Motor Imagery Classification using Feature Relevance Analysis: An Emotiv-based BCI System

J. Hurtado-Rincon*, S. Rojas-Jaramillo*,Y. Ricardo-Cespedes*
, Andrés M. Álvarez-Meza*, and Germán Castellanos-Domínguez*
*Signal Processing and Recognition Group, Universidad Nacional de Colombia sede Manizales, Manizales, Colombia
e-mail: jvhurtador,srojasj,yrcespedesv,amalvarezme,cgcastellanosd {@unal.edu.co}

***Abstract*—Brain Computer Interfaces (BCI) have been emerged as an alternative to support automatic systems able to interpret brain functions, commonly, by analyzing electroencephalography (EEG) recordings. In this work, a time-series discrimination methodology, called Motor Imagery Discrimination by Relevance Analysis (MIDRA), is presented to support the development of BCI from EEG data. Particularly, a Motor Imagery (MI) paradigm is studied, i.e., imagination of left-right hand movements. In this sense, a feature relevance analysis strategy is presented to select representing characteristics using a variability criterion. Besides, short-time parameters are estimated from EEG data by considering both time and time-frequency representations to deal with non-stationary dynamics. MIDRA is assessed on two different BCI databases, a well-known MI data and an Emotiv-based dataset. Attained results showed that MIDRA enhances the BCI system performance in comparison with benchmark methods by suitable ranking the input feature set. Moreover, applying MIDRA in a BCI based on the Emotiv device is a straightforward alternative for dealing with MI paradigms.**

I. Introduction

Brain Computer Interfaces (BCI) have been emerged as an alternative to support the development of devices able to measure and interpret brain functions [1], [2]. Traditionally, BCI are used to help people with disability by means of the analysis of the human sensorimotor functions, which are based on the paradigm in cognitive neuroscience named as Motor Imagery (MI), e.g., imagination of hand movements, whole body activities, relaxation, etc. [3]. Commonly, electroencephalography - EEG recordings are employed in BCI to extract the brain electrical activity for monitoring the process of thinking. Nonetheless, the analysis of EEG signals exhibits some challenges regarding to intrinsic non-stationary data dynamics, being necessary to construct a suitable EEG discrimination methodology for enhancing BCI system accuracy. Moreover, representing EEG time-series leads to high-dimensional spaces that include redundant information about the studied cognitive paradigm.

Devoted to feature representation for BCI systems using EEG data, different methods have been proposed, such as: 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 (CWT and DWT) based features [4], [5], [6], [7], [8], [9]. Although many features may be extracted from EEG, their main restriction is how to extract relevant features due to they may contain redundant information. Particularly, several approaches have been used to identify the relevance of the computed features in BCI systems, as given in [4], [10], [5], [11]. 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 over-fitted results.

Here, a time-series discrimination methodology, termed Motor Imagery Discrimination by Relevance Analysis (MIDRA), is proposed as a tool to support the development of BCI from EEG recordings. Regarding this, a variability based criterion is employed to analyze the relevance of EEG features. Particularly, short-time features are estimated from EEG data by considering both time and time-frequency representations to deal with the non-stationary nature of EEG recordings. Namely, spectral, Hjorth, and wavelet based characteristics are computed from each EEG channel to learn the main patterns in MI paradigms. MIDRA allows ranking the input feature set to highlight the most important information that helps to recognize different MI classes, i.e., imagination of left-right hand movements. Our approach is tested on two different BCI databases, a well-known MI data and an Emotiv-based dataset. Attained results showed that MIDRA enhances the BCI system performance in comparison with benchmark methods by suitable ranking the input feature set. Moreover, applying MIDRA in a BCI based on the Emotiv device is a straightforward alternative for dealing with MI paradigms.

II. Motor Imagery Discrimination based on Relevance Analysis

A. EEG Feature Representation

Let $\Psi{=}\{\boldsymbol{Y}_r : r{=}1,\dots,R{\in}\mathbb{N}\}$ be a set of R EEG data trials, where $\boldsymbol{Y}_r{\in}\mathbb{R}^{C\times T}$ is the r-th trial matrix with $C{\in}\mathbb{N}$ channels and $T{\in}\mathbb{N}$ samples with sample frequency $F_s{\in}\mathbb{R}$. Moreover, let $l_r{\in}\{-1,+1\}$ be the Motor Imagery (MI) paradigm condition (class label) of $\boldsymbol{Y}_r$ [10]. To carry out discrimination of MI paradigms, we calculate long-term features from short-time parameters characterizing EEG, aiming to deal with the EEG non-stationary dynamics. Namely, given an EEG trial matrix $\boldsymbol{Y}$ the following short-time parameters are computed.

Spectral parameters. Power Spectral Density (PSD) vector $\boldsymbol{p}{\in}\mathbb{R}^{B_s}$ is computed for each channel $\boldsymbol{y}{\in}\mathbb{R}^{T}$ in $\boldsymbol{Y}$, where $B_s{\in}\mathbb{N}$ is the number of frequency bins. Thereby, $\boldsymbol{y}$ is split into

978-1-4799-7666-9/14/$31.00

$M \in \mathbb{N}$ overlapped segments of length $L \in \mathbb{N}$, and a piecewise stationary assumption is imposed by means of a smooth time weighting window $\boldsymbol{w} \in \mathbb{R}^L$ due to the non-stationary nature of EEG data [4]. Hence, the windowed segments $\boldsymbol{v}^m \in \mathbb{R}^L$ ($m=1, \ldots, M$) are extracted. Thus, $\boldsymbol{p}$ is estimated from the modified periodogram vector $\boldsymbol{u} \in \mathbb{R}^{F_s/2}$ by using the Discrete Fourier Transform.

Hjorth parameters. The following time-domain based parameters are estimated from each windowed segment $\boldsymbol{v}^m$ [5]. *i) Activity*: $\boldsymbol{\sigma}^2_{\boldsymbol{v}} \in \mathbb{R}^M$, where each element is directly described by the signal power variance $\sigma^2_m = \mathrm{var}(\boldsymbol{v}^m)$. *ii) Mobility*: $\boldsymbol{\phi}_{\boldsymbol{v}} \in \mathbb{R}^M$, where each element is a measure of the signal mean frequency $\phi_m = \sqrt{\mathrm{var}(\boldsymbol{v}^{m\prime})/\mathrm{var}(\boldsymbol{v}^m)}$, being $\boldsymbol{v}'$ the derivative of $\boldsymbol{v}$. *iii) Complexity*: $\boldsymbol{\vartheta}_{\boldsymbol{v}} \in \mathbb{R}^M$, having each element measuring frequency variations as the deviation of the signal from the sine shape as $\vartheta_m = \phi'_m/\phi_m$, where ϕ'_m is the derivative of ϕ_m.

Time-Frequency parameters. Wavelet-based transform are employed to analyze EEG data holding non-stationary spectral components [4]. Here, both Continuous Wavelet Transforms (CWT) and Discrete Wavelet Transforms (DWT) are considered. CWT-based transformation quantifies similarity between a given equally sampled time-series at time spacing $\delta_t \in \mathbb{R}$ and a fixed base function $\gamma(\eta)$ (mother wavelet), which is ruled by the dimensionless parameter $\eta \in \mathbb{R}$. Therefore, the CWT vector $\boldsymbol{\varsigma}^g \in \mathbb{C}^T$ is extracted from channel $\boldsymbol{y}$ at scale $g \in \mathbb{R}$ as follows: $\varsigma^g_t = \sum_{\tau=1}^{T} y_\tau \gamma^*((\tau - t)\delta_t/g)$, where $(^*)$ notes the complex conjugate. In this work, both procedures of Wavelet scaling g and translating through the localized time index $t \in T$ are used to model amplitude time variations. Alike, the DWT can provide precise time-frequency information about $\boldsymbol{y}$ as a multi-resolution and non-redundant representation [6]. In this sense, the detail vector $\boldsymbol{b}^j \in \mathbb{C}$ at level j as: $b^j_t = \sum_{k \in \mathbb{Z}} a_{j,k}\psi_{j,k}(t)$, where $a_{j,k} = \sum_{t \in T} y_t h_{j,k}(t)$; $a_{j,k} \in \mathbb{C}$, and being $h_{j,k}(t) \in \mathbb{C}$ the impulse response of a given wavelet filter. The DWT of $\boldsymbol{y}$ is computed for a given wavelet $\psi(\cdot)$ as $y_t = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} a_{j,k}\psi_{j,k}(t)$. Regarding this, provided DWT-based sub-bands at different scales are employed to highlight time-frequency information about $\boldsymbol{y}$. Once all the above short-time parameters are computed, some of their statistical measures can be considered to extract an input feature representation matrix $\boldsymbol{X} \in \mathbb{R}^{R \times D}$, being $D \in \mathbb{N}$ the number of considered features.

B. Relevance Analysis by Variability-based Criterion

Each row vector $\boldsymbol{x} \in \mathbb{R}^D$ in $\boldsymbol{X}$ holds a subset of $D = C \times Q$ concatenated features extracted from all provided EEG channels of a given MI trial, with $Q \in \mathbb{N}$ the number of features extracted from a single channel. Usually, provided feature representation $\boldsymbol{X}$ has a huge dimension, being necessary to extract the most discriminative features. So, a low-dimensional representation of the original feature space is accomplished commonly by using the well-known Principal Component Analysis (PCA) algorithm, which projects the data matrix over directions with the largest variance. Here, we propose to measure the relevance of each feature by taking advantage of the PCA mapping. Specifically, given 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 a relevance vector $\boldsymbol{\rho} \in \mathbb{R}^D$ as follows:

$$\boldsymbol{\rho} = \boldsymbol{E}\{|\varrho_d \boldsymbol{\alpha}_d| : \forall d \in D' \leq D\}$$

where $\varrho_d \in \mathbb{R}^+$ and $\boldsymbol{\alpha}_d \in \mathbb{R}^D$ are the eigenvalues and eigenvectors of the covariance matrix $\boldsymbol{X}^\top \boldsymbol{X}/D$, respectively [12]. The main assumption behind the relevance measure introduced in Eq. (II-B) is that the largest values of ρ_d points out to the best input attributes since they show higher overall correlations with 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 $\boldsymbol{\rho}$ can be employed to rank the original features according to its contribution for retaining the input data variability. From such a rank further learning stages for classifying MI paradigms can be supported by revealing representative EEG features.

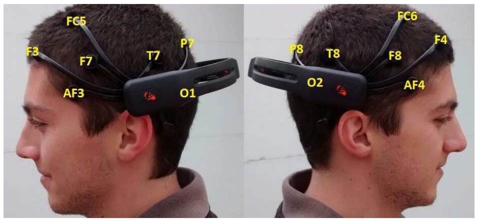
III. Experimental set-up

In order to assess the proposed Motor Imagery Discrimination methodology based on Relevance Analysis (MIDRA) as a tool to support BCI systems, experimental testing is carried out using two MI databases, which are based on the cognitive neuroscience paradigm, i.e., imagination of left-right hand movements [3].

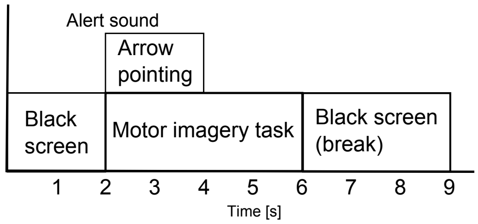
i) D1 - Dataset I of BCI competition IV (2008). The EEG signals are provided by the Berlin Brain-Computer Interface group for a BCI. Seven subjects were studied who were instructed to perform left-right MI paradigms, i.e., imagination of left-right motor action. Signals are band-pass filtered between 0.05 and $200\,Hz$ and then digitized at $1000\,Hz$. Moreover, 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 session is performed without feedback. The database contains 100 repetitions for each MI class per person. The interest section is 4 s during, when subject is instructed to perform the MI task indicated by a pointing arrow on a screen. These periods are interleaved with 2 s of blank screen and 2 s with a fixation cross in the center of the screen.

ii) D2 - Emotiv EPOC headset-based Dataset. The Emotiv EPOC sensor is a portable device that allows recording 14 high resolution EEG channels using a sample frequency of $128\,Hz$. Such a device considers the well-known 10-20 EEG electrode position, yielding to the channel distribution presented in Figure 1(a). For concrete testing two subjects are recorded. A MI protocol is carry out as described in Figure 1(b). Four runs were taken for each subject on the same day to record 40 trials in each run. Consequently, 160 trials are obtained per subject (80 corresponds to imagination of right movement and 80 to left one). Here, we aim to test the MIDRA capability for developing a straightforward BCI based on Emotiv.

MIDRA set up. The main sketch of MIDRA is presented in Figure 2. We aim to highlight latent patterns in provided EEG recordings as a time-series discrimination methodology. Regarding this, MIDRA is composed of the following stages:



(a) Emotiv EPOC channels



(b) MI protocol

Fig. 1. Emotiv EPOC dataset protocol

i) EEG preprocessing, *ii*) EEG feature representation, and *iii*) feature relevance analysis.

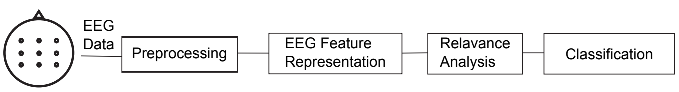


Fig. 2. MIDRA main sketch

With regard to the preprocessing stage, *D1* signals are preprocessed using the Common Spatial Patterns (CSP) and the Empirical Mode Decomposition (EMD) algorithms to highlight the main EEG temporal dynamics [13]. Since Emotiv-based experiments intend to prove MIDRA advantages for extracting relevant EEG patterns in a straightforward BCI the CSP and EMD algorithms are not applied in *D2*. Instead, EEG signals of a given subject are normalized to have zero mean along each channel and along time.

In addition, the PSD, Hjort, CWT, and DWT-based short-time parameters are computed as described in section § II-A. Since the band of interest to be analyzed in the MI paradigm is $8-30\,Hz$ as to include both the μ and β rhythms [5], the segment length value L, needed to estimate the PSD and Hjort parameters, is fixed to get a edge frequency $F_r = 8\,Hz$ ($L > F_s/F_r$ [14]). Besides, short-time instantaneous CWT amplitudes are computed using two Morlet wavelets to extract either μ or β bands: one centered at 10 and another at $22\,Hz$. Accordingly, the CWT-based feature vectors, $\boldsymbol{\varsigma}^{g_\mu}$ and $\boldsymbol{\varsigma}^{g_\beta}$, are computed for $g_\mu = 73$ and $g_\beta = 35$ to include μ and β rhythms. In case of the DWT, the following mother wavelets are considered: Symlet (orders 3, 4, 5, 6 and 7), Daubechies (orders 3, 4 and 5), and Coiflets (orders 2 and 3) [15], [16]. From each provided mother wavelet, the detail coefficient vector $\boldsymbol{b}^j$ is estimated. Again, aiming to include μ and β rhythms, the second and the third levels are employed from each detail coefficient vector. To extract the input feature set from all computed short-time parameters, Table I shows their statistical measures that are considered in accordance to [4], [5]. As a result, the feature representation matrix $\boldsymbol{X}$ holds the dimensions $R = 200$ and $D = 5183$ in *D1* ($D = 87 \times 59$), and $R = 80$ and $D = 1218$ in *D2* ($D = 87 \times 14$), where R is the number of trials and D is the number of measured features for each channel.

Furthermore, the number of dimensions D' in PCA is calculated as to have 95% of explained variance to set-up the MIDRA relevance analysis stage described in section § II-B. Afterwards, the inferred relevance vector ρ is employed to rank the original features. Then, we generate a learning curve by adding one by one the sorted features based on the $\boldsymbol{\rho}$ ranking. A 10-fold cross-validation scheme is used to validate the classification performance by using a k-nearest neighbor classifier. Moreover, the k value is fixed according to the training set accuracy.

TABLE I. EXTRACTED FEATURES FOR A GIVEN EEG CHANNEL

Parameter	Features			# Feat/ch
PSD	$\Vert\boldsymbol{S}_\mu\Vert_2^2$	$\boldsymbol{E}\{\boldsymbol{S}_\mu\}$	$\mathrm{var}(\boldsymbol{S}_\mu)$	6
	$\Vert\boldsymbol{S}_\beta\Vert_2^2$	$\boldsymbol{E}\{\boldsymbol{S}_\beta\}$	$\mathrm{var}(\boldsymbol{S}_\beta)$	
Hjort	$\max(\boldsymbol{\sigma}_{\boldsymbol{v}}^2)$	$\boldsymbol{E}\{\boldsymbol{\sigma}_{\boldsymbol{v}}^2\}$	$\mathrm{var}(\boldsymbol{\sigma}_{\boldsymbol{v}}^2)$	9
	$\max(\boldsymbol{\phi}_{\boldsymbol{v}})$	$\boldsymbol{E}\{\boldsymbol{\phi}_{\boldsymbol{v}}\}$	$\mathrm{var}(\boldsymbol{\phi}_{\boldsymbol{v}})$	
	$\max(\boldsymbol{\vartheta}_{\boldsymbol{v}})$	$\boldsymbol{E}\{\boldsymbol{\vartheta}_{\boldsymbol{v}}\}$	$\mathrm{var}(\vartheta_{\boldsymbol{v}})$	
CWT	$\max(\lvert\boldsymbol{\varsigma}^{g_\mu}\rvert)$	$\boldsymbol{E}\{\lvert\boldsymbol{\varsigma}^{g_\mu}\rvert\}$	$\mathrm{var}(\lvert\boldsymbol{\varsigma}^{g_\mu}\rvert)$	6
	$\max(\lvert\boldsymbol{\varsigma}^{g_\beta}\rvert)$	$\boldsymbol{E}\{\lvert\boldsymbol{\varsigma}^{g_\beta}\rvert\}$	$\mathrm{var}(\lvert\boldsymbol{\varsigma}^{g_\beta}\rvert)$	
DWT	$\max(\lvert\boldsymbol{b}^{j\mu}\rvert)$	$\boldsymbol{E}\{\lvert\boldsymbol{b}^{j\mu}\rvert\}$	$\mathrm{var}(\lvert\boldsymbol{b}^{j\mu}\rvert)$	66
	$\max(\lvert\boldsymbol{b}^{j\beta}\rvert)$	$\boldsymbol{E}\left\{\lvert\boldsymbol{b}^{j\beta}\rvert\right\}$	$\mathrm{var}(\lvert\boldsymbol{b}^{j\beta}\rvert)$	
Total features per channel				87

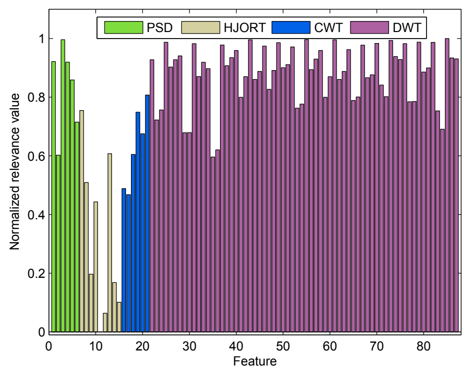
Notations: $\Vert\cdot\Vert_2$(2-norm); $\boldsymbol{E}\{\cdot\}$(expected value);$\max$(maximum); var(variance).

IV. RESULTS AND DISCUSSION

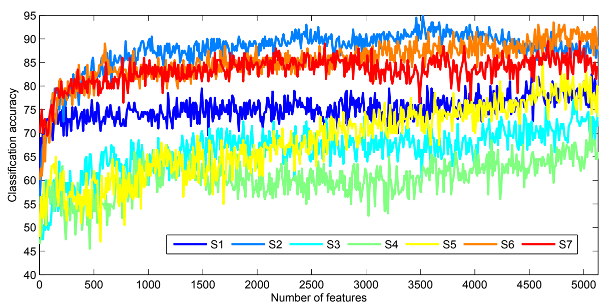
Figures 3 and IV show the attained MIDRA results for *D1* and *D2* datasets, respectively. It can be noted how the spectral and DWT-based features present high relevance values in both databases (see Figures 3(a) and 4(a)). In this regard, since spectral parameters are estimated by highlighting the μ and β rhythms, the BCI-system is able to identify MI patterns from non-stationary EEG recordings [4]. Alike, DWT-based features provide high relevance values in terms of the introduced variability criterion because their multi-resolution decomposition allows dealing with non-stationary dynamics, i.e., MI paradigms. With respect to CWT-based features, in most of the cases, they do not add relevant information to the system, since CWT analyzes the signal neglecting possible nonstationary behaviors. In the case of Hjort features although measures of the activity parameter can be considered influential, overall, Hjort-based parameters exhibit low-relevance values. In fact, due to second order statistics are only considered when calculating such features EEG nonstationary patterns are not modeled properly by them.

Now, from Figures 3(b) and 4(b), after 1000 and 400 relevant features for testing of *D1* and *D2* datasets, respectively, the MI discrimination performance becomes almost similar according to the MIDRA-based learning curves. However, in some cases, the BCI performance curves present local minimums when adding new relevant features, while then the classification accuracy grows up again. This behavior may be explained by the fact that some features may represent highly relevant attributes, but they involve redundant information, i.e., the needling phenomenon.

Finally, Table II presents the best MIDRA classification accuracy for each subject in *D1*. It can be notice that, in most of the cases, MIDRA outperforms state of the art approaches. Moreover, using only the first 1000 relevant features in *D1* (MIDRAS in Table II) is able to obtain an acceptable BCI performance. Similarly, MIDRA and MIDRAS (using the first 400 relevant features) achieve an appropriate MI classification performance in *D2*.



(a) Feature relevance values



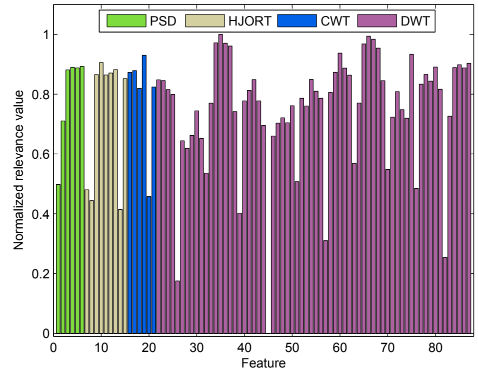
(b) Learning curve

Fig. 3. *D1* MIDRA results

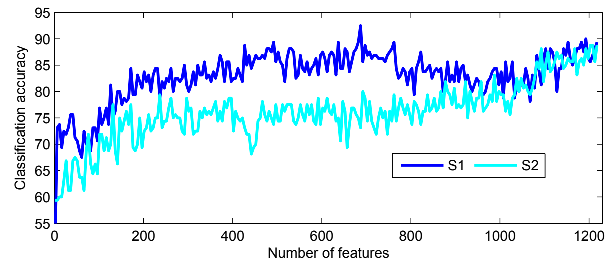
TABLE II. *D1* CLASSIFICATION ACCURACY [%]

S	He[10]	Zhang[17]	MIDRAS	MIDRA
1	67.7±02.2	77.2±0.0	77±05.87	84± 8.8
2	70.7±01.2	70.8±0.0	87.5±08.6	95±5.8
3	83.9±01.3	-	63±08.2	78.5±11.8
4	93.0±01.2	-	58±12.3	73.5±5.8
5	93.2±1.20	-	75.5±12.1	78.5±4.7
6	-	76.8±0.0	94.5±5.9	89.5±08.3
7	-	80.0±0.0	84.5±06.43	90.0±5.5
Mean	81.7±12.1	76.2±03.9	76.5±12.1	85.8±8.0

(–) Not provided



(a) Feature relevance values



(b) Learning curve

Fig. 4. *D2* (Emotiv data) MIDRA results. S1 performance: MIDRAS 86.2 ± 7.7; MIDRA 92.5 ± 4.4. S2 performance: MIDRAS 90.0 ± 7; MIDRA 90 ± 6.

V. CONCLUSIONS

We presented a MI discrimination methodology, termed MIDRA, to support the analysis of EEG data in BCI systems. For such a purpose, a relevance analysis strategy was introduced by using a variability based criterion to rank the contribution of EEG features for classifying a MI paradigm, i.e., imagination of left-right hand movements. In addition, to model MI paradigm from non-stationary EEG time-series, short-time parameters were employed, namely: spectral, Hjorth, and time-frequency based approaches were studied. Moreover, a k nearest-neighbor classifier was trained by taking advantage of the estimated feature relevance rank, and the BCI-system was validated by a 10-fold cross validation methodology. Experimental results carried out on two different BCI databases, a well-known public MI data and an Emotiv-based dataset built by us (following a standard MI protocol), showed that the precision of the BCI system increases significantly when using the proposed approach (MIDRA) in comparison with benchmark methods. In addition, achieved results also showed that the calculated features using the spectral and the time-frequency based representations allow providing a relevant feature set. Hence, applying MIDRA in a BCI system based on the Emotiv EPOC device is a straightforward alternative for dealing with MI paradigms. As future work, it would be interesting to identify other MI movements and to introduce more elaborate decomposition methodologies in terms of description of non-stationary signals. Furthermore, hidden inter-channel relationships should be estimated to enhance extraction of MI patterns.

Acknowledgments. Under grants provided by a Ph.D. scholarship and the project "Sistemas de interacción cerebro - máquina (BMI) a partir de registros de actividad eléctrica cerebral" funded by Colciencias.

References

[1] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment," *Neural Syst. and Rehab. Eng., IEEE Trans. on*, vol. 15, no. 4, pp. 473–482, 2007.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[3] B. Z. Allison, E. W. Wolpaw, and J. R. Wolpaw, "Brain-computer interface systems: progress and prospects," *Exp. Rev. of Med. Dev.*, vol. 4, no. 4, pp. 463–474, 2007.

[4] R. Corralejo, R. Hornero, and D. Álvarez, "Feature selection using a genetic algorithm in a motor imagerybased brain computer interface," in *IEEE EMBC*, 2011.

[5] G. Rodríguez and P. J. García, "Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces," *Medical systems*, vol. 36, no. 1, pp. 51–63, 2012.

[6] O. Carrera-Leon, J. M. Ramirez, V. Alarcon-Aquino, M. Baker, D. D'Croz-Baron, and P. Gomez-Gil, "A motor imagery bci experiment using wavelet analysis and spatial patterns feature extraction," in *Engineering Applications (WEA), 2012 Workshop on*. IEEE, 2012, pp. 1–6.

[7] J. F. D. Saa and M. Çetin, "A latent discriminative model-based approach for classification of imaginary motor tasks from eeg data," *Journal of neural engineering*, vol. 9, no. 2, p. 026020, 2012.

[8] A. Alvarez-Meza, L. Velasquez-Martinez, and G. Castellanos-Dominguez, "Feature relevance analysis supporting automatic motor imagery discrimination in eeg based bci systems," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 7068–7071.

[9] H. Higashi and T. Tanaka, "Common spatio-time-frequency patterns for motor imagery-based brain machine interfaces," *Computational intelligence and neuroscience*, vol. 2013, 2013.

[10] Wei-He, P. Wei, L. Wang, and Y. Zou, "A novel emd-based common spatial pattern for motor imagery brain-computer interface," in *IEEE EMBC*, 2012.

[11] S. Darvishi, M. C. Ridding, D. Abbott, and M. Baumert, "Investigation of the trade-off between time window length, classifier update rate and classification accuracy for restorative brain-computer interfaces," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 1567–1570.

[12] J. Orozco, S. Murillo, A. Álvarez, J. Arias, E. Trejos, J. Vargas, and G. Castellanos, "Automatic selection of acoustic and non-linear dynamic features in voice," in *INTERSPEECH*, 2011.

[13] L. F. Velásquez-Martínez, A. M. Álvarez-Meza, and 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.

[14] A. Teixeira, A. M. Tomé, M. Boehm, C. Puntonet, and E. Lang., "How to apply nonlinear subspace techniques to univariate biomedical time series," *IEEE Trans. on Instrument. and Measur.*, vol. 58, no. 8, pp. 2433–2443, 2009.

[15] A. Figliola and E. Serrano, "Analysis of physiological time series using wavelet transforms," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 16, no. 3, pp. 74–79, 1997.

[16] R. Corralejo, R. Hornero, and D. Alvarez, "Feature selection using a genetic algorithm in a motor imagery-based brain computer interface," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 7703–7706.

[17] H. Zhang, C. Guan, K. K. Ang, and C. Wang, "Bci competition iv–data set i: learning discriminative patterns for self-paced eeg-based motor imagery detection," *Frontiers in neuroscience*, vol. 6, 2012.