{n,m}^{B} = \begin{cases}\frac{1}{N}-\frac{1}{N_c}, & \text{if } (\mathbf{x}_n,\mathbf{x}_m) \text{ belong to } \mathcal{C}_c,\\ \frac{1}{N}, & \text{if } (\mathbf{x}_n,\mathbf{x}_m) \text{ belong to different classes,}\end{cases} \tag{12}$$

$$V_{n,m}^{W} = \begin{cases}\frac{1}{N_c}, & \text{if } (\mathbf{x}_n,\mathbf{x}_m) \text{ belong to } \mathcal{C}_c,\\ 0, & \text{if } (\mathbf{x}_n,\mathbf{x}_m) \text{ belong to different classes.}\end{cases} \tag{13}$$

Then, the transformation matrix in FDA is obtained by:

$$\mathbf{F}_{FDA} = \underset{\mathbf{F}\in\mathbb{R}^{D\times r}}{\arg\max}\left[\mathrm{trace}\left(\left(\mathbf{F}^{\top}\mathbf{S}_W\mathbf{F}\right)^{-1}\left(\mathbf{F}^{\top}\mathbf{S}_B\mathbf{F}\right)\right)\right]. \tag{14}$$

The solution of (14) can be obtained by solving a generalized eigenvalue problem. Let the column vectors $\{\upsilon_d\}_{d=1}^{D}$ be the generalized eigenvectors associated with their generalized eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_D$ of the following problem:

$$\mathbf{S}_B\upsilon = \lambda\mathbf{S}_W\upsilon. \tag{15}$$

Thus, the solution of the above maximization problem is given by:

$$\mathbf{F}_{FDA} = (\upsilon_1|\upsilon_2|\ldots|\upsilon_r). \tag{16}$$

Finally, it should be noted that, for a binary classification problem, $\mathbf{F}_{FDA}$ is defined by a D-dimensional vector instead of a $D\times r$ matrix. In this case, FDA projects the original data down to one dimension that gives a large separation between the projected class means while also giving a small variance within each class.

3.4. Local Fisher Discriminant Analysis (LFDA)

With the aim to extend the advantages of LPP into a supervised DR (such as FDA), the Local Fisher Discriminant Analysis (LFDA) method has been proposed (Sugiyama, 2007). In particular, LFDA computes the between-class covariance matrix and the within-class covariance matrix in a local manner. Based on (10) and (11), LFDA defines both the local between-class covariance matrix and the local within-class covariance matrix:

$$\widetilde{\mathbf{S}}_B = \frac{1}{2}\sum_{n,m=1}^{N}\widetilde{V}_{n,m}^{B}(\mathbf{x}_n-\mathbf{x}_m)(\mathbf{x}_n-\mathbf{x}_m)^{\top}, \tag{17}$$

$$\widetilde{\mathbf{S}}_W = \frac{1}{2}\sum_{n,m=1}^{N}\widetilde{V}_{n,m}^{W}(\mathbf{x}_n-\mathbf{x}_m)(\mathbf{x}_n-\mathbf{x}_m)^{\top}, \tag{18}$$

where

$$\widetilde{V}^{B}_{n,m} = \begin{cases} A_{n,m}\left(\frac{1}{N} - \frac{1}{N_c}\right), & \text{if } (\mathbf{x}_n, \mathbf{x}_m) \text{ belong to } \mathcal{C}_c, \\ \frac{A_{n,m}}{N}, & \text{if } (\mathbf{x}_n, \mathbf{x}_m) \text{ belong to different classes}, \end{cases} \quad (19)$$

$$\widetilde{V}^{W}_{n,m} = \begin{cases} \frac{A_{n,m}}{N_c}, & \text{if } (\mathbf{x}_n, \mathbf{x}_m) \text{ belong to } \mathcal{C}_c, \\ 0, & \text{if } (\mathbf{x}_n, \mathbf{x}_m) \text{ belong to different classes}. \end{cases} \quad (20)$$

In the above equations, symbols with tilde are used to denote the local versions of covariance matrices and weights. In LFDA, the affinity $A_{n,m}$ weights the values for sample pair $(\mathbf{x}_n, \mathbf{x}_m)$ in the same class. In particular, $A_{n,m}$ is computed by (6) as in LPP but in class-wise manner in order to preserve the within-class local structure of the data. The smaller the affinity value $A_{n,m}$, the farther away the pair $(\mathbf{x}_n, \mathbf{x}_m)$ and, then, it has less influence on $\widetilde{\mathbf{S}}_B$ and $\widetilde{\mathbf{S}}_W$. Note that if $A_{n,m}$ is set to 1 for all sample pairs, $\widetilde{\mathbf{S}}_B$ and $\widetilde{\mathbf{S}}_W$ correspond to $\mathbf{S}_B$ and $\mathbf{S}_W$, respectively.

Given that LFDA can be seen as a localized variant of FDA (Sugiyama, 2007), the LFDA transformation matrix ($\mathbf{F}_{LFDA}$) can be obtained in a similar way than the FDA transformation matrix ($\mathbf{F}_{FDA}$):

$$\mathbf{F}_{LFDA} = \underset{\mathbf{F} \in \mathbb{R}^{D \times r}}{\arg\max} \left[\operatorname{trace}\left((\mathbf{F}^\top \widetilde{\mathbf{S}}_W \mathbf{F})^{-1} (\mathbf{F}^\top \widetilde{\mathbf{S}}_B \mathbf{F})\right)\right]. \quad (21)$$

In other words, LFDA seeks for a mapping that nearby data pairs in the same class are mapped as close together as possible in the reduced space; meanwhile, the data pairs in different classes are separated from each other. Similarly to the analytic solution of FDA defined by (15) and (16), the LFDA criterion of (21) can be maximized directly by computing the eigenvectors of a generalized eigenvalue problem of $\widetilde{\mathbf{S}}_B$ and $\widetilde{\mathbf{S}}_W$.

Finally, it should be emphasized that LFDA has been implemented in order to reduce the computational cost (Sugiyama, 2007). In particular, it has been done by defining $\widetilde{\mathbf{S}}_B$ and $\widetilde{\mathbf{S}}_W$ in alternative matrices expressions which can be efficiently computed. For specific details of the efficient computation of $\mathbf{F}_{LFDA}$, see (Sugiyama, 2007). As we will show in the experimental results, the efficient matrix-based implementation of LFDA makes feasible its application to design BCI systems.

3.5. Choosing the number of dimensions in the transformed data using LOO-CV error

Unfortunately, there is no general rule for deciding how to choose the lower dimensionality of the projected data (r). As a first alternative, the number of dimensions is often chosen according to ad hoc filter rules which may be intuitively plausible and may work well in practice (Alpaydin, 2010; Bishop, 2006). For example, in PCA, it is usually done by selecting those components that contain almost 90% of the total variance (Yu et al., 2014). Its main drawback is that filter-based rules are defined without taking into account the performance obtained by the trained classifier.

Another kind of approaches makes use of Cross-Validation (CV) procedures for choosing the r dimensions to optimize a predefined criterion (Alpaydin, 2010; Bishop, 2006), such as the CV misclassification rate. This case follows a wrapper-based procedure and, then, a large iterative computations under CV techniques must be done by training and evaluating the classifier for each possible transformed data space. Due to the high variability in EEG features and the small number of available data samples in the BCI experiments (Lotte et al., 2007), the Leave-One-Out (LOO) may have to be used as CV technique (McFarland et al., 2011). Whereas, the LOO-CV approach has been discarded in most implementations for BCI because it is very costly in computational terms (McFarland et al., 2011; Popescu, Fazli, Badower, Blankertz, & Müller, 2007). However, one of our recent works shows that LOO-CV can be computed with very reduced complexity for designing of BCI systems based on linear discriminants (Rodríguez-Bermúdez et al., 2013a). It has been achieved by computing the Allen's PREdiction Sum of Squares (PRESS) statistic (Rodríguez-Bermúdez et al., 2013a), which provides a fast estimation of the LOO-CV error for linear models (Bontempi, 2011). In this paper, we propose to extend this efficient methodology for choosing the number of dimensions in the transformed data space.

Consider that the original input data $\mathbf{X}$ is projected into $\mathbf{Z}$ with dimensionality r and, then, for each sample $\mathbf{z}_n$, the output of the linear discriminant in the projected space is $o_n = \mathbf{v}^\top \mathbf{z}_n$, being $\mathbf{v}$ the obtained weight vector of the linear discriminant in the embedded space. Thus, the PRESS statistic is computed as follows:

$$E_{PRESS,r} = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{t_n - o_n}{1 - hat_{n,n}} \right)^2. \quad (22)$$

$hat_{n,n}$ denotes the nth value of the HAT-matrix (Bontempi, 2011) in the transformed space with r dimensions, which transforms $\mathbf{t}$ into $\mathbf{o}$:

$$\mathbf{o} = \mathbf{Z}\mathbf{w} = \mathbf{Z}\mathbf{Z}^\dagger \mathbf{t} = \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{t} = \mathrm{HAT} \cdot \mathbf{t}. \quad (23)$$

Note that (22) only needs the diagonal of the HAT-matrix, which can be easily obtained by the row-wise dot-product between $\mathbf{Z}$ and $\mathbf{Z}^\dagger$. The optimal number of dimensions can be found by estimating the LOO error for different transformation spaces and, then, selecting the dimensionality (r^*, with $r^* \leqslant D$) such that minimizes the LOO error:

$$r^* = \underset{r \in \{1,\ldots,D\}}{\arg\min} E_{PRESS,r}. \quad (24)$$

4. Experiments

As it is described in Section 2.1, this work experimentally analyzes five sets of EEG data extracted from five different novice users of MI-based BCI systems. The experiments have been done in order to show that the appropriate use of FS and FT methods in BCI data allows to increase the obtained performance by state-of-the-art techniques for designing MI-based BCI systems from first sessions with novice users. For doing this, we consider several well-known and widespread FE procedures to construct different EEG feature spaces, which are used as input data to train the linear discriminant (the most widely-used classifier in BCI systems) using the Least Squares (LS) approach. Firstly, we have evaluated the linear classifier using the extracted features from each individual method (BP, HJ and AAR) and, also, with all input features. Then, and from the input data space defined by the three FE methods, standard FS procedures are applied: sequential backward selection (SBS) and sequential forward selection (SFS). Lastly, we explore the three FT techniques: PCA, LPP and LFDA. Note that FDA is logically not included in the experiments as feature transformation method because it gives an equivalent solution than linear discrimination with all input features, as we have above mentioned. The implemented methodology for designing BCI systems has been done in MATLAB 7.11(R2010b) environment and, particularly, its FE module makes uses of the BioSig Toolbox. With respect to the FT techniques, we use the implementation given in van der Maaten et al. (2009) for PCA and LPP; whereas, the LFDA has been implemented according to Sugiyama (2007). All simulations have been carried out in the same computer with 4 GB of memory and 2.67 GHz processor.

For the performance evaluation of the different analyzed approaches, a LOO-CV procedure has been used to make fair experimental comparisons: it makes the best use of the available data and it is insensitive to random effects of sub-sampling. Although the computational cost of the traditional LOO-CV approach can

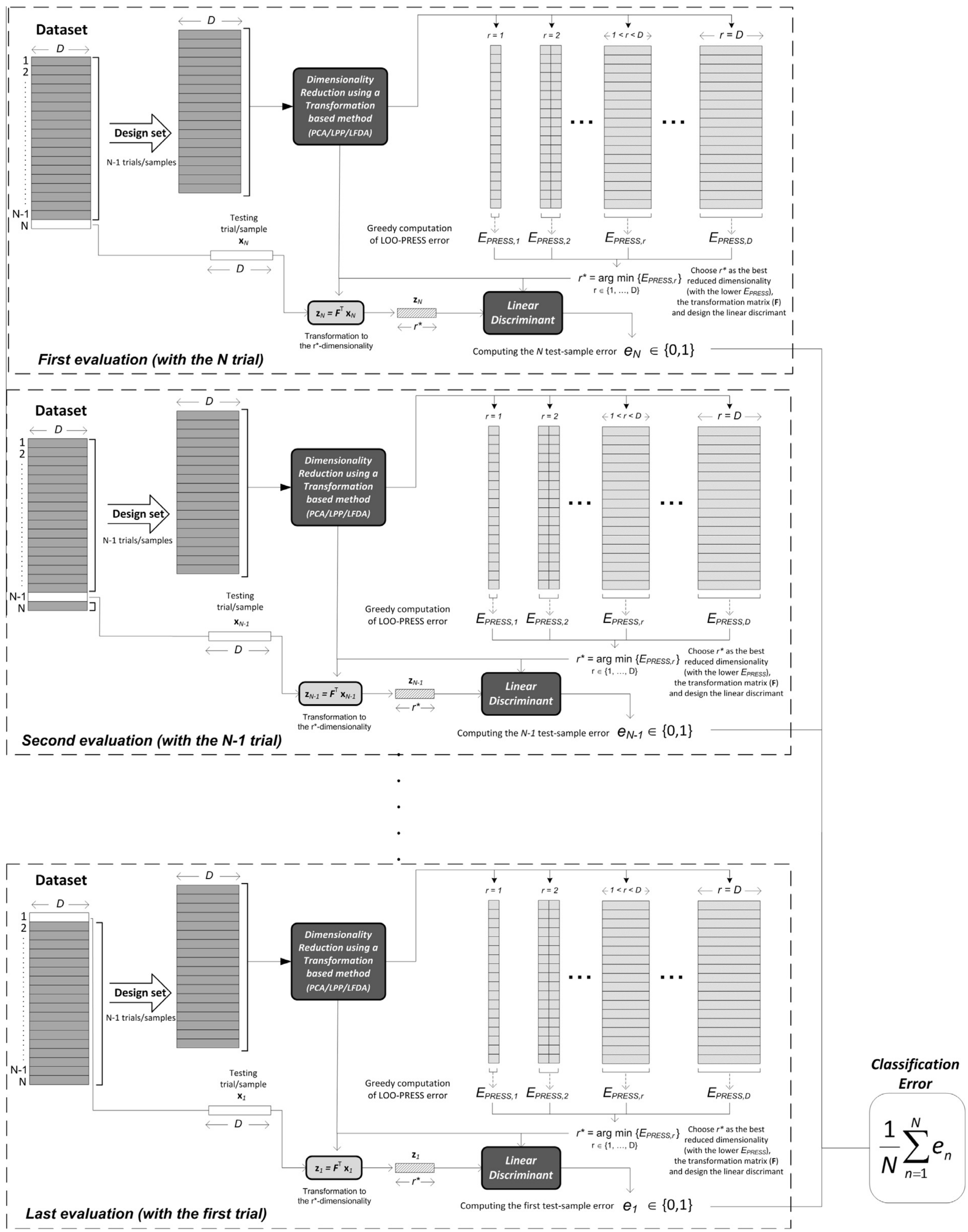

Fig. 2. Scheme of the procedure for performance evaluation of the proposed methodology using a nested LOO-CV.

be very high, this work computes the LOO-CV error (without doing the corresponding iterations) in a fast way thanks to the closed matrix-based formula for the Allen's PRESS statistic. FS (SBS and SFS) and FT (PCA, LPP and LFDA) methods need to perform model selection, i.e., they need to choose the most appropriate input variables and its corresponding weight values for a linear classifier. Therefore, a nested CV approach has to be applied: an inner CV loop to perform the feature/model selection and an outer separate CV to compute an estimate of the classification error. In principle, it is best to apply a nested LOO-CV approach. However, the nested LOO-CV is impractical for search-based feature selection procedures (SFS and SBS) and, then, the resubstitution technique has been applied for the inner CV loop. In contrast, in our proposed framework with feature transformation methods, the nested LOO

Table 1
LOO-CV classification accuracy results (in %) using a linear discriminant for the five subjects by considering the three individual FE methods (BP, HJ and AAR) and all extracted features. After that, from the input data space defined by all features, FS procedures are evaluated: SBS and SFS. Lastly, we consider FT approaches: PCA, LPP and LFDA. In each subject, best decision result is shown in bold face. Note that the total LOO-CV computational time (in seconds) is also shown in brackets.

Features	Subject A	Subject B	Subject C	Subject D	Subject E	Mean
BP	55.0(1.1e−2)	58.7(1.2e−2)	53.7(1.1e−2)	56.2(1.1e−2)	76.2(1.1e−2)	59.9(1.1e−2)
HJ	57.5(3.8e−2)	50.7(3.7e−2)	56.2(3.7e−2)	50.2(3.8e−2)	66.2(3.7e−2)	58.2(3.7e−2)
AAR	50.7(4.1e−2)	53.2(4.0e−2)	57.5(4.0e−2)	58.7(4.1e−2)	57.5(4.0e−2)	55.5(4.0e−2)
All	56.3(4.7e−2)	58.8(4.7e−2)	51.3(4.7e−2)	58.7(4.7e−2)	71.2(4.7e−2)	59.3(4.7e−2)
SBS	57.5(6.6e−1)	58.8(4.5e−1)	53.7(6.8e−1)	70.0(6.7e−1)	76.2(7.4e−1)	63.2(6.4e−1)
SFS	65.0(3.7e−1)	72.5(5.3e−1)	63.7(3.4e−1)	60.2(3.1e−1)	76.2(3.3e−1)	67.5(3.8e−1)
PCA	63.7(7.4e−2)	70.0(7.4e−2)	60.0(7.4e−2)	65.1(7.3e−2)	65.0(7.5e−2)	64.7(7.4e−2)
LPP	62.5(5.8e−2)	70.0(5.8e−2)	61.3(5.9e−2)	60.0(5.8e−2)	71.2(5.9e−2)	65.0(5.8e−2)
LFDA	78.8(5.4e−2)	77.5(5.3e−2)	67.5(5.4e−2)	78.8(5.5e−2)	83.8(5.4e−2)	77.3(5.4e−2)

CV is feasible given that, for each sample of the outer CV loop, the inner CV loop is done directly using the Allen's PRESS statistic.

Fig. 2 gives a graphical scheme of the performance evaluation using a nested LOO-CV. From this figure, for each trial of the outer loop, the dataset composed of the remaining $N-1$ samples is iteratively transformed by the corresponding FT method (PCA or LPP or LFDA) from one dimension to the original input dimensionality (22 features in this paper) and, then, we choose the number of dimensions that gives the lower PRESS statistic.

4.1. Results and discussions

Table 1 shows the LOO-CV accuracy results (in %) obtained by a LS linear discriminant with different approaches to define the EEG feature data space and, also, it is shown its corresponding computational time in brackets.

According to these results, there is not an individual FE method that provides the best performance for all subjects. In particular, the most discriminative features are the BP features for Subjects B and E; the HJ features for Subject A; and the AAR features for Subjects C and D. From the last column of Table 1, the BP method can be considered on average more suitable than HJ and AAR. The use of all input features does not ensure an improvement of the classification accuracy results. In fact, the simultaneous use of BP, HJ and AAR features as inputs of the linear classifier worsens the obtained performance for Subjects A, C and E. Besides, from the first four rows of Table 1, it should be emphasized that the obtained results by conventional approaches in these BCI novice users are logically not very good -except in Subject E- because it was the first experience with the BCI system and, then, these first acquired EEG signals are of poor quality for classification purposes. As it is widely known in the BCI literature, poor classification results usually tend to produce a rejection in the subject to reuse the system. Due to this work explores FS and FT techniques in order to exploit the discriminative information in the three FE methods and, then, better classification results can be given by the linear discriminant.

First, standard sequential search procedures (SBS and SFS) are evaluated and, according to the mean results of Table 1, both of them increase the classifier performance. These results show the usefullness of reducing the original input data space defined by the three FE methods. In particular, SFS provides better results than SBS for most subjects, except Subject D. Obviously, both procedures entail an additional computational time, being SBS slower than SFS because it starts the sequential search by considering all features. After that, FT methods (PCA, LPP and LFDA) are explored and, considering the mean results of Table 1, all of them gives better performance results with respect to the original input feature space. Considering the unsupervised FT methods, LPP slightly outperforms PCA in average but, in some subjects (A and D), PCA works better. Overall, the best approach is LFDA: this supervised FT method clearly outperforms the obtained results in all cases. The improvement is very high in classification terms and without increasing significantly the computational requirements. According to these experimental results, LFDA is a promising method for DR in MI task classification.

Once the performance results have been discussed, Fig. 3 depicts the LOO-CV classification accuracy rate by a linear discriminant as function of the dimensionality of the projected spaces (r) with PCA, LPP and LFDA. As it is expected due to high variability of the EEG signals among different subjects, the evolution/trend of the classification performance with respect to r in each method varies a lot for each subject. For example, the obtained classification accuracies with only the first transformed variable are very different and poor in Subject A; meanwhile, in Subject C, the classifier provides similar performance results with the first variable from PCA, LPP and LFDA. Naturally, the solutions of PCA, LPP and LFDA coincide when the embedding dimensionality is equal to 22 (the dimensionality of the original feature space). From these graphs, LFDA has excellent results and it overall outperforms PCA and LPP for the five subjects under study. Besides, the advantages of LFDA are clearer for lower dimensionalities and, thus, it tends to provide very reduced datasets that are more easily separable by the linear classifier.

Finally, and in order to give an illustrative overview of the behavior of the FT methods in data visualization tasks and its corresponding projected spaces, Fig. 4 includes the 2-D scatter plots obtained by PCA, LPP and LFDA in Subject E. For each method, the two-dimensional space is defined by those first two variables which have been selected according to the LOO-CV technique introduced in Section 3.5. As it has been mentioned in Section 3, PCA and LPP do not include the class label information and, thus, its projected feature spaces may be not appropriate for classification purposes -as examples, see Fig. 4(a) and (b). In contrast, the supervised feature transformation obtained by LFDA separates the two classes reasonably well -see Fig. 4(c) and it makes easier the classification stage.

5. Conclusions and further work

The performance of a MI-based BCI system depends on the quality of available EEG data and the user attitude. These aspects are critical for untrained subjects during its first experiments, where the classification results are not usually good due to the poor quality of the acquired EEG signals. Therefore, it is an important issue to adaptively design BCI systems for each subject aimed at providing a good enough performance from the beginning and, then, those BCI users with initial difficulties can reduce their frustration/rejection feelings. Besides, due to the high variability of EEG signals, the BCI system should be able to adapt its internal

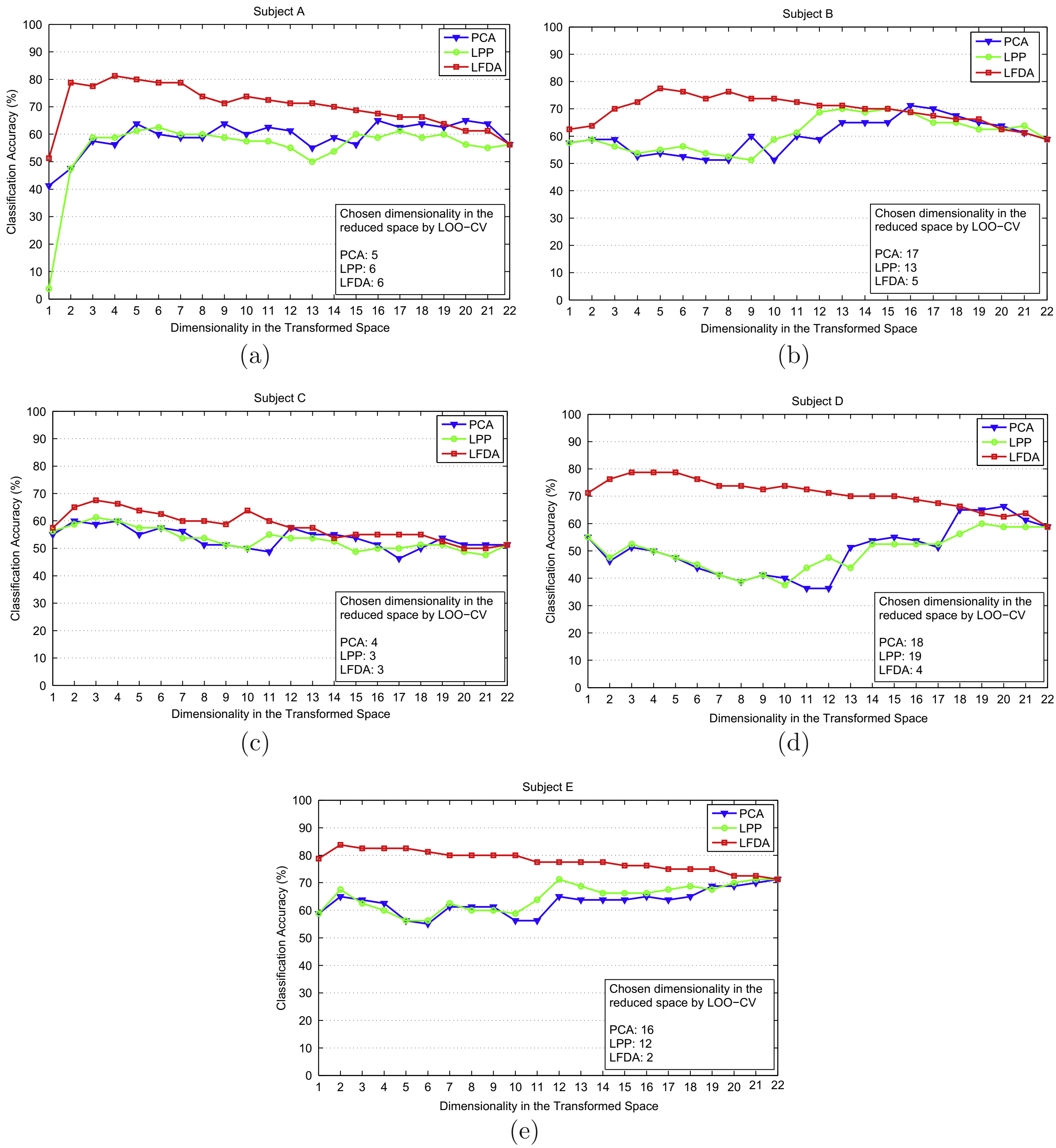

Fig. 3. Evolution of the test classification accuracy (in the outer loop of the LOO-CV procedure) with respect the number of dimensions in the transformed space by PCA, LPP and LFDA for the five subjects under study (A-E). In addition, for each subject, it is shown the average number of chosen dimensions in each mapping according to the PRESS statistic.

structure/operation for each new trial as far as possible. In order to achieve these goals, we have studied dimensionality reduction methods for the adaptive design of BCI systems from first sessions with novice users. Sequential backward and forward selection (SFS and SBS) procedures have been analyzed, along with three transformation-based methods: PCA, LPP and LFDA. Particularly, it should be remarked that, according to the state-of-the-art, this is the first work that analyzes and applies LFDA for transforming EEG features. Besides, as an another important contribution, we introduce an efficient LOO-CV technique for choosing the embedded dimensionality in the projected space, which makes suitable the adaptation of the BCI system for each trial.

Experiments have been done on five BCI-EEG datasets from different novice users in their first motor imagery sessions. The original input feature space is defined by three feature extraction methods: Band Power, Hjorth and AAR. An experimental comparative study has been conducted on feature selection (SFS and SBS) and transformation-based (PCA, LPP and LFDA) methods. Overall,

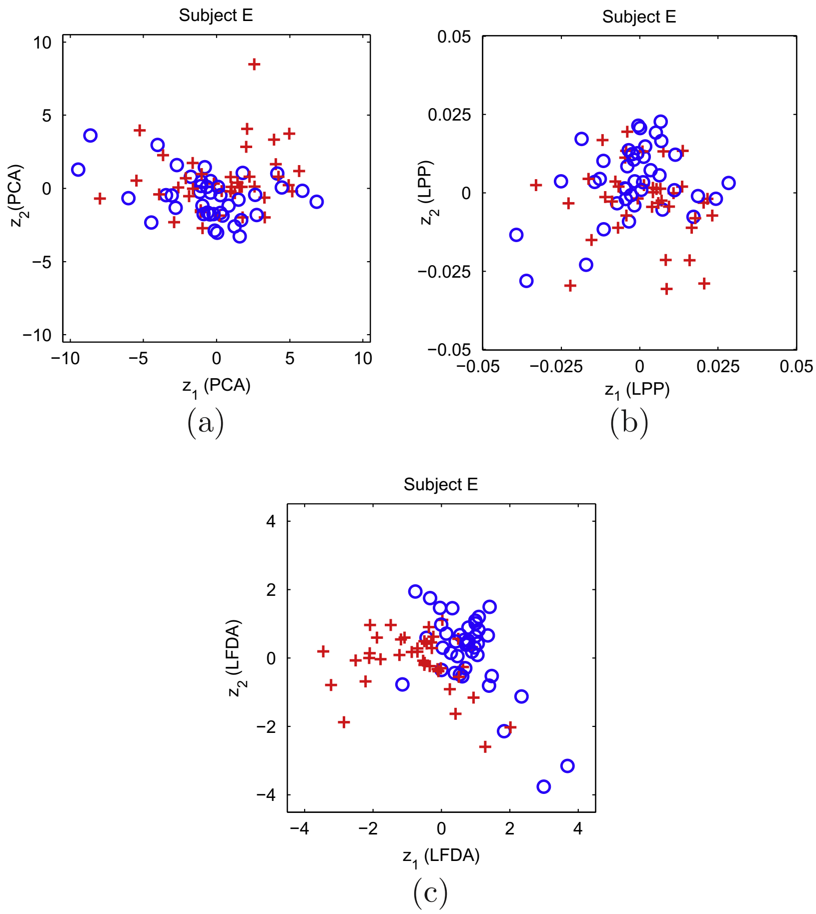

Fig. 4. Two-dimensional scatter plot by considering the two most significant variables in the transformed space for Subject E. (a): PCA, (b): LPP, (c): LFDA. Input vectors from C_1 and C_2 are denoted with 'circles' and 'crosses', respectively.

obtained off-line results show that the use of dimensionality reduction techniques increases the classifier performance with respect to the original feature space. Particularly, it has been observed that LFDA provides excellent classification results for all subjects with reduced computational complexity.

We believe that these promising results given by the LFDA approach can help to enhance the adaptive design of BCI systems and, particularly, with novice subjects. Future research will pay much attention to on-line experiments with more EEG channels and subjects. Besides, a detailed analysis of different time course approaches during feature extraction is also an ongoing study. Another further research work is the extension to other BCI paradigms and applications; in particular, vigilance estimation (Shi & Lu, 2013), emotion recognition (Wang, Nie, & Lu, 2014) and mental fatigue (Liu, Zhang, & Zheng, 2010).

References

Ahn, M., Cho, H., Ahn, S., & Jun, S. C. (2013). High theta and low alpha powers may be indicative of BCI-illiteracy in motor imagery. *PLoS One, 8*(11), e80886. http://dx.doi.org/10.1371/journal.pone.0080886.

Alpaydin, E. (2010). *Introduction to machine learning*. The MIT Press.

Anderson, C., Stolz, E., & Shamsunder, S. (1998). Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Transactions on Biomedical Engineering*, 0018-9294, *45*(3), 277–286.

Atyabi, A., Luerssen, M. H., & Powers, D. M. W. (2013). PSO-based dimension reduction of EEG recordings: Implications for subject transfer in BCI. *Neurocomputing, 119*(0), 319–331.

Bashashati, A., Fatourechi, M., Ward, R. K., & Birch, G. E. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering, 4*(2), R32–R57.

Bhattacharyya, S., Sengupta, A., Chakraborti, T., Konar, A., & Tibarewala, D. (2014). Automatic feature selection of motor imagery EEG signals using differential evolution and learning automata. *Medical & Biological Engineering & Computing, 52*(2), 131–139.

Billinger, M., Brunner, C., & Neuper, C. (2010). Classification of adaptive autoregressive models at different sampling rates in a motor imagery-based BCI. In *Fourth international BCI meeting*. Pacific Grove, CA, USA.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. 0387310738. Springer.

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The Berlin brain-computer interface: Non-medical uses of BCI technology. *Frontiers in Neuroscience, 4*, 1–17.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Muller, K.-R. (2008). Optimizing spatial filters for Robust EEG single-trial analysis. *IEEE Signal Processing Magazine, 25*(1), 41–56.

Bontempi, G. (2011). Statistical foundations of machine learning. Université Libre de Bruxelles.

Cabrera, A., Farina, D., & Dremstrup, K. (2010). Comparison of feature selection and classification methods for a brain-computer interface driven by non-motor imagery. *Medical and Biological Engineering and Computing*, 0140-0118, *48*, 123–132.

Chen, H.-L., Liu, D.-Y., Yang, B., Liu, J., & Wang, G. (2011). A new hybrid method based on local Fisher discriminant analysis and support vector machines for hepatitis disease diagnosis. *Expert Systems with Applications, 38*(9), 11796–11803.

Chung, F. R. K. (1997). *Spectral graph theory*. 0821803158. Providence, R.I.: American Mathematical Society.

Corralejo, R., Hornero, R., & Alvarez, D. (2011). Feature selection using a genetic algorithm in a motor imagery-based Brain Computer Interface. In *2011 Annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 7703 –7706).

Delgado-Saa, J. F., & Cetin, M. (2011). Hidden conditional random fields for classification of imagery motor tasks from EEG data. In *19th European signal processing conference (EUSIPCO 2011)* (pp. 1377–1381). Barcelona, Spain.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification*. Wiley Interscience.

Erfanian, A., & Erfani, A. (2004). EEG-based brain-computer interface for hand grasp control: Feature extraction by using ICA. In *Ninth annual conference of the international FES society*.

Fruitet, J., McFarland, D. J., & Wolpaw, J. R. (2010). A comparison of regression techniques for a two-dimensional sensorimotor rhythm-based brain-computer interface. *Journal of Neural Engineering, 7*(1), 16003.

Fukunaga, K. (1990). Introduction to statistical pattern recognition. *Computer science & scientific computing*. 0122698517. Academic Press.

Gan, J. (2006). Feature dimensionality reduction by manifold learning in brain-computer interfaces. In *Third international workshop on brain-computer interfaces* (pp. 28–29). Graz, Austria.

Gao, J., Kwan, P. H., & Huang, X. (2009). Comprehensive analysis for the local Fisher discriminant analysis. *International Journal of Pattern Recognition and Artificial Intelligence, 23*(6), 1129–1143.

Garrett, D., Peterson, D., Anderson, C., & Thaut, M. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11*(2), 141–144.

Guerrero-Mosquera, C., Verleysen, M., & Navia Vazquez, A. (2010). EEG feature selection using mutual information and support vector machine: A comparative analysis. In *32nd Annual international IEEE EMBC conference* (pp. 4946–4949).

Guger, C., Schlogl, A., Neuper, C., Walterspacher, D., Strein, T., & Pfurtscheller, G. (2001). Rapid prototyping of an EEG-based brain-computer interface (BCI). *IEEE Transactions on Rehabilitation Engineering Neural Systems and Rehabilitation Engineering, 9*(1), 49–58.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 1532-4435*, 3*, 1157–1182.

He, L., Hu, Y., Li, Y., & Li, D. (2013). Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG. *Neurocomputing, 121*, 423–433.

He, X., & Niyogi, P. (2004). Locality preserving projections. In S. Thrun, L. Saul, & B. Scholkopf (Eds.). *Advances in neural information processing systems* (Vol. 16). Cambridge, MA: MIT Press.

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 0013-4694*, 29*(3), 306–310.

Hsu, W.-Y. (2012). Fuzzy hopfield neural network clustering for single-trial motor imagery EEG classification. *Expert Systems with Applications*, 0957-4174*, 39*(1), 1055–1061.

Hsu, W. -Y. (2013). Improving classification accuracy of motor imagery EEG using genetic feature selection. *Clinical EEG and Neuroscience* (in press). http://dx.doi.org/10.1177/1550059413491559.

Huster, R. J., Mokom, Z. N., Enriquez-Geppert, S., & Herrmann, C. (2014). Brain-computer interfaces for EEG neurofeedback: Peculiarities and solutions. *International Journal of Psychophysiology, 91*(1), 36–45.

Koprinska, I. (2009). Feature Selection for brain-computer interfaces. In *International workshop on new frontiers in applied data mining (PAKDD), LNCS* (Vol. 5669, pp. 106–117).

Krusienski, D. J., Grosse-Wentrup, M., Galán, F., Coyle, D., Miller, K. J., Forney, E., et al. (2011). Critical issue in state-of-the-art brain-computer interface signal processing. *Journal of Neural Engineering, 8*, 1–8.

Lee, F., Scherer, R., Leeb, R., Schlgl, A., Bischof, H., & Pfurtscheller, G. (2004). Feature mapping using PCA, locally linear embedding and isometric feature mapping for EEG-based brain computer interface. In *Proceedings of the 28th AAPR workshop digital imaging in media and education* (pp. 189–196).

Lee, J. A., & Verleysen, M. (2007). *Nonlinear dimensionality reduction*. Springer.

Liang, N.-Y., Saratchandran, P., Huang, G.-B., & Sundararajan, N. (2006). Classification of mental task from EEG signals using extreme learning machine. *International Journal of Neural Systems, 16*(01), 29–38.

Liu, J., Zhang, C., & Zheng, C. (2010). EEG-based estimation of mental fatigue by using KPCA-HMM and complexity parameters. *Biomedical Signal Processing and Control, 5*(2), 124–130.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 1741-2560*, 4*(2), R1–R13.

McCann, M. T., Thompson, D. E., Syed, Z. H., & Huggins, J. E. (2014). Electrode subset selection methods for an EEG-based P300 brain-computer interface. In *Disability and Rehabilitation: Assistive Technology* (pp. 1–5). http://dx.doi.org/10.3109/17483107.2014.884174.

McFarland, D. J., Sarnacki, W. A., & Wolpaw, J. (2011). Should the parameters of a BCI translation algorithm be continually adapted? *Journal of Neuroscience Methods, 199*(1), 103–107.

Muller, K.-R., Anderson, C., & Birch, G. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11*(2), 165–169.

Müller, K.-R., Krauledat, M., Dornhege, G., Curio, G., & Blankertz, B. (2004). Machine learning techniques for brain-computer interfaces. *Biomedical Technology, 49*(1), 11–22.

Naeem, M., Brunner, C., & Pfurtscheller, G. (2009). Dimensionality reduction and channel selection of motor imagery electroencephalographic data. *Computational Intelligence and Neuroscience, 2009*, 537504. http://dx.doi.org/10.1155/2009/537504.

Nicolas-Alonso, L., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors, 2*, 1211–1279.

Oh, S.-K., Yoo, S.-H., & Pedrycz, W. (2013). Design of face recognition algorithm using PCA-LDA combined for hybrid data pre-processing and polynomial-based RBF neural networks: Design and its application. *Expert Systems with Applications, 40*(5), 1451–1466.

Pfurtscheller, G., Brunner, C., Schlgl, A., & da Silva, F. L. (2006). Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage, 31*(1), 153–159.

Popescu, F., Fazli, S., Badower, Y., Blankertz, B., & Müller, K. R. (2007). Single trial classification of motor imagination using 6 dry EEG electrodes. *PloS One, 2*(7), e637+.

Rejer, I., & Lorenz, K. (2013). Genetic algorithm and forward method for feature selection in EEG feature space. *Journal of Theoretical and Applied Computer Science, 7*(2), 72–82.

Rodríguez-Bermúdez, G., García-Laencina, P. J., & Roca-Dorda, J. (2013a). Efficient automatic selection and combination of EEG features in least squares classifiers for motor-imagery brain computer interfaces. *International Journal of Neural Systems, 23*(04), 1350015.

Rodríguez-Bermúdez, G., García-Laencina, P., Roca-González, J., & Roca-Dorda, J. (2013b). Efficient feature selection and linear discrimination of EEG signals. *Neurocomputing, 115*, 161–165.

Sabeti, M., Boostani, R., Katebi, S., & Price, G. (2007). Selection of relevant features for EEG signal classification of schizophrenic patients. *Biomedical Signal Processing and Control*, 1746-8094*, 2*(2), 122–134.

Sanei, D., & Chambers, J. (2008). *EEG signal processing*. John Wiley & Sons.

Sardouie, H. S., & Shamsollahi, M. B. (xxxx). Discriminating MEG signals recorded during hand movements using selection of efficient features. *Frontiers in Neuroscience*, 6 (42). http://dx.doi.org/10.3389/fnins.2012.00042.

Schlögl, A. (2000). *The electroencephalogram and the adaptive autoregressive model: Theory and applications*. Shaker Verlag.

Serre, D. (2002). *Matrices: Theory and applications*. New York: Springer.

Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P., & Müller, K.-R. (2006). Towards adaptive classification for BCI. *Journal of Neural Engineering, 3*(1), R13–R23.

Shi, L.-C., & Lu, B.-L. (2013). EEG-based vigilance estimation using extreme learning machines. *Neurocomputing, 102*, 135–143.

Siuly, S., Li, Y., & Wen, P. (2014). Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain-computer interface. *Computer Methods and Programs in Biomedicine* (in press). http://dx.doi.org/10.1016/j.cmpb.2013.12.020.

Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *Journal of Machine Learning Research, 8*, 1027–1061.

Tong, S., & Thankor, N. V. (2009). *Quantitative EEG analysis methods and clinical applications*. Artech House.

van der Maaten, L. J. P., Postma, E. O., & van den Herik, H. J. (2009). Dimensionality reduction: A comparative review. Technical report, Tilburg University.

Vidaurre, C., Sannelli, C., Müller, K.-R., & Blankertz, B. (2011). Co-adaptive calibration to improve BCI efficiency. *Journal of Neural Engineering, 8*(2), 025009 (8pp).

Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing, 129*, 94–106.

Yu, J.-B. (2011). Bearing performance degradation assessment using locality preserving projections. *Expert Systems with Applications, 38*(6), 7440–7450.

Yu, X., Chum, P., & Sim, K.-B. (2014). Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system. *Optik – International Journal for Light and Electron Optics*, 0030-4026*, 125*(3), 14981502.

Zelnik-Manor, L., & Perona, P. (2005). Self-tuning spectral clustering. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.). *Advances in neural information processing systems* (Vol. 17, pp. 1601–1608). Cambridge, MA: MIT Press.

Zhang, R., Xu, P., Liu, T., Zhang, Y., Guo, L., Li, P., et al. (2013). Local temporal correlation common spatial patterns for single trial EEG classification during motor imagery. *Computational and Mathematical Methods in Medicine, 2013*, 591216. http://dx.doi.org/10.1155/2013/591216 (7 pages).