# 33% Classification Accuracy Improvement in a Motor Imagery Brain Computer Interface

**E. Bou Assi[1], S. Rihana[2], M. Sawan[1]**

[1]Polystim Neurotech Lab, Institute of Biomedical Engineering, Polytechnique Montreal, QC, Canada; [2]Biomedical Engineering Department, Holy Spirit University of Kaslik, Jounieh, Lebanon

**Correspondence to:** E. Bou Assi, elie.bou-assi@polymtl.ca

## ABSTRACT

A right-hand motor imagery based brain-computer interface is proposed in this work. Such a system requires the identification of different brain states and their classification. Brain signals recorded by electroencephalography are naturally contaminated by various noises and interferences. Ocular artifact removal is performed by implementing an automatic method "Kmeans-ICA" which does not require a reference channel. This method starts by decomposing EEG signals into Independent Components; artefactual ones are then identified using Kmeans clustering, a non-supervised machine learning technique. After signal preprocessing, a Brain computer interface system is implemented; physiologically interpretable features extracting the wavelet-coherence, the wavelet-phase locking value and band power are computed and introduced into a statistical test to check for a significant difference between relaxed and motor imagery states. Features which pass the test are conserved and used for classification. Leave One Out Cross Validation is performed to evaluate the performance of the classifier. Two types of classifiers are compared: a Linear Discriminant Analysis and a Support Vector Machine. Using a Linear Discriminant Analysis, classification accuracy improved from 66% to 88.10% after ocular artifacts removal using Kmeans-ICA. The proposed methodology outperformed state of art feature extraction methods, namely, the mu rhythm band power.

## 1. INTRODUCTION

The first Brain-Computer Interface (BCI) based system was proposed by Vidal in 1973 hundred years after the discovery of brain electrical activity [1]. Motor imagery based BCI is of particular importance

since it is a purely cognitive method that does not require an external stimulus or excitation such as in P300 or Evoked Electroencephalography (EEG) potential based BCIs. Such a system translates human thoughts into commands bypassing the peripheral nervous system. In an attempt to control a BCI, the user may modulate brain activity patterns. These should be automatically identified by the BCI classification algorithm and translated into commands. Although much effort has been put towards a better performance of BCI systems, the ability to accurately identify different motor imagery states remains elusive [2]. EEG is considered as the gold standard to capture electrical brain currents [3] flowing during synaptic excitations of pyramidal neurons in the cerebral cortex [4]. In this work, non-invasive EEG is used for recording brain's electrical activity by placing electrodes along the scalp. Therefore, signals are contaminated by external and internal inferences and need to be processed before analysis and feature extraction. These noises could be ocular artifacts, Electrocardiogram (ECG), Electromyogram (EMG) and many others. Ocular artifacts, mainly eye blinking, mostly affect the performance of an EEG based BCI system since they overlap with the EEG frequency of interest [5]. While other artifacts such as EMG can be minimized by asking the subject to be relaxed during the acquisition, it is impossible for almost all people to control their eye blinking. Additionally, it has been recently shown that asking the subjects to avoid blinking can induce an additive cognitive task introducing a bias in certain experiments [6]. Two main strategies have been proposed to attenuate this overlapping artifact. The first consists of a visual identification and rejection of epochs contaminated by ocular artifacts. This method is not accurate since it is subjective (related to the observer) and is usually accompanied by a loss of data [7]. The second strategy relies on conventional Electrooculogram (EOG) correction methods such as regression [8], linear filtering [9], Principal Components Analysis (PCA) [10] and Independent Component Analysis (ICA). ICA has shown to be very effective for ocular artifacts filtering with results favorable to those obtained with PCA or regression based methods [6]. However, this blind source separation (BSS) based technique requires the identification of artefactual components. Many researches have been conducted in this field; the traditional approach is a manual identification of artefactual Independent Components (ICs) by interpreting activation time series, scalp topographies, and frequency distribution. Although this approach may show usable for offline EEG preprocessing and analysis, it is not suitable for online BCI applications [11]. Recent studies attempted to automate the identification of artefactual ICs by adding a reference channel such as an EOG to the acquisition montage [12]. Since EOG measures the corneo-retinal potential between the front and the back of the eye, the relations of each IC and this channel may allow the semi-automatic identification of ocular artifacts. However, placing electrodes on the face wouldn't be esthetically desirable in real-life BCI applications besides increasing the complexity of the acquisition scheme. In this work, "Kmeans-ICA", a fully automatic filtering method based on ICA was implemented to remove overlapping ocular artifacts in a BCI framework. This method extracts simple features from ICs' time series and does not require adding an EOG channel. The average correlation of each IC with prefrontal electrodes (most contaminated by EOG noise), the maximum value of each IC, and the ratio between the peak amplitude and the variance of each IC are used as inputs to a non-supervised clustering algorithm [5].

After signal preprocessing, a right-hand motor imagery based BCI is implemented. The BCI should be able to recognize and classify considered mental states. It starts by extracting the band power, the phase locking value (PLV), and the coherence based on a discrete wavelet transform (DWT) decomposition. These physiologically related features are then passed into a statistical test to choose the most relevant ones. The features displaying a statistically significant difference between relax and motor imagery states are fed into a linear discriminant analysis (LDA) and a support vector machine (SVM) for classification. Recognition rate and accuracy substantially improved after removing ocular interferences.

## 2. MATERIALS AND METHODS

As already mentioned, a right-hand motor imagery based BCI was implemented in this work. In what follows, the materials and methods used to implement this system are presented in a traditional signal processing framework. Figure 1 summarizes signal processing techniques adopted in this work, starting
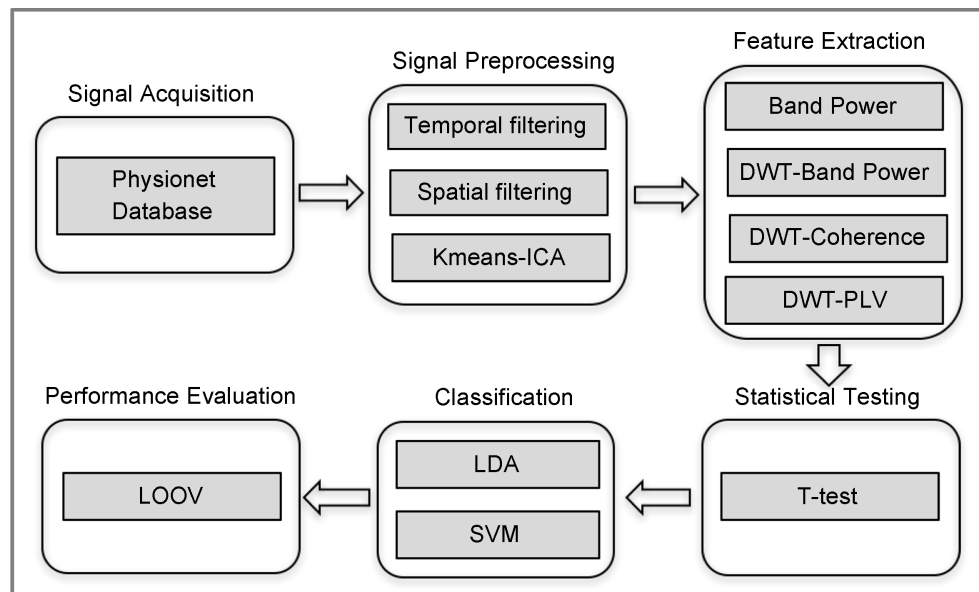
**Figure 1.** Framework of the proposed BCI system. Signal preprocessing and feature extraction are performed prior to statistical testing. Selected features are adopted for classification using a LDA and SVM based on a LOOV scheme. ICA: Independent Component Analysis; DWT: Discrete Wavelet Transorm; PLV: Phase Locking Value; T-test: Student test; LDA: Linear Discriminant Analysis; SVM: Support Vector Machine; LOOV: Leave One Out Corss Validation.

from signal acquisition until classification. Interestingly, two challenges faced in the BCI research are conducted here: the first consists of adequate preprocessing of EEG signals especially those contaminated with ocular artifacts. The second tackles the identification of the right-hand motor imagery state. Therefore, we will start by presenting the pre-processing techniques while focusing on the proposed Kmeans-ICA ocular artifact rejection method. Then feature extraction, classification and performance evaluation methods are presented and discussed.

## 2.1. Signal Acquisition

An international database, "PhysioNet.org", is used in this study. It contains several datasets covering different physiological effects. The "EEG Motor Movement/Imagery dataset" is selected. This database was recorded using the international BCI2000 instrumentation system [13]. It contains 64 channels scalp EEG recordings sampled at 160 Hz. The total database includes 109 subjects who performed 14 motor imagery trials. Electrodes were positioned as per the 10-10 international system, a standard of the American clinical neurophysiology society and the international federation of clinical neurophysiology [14]. In this work, the right-hand motor imagery and relax signals recorded while eyes open are adopted. For more details about the data acquisition experiment, readers are referred to [13].

## 2.2. Signal Preprocessing

### 2.2.1. Filtering

EEG recordings were first band pass filtered [1 - 30 Hz] using a finite impulse response filter (FIR). This ensures including the motor imagery frequency spectrum, mainly the beta rhythm (13 - 30 Hz) and the mu rhythm (7 - 12 Hz). The main advantage of FIR filters is that they are stable and that they generally

yield a linear phase response. Moreover, to remove this linear phase response and achieve a zero-phase filtering, the forward backward technique was adopted by applying the FIR filter to the signal, inverting it in time, reapply the FIR filter and invert the signal back. In addition, a spatial filtering was used to increase the Signal to Noise Ratio (SNR) what may lead to a better classification of the motor imagery states. Different types of spatial filtering were proposed in the literature but several studies have shown that Common Average Reference (CAR) is the best suited for BCI applications [15]. The spatial filter should be designed in a way to maximally accentuate the control signal while attenuating non-EEG artifacts. This can be achieved by subtracting the common electrical brain activity to a defined point of interest (1).

$$\hat{V}_i = V_i - \frac{1}{N} \sum_{i \in \omega_i} V_j \tag{1}$$

where $V_i$ and $\hat{V}_i$ are electrode potentials respectively prior and after filtering respectively, and $\omega_i$ is a group of the $N$ surrounding channels.

### 2.2.2. Ocular Artifacts Rejection

Our group has recently proposed a non-supervised ocular artifact rejection method combining ICA and Kmeans clustering [7]. In this work, we validate the proposed methodology and use it within the context of a motor imagery classification enforced by statistical assessment and adequate performance evaluation. Due to the overlapping nature of ocular artifacts, advanced signal processing techniques based on BSS have been introduced [16]. A BSS method is based on having $N$ unknown sources $S$, an unknown operator $A$, and $N$ observed signals $X$ (2).

$$X = AS \tag{2}$$

ICA was proposed within the framework of BSS techniques. The main idea consists of finding the weights matrix $A$ that will separate the original multichannel signal into non-Gaussian, statistically independent components. While several ICA algorithms exist, the Extended-Infomax algorithm has been used in this work. It had been previously demonstrated that Extended-Infomax outperformed other ICA algorithms, especially for signals containing sup and sub-Gaussian components simultaneously [16]. Some ICs are typically ocular related electrical activity (artifacts) and should be removed. This is traditionally done manually by studying activations' frequency content and scalp topographies [11]. Our main concern is to automatically find artefactual ICs and omit them from the signal's reconstructed part.

### 2.2.3. Automatic Artifact Identification—Feature Extraction

To automate the process of Ocular Artifact identification, three features were adopted and introduced into a Kmeans clustering algorithm [7]. The first feature consists of computing the average of cross correlation between each IC with pre-frontal electrodes (FP1 and FP2) as shown in (3). Our motivation behind using this feature is that frontal electrodes are the most contaminated by EOG noise [4].

$$\text{Avrg}(i) = \left( R_{\text{fp}_1, \text{IC}(i)} + R_{\text{fp2}, \text{IC}(i)} \right) \Big/ 2 \tag{3}$$

The cross correlation is implemented as a measure of similarity between two-time series. $R_{ab}$ is the cross correlation between the two signals $a$ and $b$ (4). The closer the absolute value is to 1 the more the signals are correlated.

$$R_{ab} = \sum_{n=-\infty}^{+\infty} a[n] * b[n+k] \tag{4}$$

where $k$ is the time delay between $a$ and $b$.

The second feature is the distribution ratio; it consists of dividing the maximum amplitude of an IC by its standard deviation (5). Adopting such a characteristic is based on the fact that non-artefactual EEG activity is tightly distributed around its mean [4].

$$\text{Dist}_{\text{Ratio}} = \frac{\max \left( IC(i) \right)}{\delta^2 \left( IC(i) \right)} \tag{5}$$

The third one is the maximum value of each IC. Our assumption behind proposing it relies on the fact that ocular artifacts are higher in amplitude than normal neurological signals and their peak are commonly centered [5]. This resulted in three feature vectors that will be used as inputs to a Kmeans clustering algorithm to identify artefactual ICs.

### 2.2.4. Kmeans Clustering—Classifying ICs

A Kmeans clustering algorithm was used in this study to classify ICs into artefactual and physiological components [7]. The Kmeans clustering algorithm performs an iterative minimization of the sum of squared distances from all elements in a cluster to the cluster center [17]. The mathematical principle of this algorithm, applied to our framework is briefly explained. Let's consider a set of ICs (6).

$$\text{ICs}\{\text{IC}(1), \text{IC}(2), \text{IC}(3), \cdots, \text{IC}(l)\} \tag{6}$$

The algorithm starts by dividing this set into 2 clusters (7) (8):

$$\text{IC}1(1), \text{IC}1(2), \text{IC}1(3), \cdots, \text{IC}1(l_1) \tag{7}$$

$$\text{IC}2(1), \text{IC}2(2), \text{IC}2(3), \cdots, \text{IC}2(l_2) \tag{8}$$

Then the local means of each group are computed and the new grouping is defined as (9) (10)

$$\left|\text{IC}1(k) - \text{C}1\right| < \left|\text{IC}1(k) - \text{C}2\right| \ \text{ for } k = 1, \cdots, l_1; \tag{9}$$

$$\left|\text{IC}2(k) - \text{C}2\right| < \left|\text{IC}2(k) - \text{C}1\right| \ \text{ for } k = 1, \cdots, l_2; \tag{10}$$

where Ci is the center of cluster i.

In such a way, ICs in cluster 1 are closer to C1 and ICs in cluster 2 are closer to C2.

### 2.2.5. Evaluation of the Proposed Method

Since the proposed approach consists of a classification of the ICs, we started by assessing the accuracy of the proposed method in identifying artefactual components. Traditional performance descriptors were used: accuracy, sensitivity, specificity, and precision. A true detection of an eye blinking artifact is considered as a true positive event. Accuracy refers to the percentage of correct predictions. Specificity determines the percentage of correctly negative labeled samples classified as negative, while sensitivity indicates the percentage of positive labeled samples classified as positive and precision refers to the positive predictions that are correctly classified. Therefore, the sensitivity of the test represents its ability to detect the presence of the event, in this part of the paper, the eye blinking artifact, while the specificity quantifies its performance in identifying the absence of the eye blinking artifact. Performance descriptors were computed over 640 samples (a total of 6,400,000 data points).

### 2.3. Motor Imagery Based BCI: Feature Extraction

After signal preprocessing, a right-hand motor imagery based BCI was implemented. It starts by extracting features able to show relevant information about different brain states. Each of these features is passed through a statistical test to verify it represents a statistically significant difference between motor imagery and relax states. Four features were used in this work. First, the band power of the mu rhythm, a state of art feature in motor imagery classification was computed and used as an input to the classifier. The three remaining features combine the Discrete Wavelet Transform (DWT) with respectively the Band Power, the PLV, and the coherence. These features were chosen since they provide information about properties (power and frequency cross correlation) and dynamics (phase locking value) of brain mental states and thus will contribute to the design of an interpretable BCI.

### 2.3.1. Band Power Feature

Mu rhythm band power is calculated over electrode C3 in this project since the main goal is to track

the event related desynchronization in the contralateral brain region (11).

$$X(t) = \left\langle x^2(t) \right\rangle_t \tag{11}$$

Motor imagery (planning of a movement) and movement are respectively characterized by the occurrence of Event Related Desynchronization (ERD) and Event Related Synchronization (ERS). ERD occurs contralaterally in the mu and beta rhythms, while ERS occurs in the beta rhythm, at the ipsilateral side [4]. The Power Spectral Density (PSD) shows a decrease during an ERD and an increase during an ERS [18].

### 2.3.2. Discrete Wavelet Transform

The wavelet transform is a multi-resolution decomposition allowing signal analysis at multiple scales and resolutions simultaneously so that low frequencies could be examined with high frequency resolution and high frequencies could be studied considering a high temporal resolution (12).

$$W_x(s,u) = \int_{-\infty}^{+\infty} X(t) \varnothing_{u,s}(t) \, dt \tag{12}$$

Each wavelet is a translated and scaled version of the mother wavelet (13).

$$\varnothing_{a,b}(t) = \frac{1}{\sqrt{(a)}} \varnothing\left(\frac{t-b}{a}\right) \tag{13}$$

In this study, the DWT was implemented in a discrete way in terms of a series of discrete low- and high-pass filters, decomposing the signal into approximations and details. This decomposition is well described by the Mallat algorithm as shown in **Figure 2**.

Equation (14) depicts the frequency bounds of detail levels obtained by wavelet transforming an EEG signal sampled at $F_s$.

$$\left[ 0, \frac{F_s}{2^{L+1}} \right], \left[ \frac{F_s}{2^{L+1}}, \frac{F_s}{2^L} \right], \cdots, \left[ \frac{F_s}{2^2}, \frac{F_s}{2} \right] \tag{14}$$

The EEG signal has been decomposed by DWT using Debauchies 5 mother wavelet due to its extreme phase and high number of vanishing moments [19]. Since the Physionet database was recorded and sampled
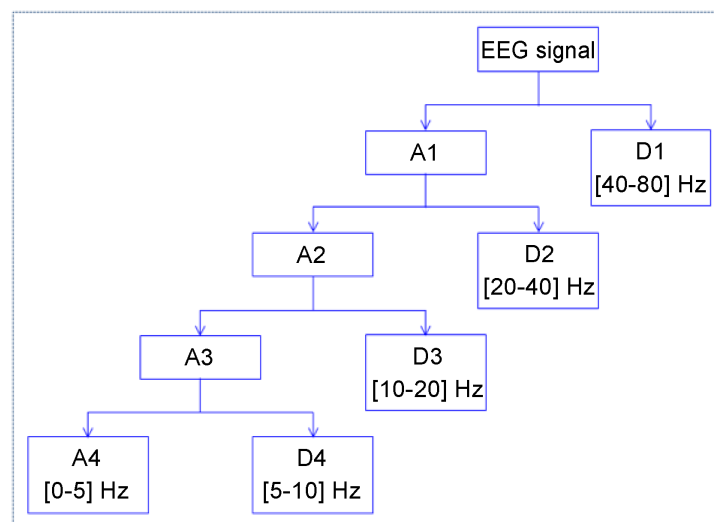


**Figure 2.** Wavelet decomposition of the EEG signal. $A_1$, $A_2$, $A_3$, $A_4$ are approximation levels of the signal while $D_1$, $D_2$, $D_3$, and $D_4$ are details coefficients.

at 160 Hz, a four levels' wavelet transform was performed to depict the Mu rhythm. Three features have been extracted from the D4 mu rhythm level. These features are the DWT-Band Power, DWT-PLV, and DWT-Coherence.

### 2.3.3. DWT-Coherence

The DWT-Coherence feature starts by a DWT decomposition and then calculate the coherence between electrodes C3 and C4 during right hand motor imagery and relax states (15). Physiologically, due to the occurrence of an event related desynchronization (ERD) on electrode C3 but not on C4 (contralateral electrode) during the right-hand motor task, the coherence is expected to decrease.

$$C_{xy} = \frac{0.5 \sum_{b_1}^{b_2} \overline{x}^* \overline{y} + \overline{x}\,\overline{y}^*}{\sqrt{\sum_{b_1}^{b_2} \overline{x}\,\overline{x}^* \cdot \sum_{b_1}^{b_2} \overline{y}\,\overline{y}^*}} \tag{15}$$

where $x$ and $y$ are the Fourier transformed temporal signals and $b_1$, $b_2$ the frequency limits of interest. Figure 3 shows 2D representation of the locations of the electrodes involved in computing this feature.

### 2.3.4. DWT-Phase Locking Value

The phase locking value is used to quantify coupling and phase synchrony between time series. It has been introduced as a feature for BCI applications [20]. Considering two signals $x(t)$ and $y(t)$ and their corresponding phase's $\varnothing_x(t)$ and $\varnothing_y(t)$, their phases' difference usually fluctuates around a constant value as shown in (16).

$$\varnothing_x(t) - \varnothing_y(t) = \text{constant value} \tag{16}$$

Two steps are performed to estimate the phase synchrony; the first consists of estimating the instantaneous phase of each EEG signal. This is usually computed by means of the analytic signal using the Hilbert Transform. For each signal $x(t)$, the analytic signal $z(t)$ is a complex function defined as follows (17):
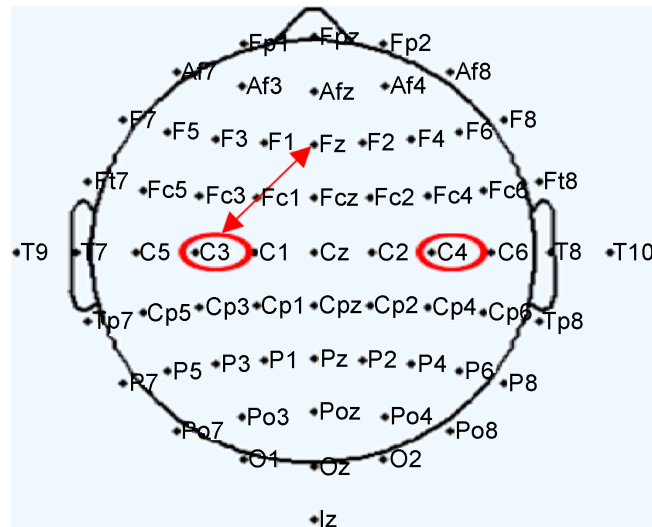
$$z(t) = x(t) + j\hat{x}(t) \tag{17}$$



**Figure 3.** Position of the electrodes involved in computing the DWT-coherence feature (circles) and the DWT-PLV feature (arrow). Electrodes are placed according to the 10-10 international positioning standard.

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} \mathrm{d}\tau \qquad (18)$$

where $\hat{x}(t)$ is the Hilbert transform of $x(t)$. The phase of the signal $x(t)$ is deduced as in (19):

$$\varnothing(t) = \arctan\left(\frac{\hat{x}(t)}{x(t)}\right) \qquad (19)$$

The second step is to evaluate the amount of phase locking by integrating the phase difference over time for each trial (20).

$$\mathrm{PLV} = \left| \left\langle e^{j\Delta\varnothing(t)} \right\rangle \right| \qquad (20)$$

If the signals are completely synchronized, $\Delta\varnothing(t)$ is a constant resulting in PLV = 1; otherwise, $\Delta\varnothing(t)$ follows a uniform distribution and the PLV tends to zero. Local and large scale synchrony are the most implemented in the literature of brain human history [21]. Local scale is computed between neighboring electrodes while large-scale involves relatively distant electrodes.

In this paper, large scale phase locking was investigated between electrodes Fz and C3. The ERD on electrode C3 may trigger a decrease in synchrony between C3 and Fz; the latter one not influenced by the ERD. Figure 3 shows a 2D mapping of the electrodes involved in computing the DWT-PLV features.

## 2.4. Statistical Test

A statistical t-test was used to verify which features are significantly different between relax and motor imagery states. The features for which a statistically significant difference is found between relax and motor imagery states will be used as inputs to the classifier. The t-test may be used under normality assumption and has been commonly used in BCI studies [22]. It has shown to be one of the most conservative tests used for detecting significant differences using a statistic based on the means of the two sets (21).

$$T = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \qquad (21)$$

A significance level of 0.05 is used in this study.

## 2.5. Classification

After quantitative features have been extracted, each signal might be represented by a feature vector and used as input to the classifier. In this study, a SVM and a LDA are compared.

### 2.5.1. Linear Discriminant Analysis

LDA is one of the most effective linear classification methods successfully implemented in BCI applications [23]. A LDA classifier is mainly based on a statistical procedure that defines a criterion function $J(w)$ that should be maximized. This function requires projecting the n-dimensional data into a line $y = w^T x$.

The first step is to compute the sample and projected mean in each class as shown respectively in (22) and (23):

$$\text{Sample } m_i = \left(\frac{1}{k_i}\right) \sum_{x \in X_i} x \qquad (22)$$

$$\text{Proj.} m_i = -\left(\frac{1}{k_i}\right) \sum_{y \in Y_i} y = \left(\frac{1}{k_i}\right) \sum_{x \in X_i} w^T x = w^T m_i \qquad (23)$$

where $k_i$ is the number of samples in class $i$ and $X_i$ is the set of vectors in the class $i$. A 2-class problem will be expressed as in (24):

$$\left| m_1 - m_2 \right| = w^{\mathrm{T}} \left| \left( m_1 - m_2 \right) \right| \tag{24}$$

Considering a class $i$, the scatter for projected samples is defined as:

$$s_i^2 = \sum\nolimits_{y \in Y_i} \left( y - m_i \right)^2 \tag{25}$$

The Linear discriminant, known as the Fisher Discriminant is the function $w^{\mathrm{T}} x$ for which $J(w)$ is at its highest possible values (26).

$$J(w) = \frac{\left| m_1 - m_2 \right|^2}{\left( s_1^2 + s_2^2 \right)} \tag{26}$$

However, in this expression, $J$ is not as function of $w$. Let's define the scatter matrices $S_i$ as shown in (27):

$$S_i = \sum\nolimits_{x \in X_i} \left( x - m_i \right) \left( x - m_i \right)^{\mathrm{T}} \tag{27}$$

The within class scatter $S_w$ and the between class scatter $S_B$ will be computed as shown respectively in (28) and (29).

$$S_w = S_1 + S_2 \tag{28}$$

$$S_B = \left( m_1 - m_2 \right) \left( m_1 - m_2 \right)^{\mathrm{T}} \tag{29}$$

Thus, $s_i$ is defined as in (30)

$$s_i = \sum\nolimits_{x \in X_i} \left( w^{\mathrm{T}} x - w^{\mathrm{T}} m_i \right)^2 = w^{\mathrm{T}} S_i w \tag{30}$$

Therefore, the criterion $J(w)$ is expressed a in (31)

$$J(w) = \frac{w^{\mathrm{T}} S_B w}{w^{\mathrm{T}} S_w w} \tag{31}$$

Using a non-parametric classifier (like LDA) gives a stable basis of comparison between several features and avoids any interference such as those provoked from the SVM training.

### 2.5.2. Support Vector Machine

Support Vector Machine (SVM) is one of the most recently adopted classifiers able to create non-linear decision boundaries [20]. It is mainly a binary classifier but it can be extrapolated to solve multi-class problems. Given a set of features with associated labels $\left\{ y_i : y_i \in \{-1,1\} \right\}$ then the training algorithm attempts to find a hyper plane allowing a maximum separation between support vectors corresponding to $y_i = 1$ and $y_i = -1$.

The simplest hyper plane is the linear one, for example, separating two classes (32):

$$\left( w \cdot x \right) + b = 0 \tag{32}$$

The optimal hyper plane is the one that maximizes the distance between the nearest training points. Figure 4 shows the classification hyper plane of a linear SVM.

The minimum distance between H1 and H2 corresponds to $2/\left| w \right|$.

Therefore, the optimization problem could be resumed as follow: Minimize $\left| w \right| 2$ subject to constraints:

$$y_i \left( w \cdot x_i + b \right) - 1 \geq 0 \tag{33}$$

Using the kernel trick, the SVM can project data into another dimensional space performing a
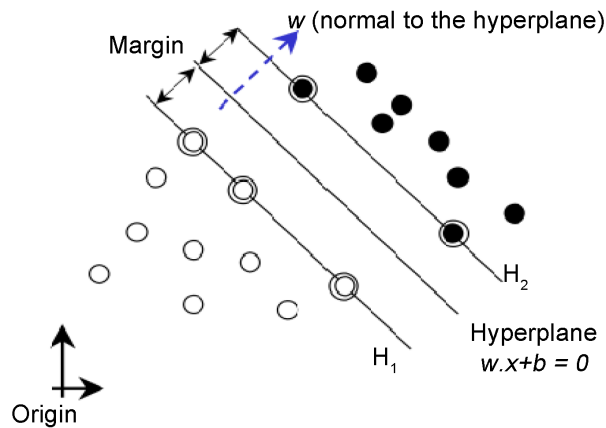
**Figure 4.** Linear kernel support vector machine.

non-linear separation.

Several kernels have been proposed such as polynomial (34), Radial Basis Function (RBF) (35) and Multi-Layer Perceptron (MLP) (36).

$$K(x, y) = (x \cdot y + 1)^p \qquad (34)$$

$$K(x, y) = \exp\left(\frac{-|x - y|^2}{2\sigma^2}\right) \qquad (35)$$

$$K(x, y) = \tanh\left(p_1 \langle x \cdot y \rangle + p_2\right) \qquad (36)$$

The SVM classifier has given good results in BCI research activities [24], however they have several hyper parameters that should be fixed such as the regularization parameter as well as the RBF width when using a non-linear SVM and they show a relatively low speed of execution what would be a constraint in an online BCI system.

## 3. RESULTS AND DISCUSSION

Results can be divided into two main parts: the first describes the performance of the ocular artifact rejection method; a case is considered for visualization purposes and quantitative evaluation methods are presented. The second covers motor imagery classification performance evaluation.

### 3.1. Ocular Artifacts Rejection Method

The ocular artifact rejection method was done using Kmeans-ICA. ICA of the 64 channels input signals was performed using the Infomax algorithm. **Figure 5** shows scalp topographies of the 64 ICs. By means of visual inspection, in the considered dataset, components 1, 3, and 6 are considered as eye blinking artifacts since their scalp maps show an intense frontal projection, typical of an ocular artifact. As already mentioned, an automatic method has been implemented in this paper for detecting these artefactual components and omitting them from the reconstructed signals.

Thus, the correlation, distribution ratio and maximum value features have been considered. Experimental results show that components 1, 3, and 6 have the highest value for each of the proposed features.

As mentioned above, IC1, IC3 and IC6 are assumed as ocular artifacts. **Figure 6** shows their scalp map projections, typical of an eye artifact. Artefactual components are also identified by a field expert based on ICs' scalp maps, power spectrum and time series. This serves as our ground truth for performance evaluation. Accuracy, precision, sensitivity, and specificity were calculated for all feature vectors.

**Table 1** shows the values of true positive (TP), true negative (TN), false positive (FP) and false
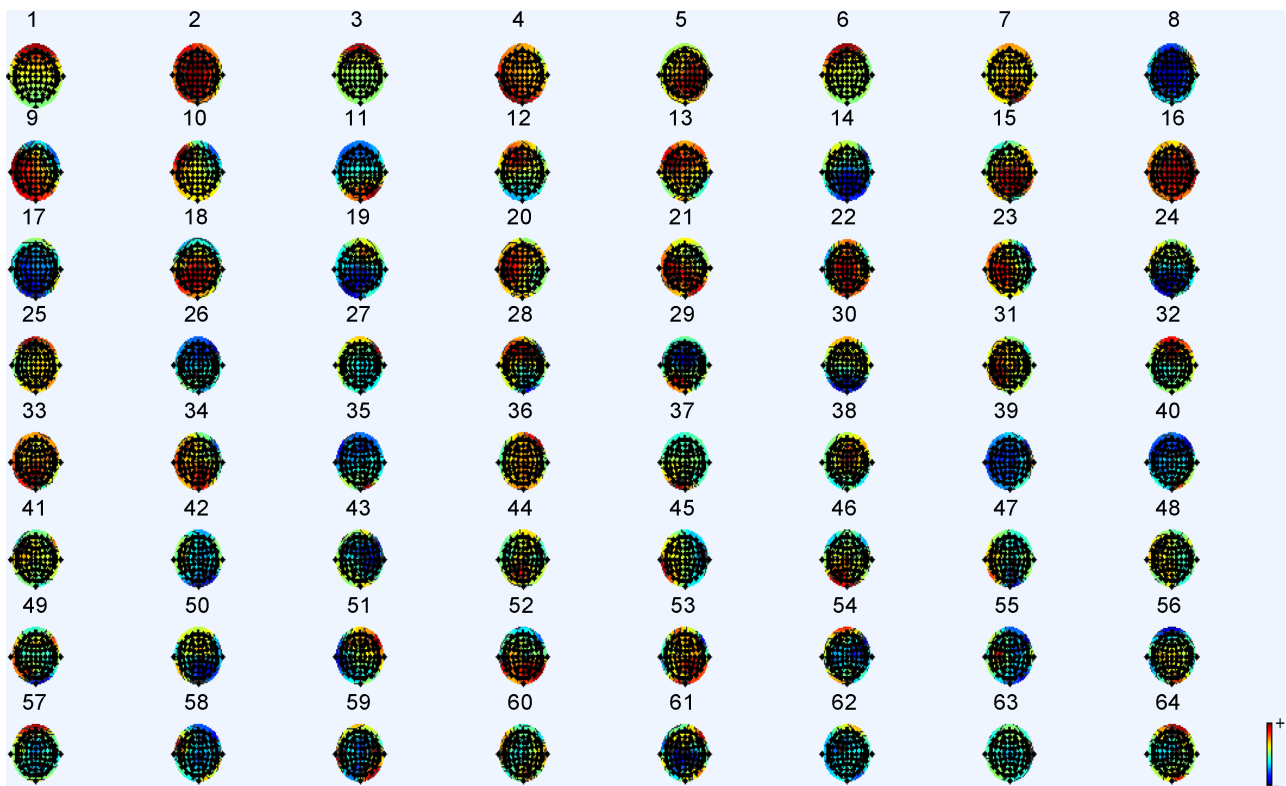
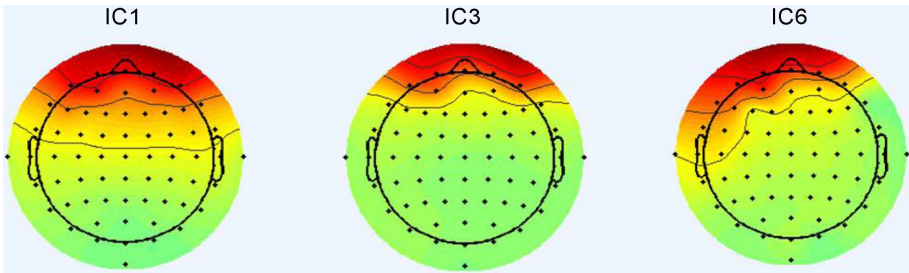**Figure 5.** Scalp topographies of the independent components.



**Figure 6.** Frontal projection typical of an ocular artifact in 2D mapping of IC1, IC2, and IC3.

**Table 1.** Values of TPs, TNs, FPs, and FNs for all features.

| Feature | Performance Measure | | | | |
|---|---|---|---|---|---|
| | TP | TN | FP | FN | Total |
| Correlation | 25 | 612 | 3 | 0 | 640 |
| Distribution Ratio | 18 | 516 | 12 | 104 | 640 |
| Maximum Value | 14 | 618 | 8 | 0 | 640 |

TP: True positive; TN: True negative; FP: False positive; FN: False negative.

negative (FN) events as calculated respectively for the classification based on the correlation, distribution ratio, and maximum value features over 640 samples from different subjects across different trials.

Table 2 shows the performance of the ocular artifact non-supervised classifier in terms of accuracy, sensitivity, specificity, and precision based on 640 samples from all subjects. Three features were used for classification; when used independently, the correlation feature gives the highest accuracy (99.54%) and will be considered for the identification of artefactual ICs in the next parts of this paper.

Prominent results were also obtained for other features especially for the maximum value, an easy to compute, simple feature. These prominent results highlight the feasibility of automatic ocular artifacts removal from EEG signals based on ICA.

### 3.2. Motor Imagery Classification

### 3.2.1. Mapping of the Band Power Feature

The four physiologically interpretable features adopted were the Band power, DWT-Band power, DWT-PLV and DWT-Coherence. A 2D mapping of the band power feature over the sensorimotor area is shown during the relax state (Figure 7(a)) and right hand motor imagery state (Figure 7(b)). The mu-rhythm band power shows a contralateral decrease (left) during the right-hand motor imagery state.

### 3.2.2. Results of the Statistical Test

In order to quantify the discriminative power of these features, significance difference has been tested between relax and motor imagery states using the t-test, a statistical significance test applied to small data sets. Table 3 shows the results of the t-test at a significance level of 0.05. The null hypothesis was rejected for Band Power, DWT-Coherence, and DWT-PLV suggesting that there is a significant difference between the two states (Table 3). Thus, these features will be use as inputs to the classifier.

**Table 2.** Results of the ROC analysis.

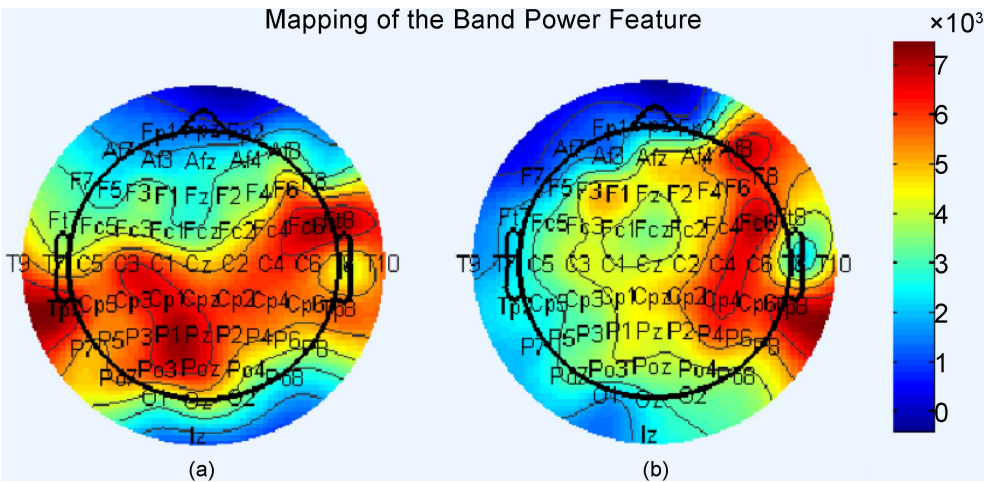| Feature | Performance Measure | | | |
|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision |
| Correlation | 99.54% | 100% | 99.51% | 89.28% |
| Distribution Ratio | 81.87% | 14.7% | 97.68% | 60% |
| Maximum Value | 98.75% | 100% | 98.72% | 64% |



**Figure 7.** 2D Mapping of the band power feature during relax (a) and right hand motor imagery (b) states. A decrease in the mu rhythm's band power is observed contralaterally (left) due to the ERD.

### 3.2.3. Classification Using a LDA

Leave-One-Out cross validation (LOOV) was performed for both classification schemes on Matlab©. This type of validation is the most used on EEG signals where the amount of data for each subject is limited. The population consists of three subjects (50 signals/trials). Table 4 shows the results of the cross validation using a LDA after ocular artifact removal based on Kmeans-ICA. Results show that the best performance is obtained after ocular artifact removal using the three features together (Table 4) achieving a relatively good performance considering non-physiological interferences (e.g. extreme values and baseline drifts) common in the Physionet database. These artifacts would highly affect the recognition rate of the classifier. Additionally, to evaluate the suitability of the "Kmeans-ICA" ocular rejection method in the context of a motor imagery classification, the LOOV based on a LDA was performed on the ocular artifact contaminated EEG signals.

Table 5 illustrates the results of LOOV without ocular artifact removal. Results show that the classification accuracy based on the three features increased from 66% to 88.10% after removing ocular artifacts by means of "Kmeans-ICA".

### 3.2.4. Classification Using a SVM

Classification based on a support vector machine was also implemented. LOOV was performed on the same signals after Kmeans-ICA filtering while using different kernel functions. A sequential minimal

**Table 3.** Results of the statistical test.

| Statistical Measure | Feature | | | |
|---|---|---|---|---|
| | Band Power | DWT-Band Power | DWT-Coherence | DWT-PLV |
| H | 1 | 0 | 1 | 1 |
| P | $6.1795e^{-07}$ | 0.6434 | 0.0068 | $4.1535e^{-04}$ |
| Result | Null Hypothesis Rejected | Null Hypothesis Conserved | Null Hypothesis Rejected | Null Hypothesis Rejected |

**Table 4.** Results of the LOOV after Kmeans-ICA using a LDA.

| Feature | Performance Measure | | | |
|---|---|---|---|---|
| | Accuracy | Specificity | Sensitivity | Precision |
| Band Power | 74% | 77.7% | 70.83% | 80.95% |
| DWT_Coherence | 71.42% | 70.58% | 65.4% | 76.2% |
| DWT_PLV | 72.72% | 65% | 64% | 66.66% |
| All Features | 88.10% | 90% | 86.36% | 90.47% |

**Table 5.** Results of the LOOV without ocular artifact removal.

| Feature | Performance Measure | | | |
|---|---|---|---|---|
| | Accuracy | Specificity | Sensitivity | Precision |
| Band Power | 66.66% | 70.58% | 64% | 76.19% |
| DWT_Coherence | 57.14% | 58.82% | 59.25% | 66.6% |
| DWT_PLV | 42.85% | 42.85% | 42.85% | 42.85% |
| All Features | 66% | 60% | 71% | 62% |

optimization algorithm was used to find the separating hyper plane using the following kernel functions: linear, polynomial, RBF and MLP. The highest accuracy was obtained using a linear kernel function based on a combination of all features (85.71%); however, using an LDA classifier gave a better accuracy (88.10%). RBF, Polynomial, and MLP kernel gave accuracies of 80.95%, 76.20%, and 73.80% respectively using the combination of all features. As mentioned, an LDA is easier to implement, simpler and less time consuming what make it suitable for online BCI applications. Worth to mention, the accuracies were obtained from signals recorded across many sessions and many subjects and are therefore subject to inter-session and inter-subject variability. Usually, in a BCI system, there is a phase of calibration, in which the classifier is well adapted to each subject. This phase will probably increase the accuracy since each system will be adapted to its user's brain states and this is an ongoing research.

Using a combination of three interpretable features as inputs to the LDA classifier, the BCI system reached a classification accuracy of 88.10% while differentiating between relax and right hand motor imagery states. Considering the Physionet signal population and for comparison, a classical BCI system based solely on the traditional, most widely used band power feature gave an accuracy of 74% what highlights the performance of the proposed system.

## 4. CONCLUSIONS

A BCI-based system was proposed for the right-hand motor imagery classification. It uses an automatic ocular artifact rejection technique that does not require adding an EOG reference channel. Its principle mainly relies on ICA to which simple and easy to implement features were added to automate the algorithm. Ocular artifacts are identified by means of a non-supervised machine learning technique: Kmeans clustering. Results show that the proposed method is effective in terms of ocular artifact identification, removal and conservation of neurological signals. This is supported by the increased accuracy of the motor imagery classification after Kmeans-ICA.

Additional preprocessing techniques were also used such as temporal and spatial filtering in order to remove other types of artifacts. Three physiologically interpretable features: band power, DWT-PLV and DWT-Coherence were used for the classification. Results show that the use of Kmeans-ICA has increased the accuracy of the classification that reached 88.10% when using a linear discriminant analysis classifier considering the inter-subject and even inter-session variability (compared to 66% without Kmeans-ICA filtering). This is due to the fact that ocular artifact removal increases the discriminant stability of the EEG data when eliminating the high influence of the EOG noise and maintaining the EEG of interest what will increase the recognition rate of the BCI system.

Two classifiers were compared: 1) a LDA and 2) a SVM with regards to BCI applications. The use of the Kmeans-ICA method could be extended to more applications other than BCI ones such as general EEG denoising. ICA is being more extensively used as a preprocessing tool in software that provides solutions for neurophysiological research; however current solutions are limited to a manual classification of the ICs. The implementation of a module based on the proposed method would be helpful and reduce the variability due to different interpretation of scalp topographies.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Vidal, J.J. (1973) Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*, **2**, 157-180. https://doi.org/10.1146/annurev.bb.02.060173.001105

2. Baali, H., Khorshidtalab, A., Mesbah, M. and Salami, M.J.E. (2015) A Transform-Based Feature Extraction Approach for Motor Imagery Tasks Classification. *IEEE Journal of Translational Engineering in Health and Medi-*

_cine_, **3**, 1-8. https://doi.org/10.1109/JTEHM.2015.2485261

3. Delgado Saa, J.F. and Cetin, M. (2013) Discriminative Methods for Classification of Asynchronous Imaginary Motor Tasks from EEG Data. _IEEE Transactions on Neural Systems and Rehabilitation Engineering_, **21**, 716-724. https://doi.org/10.1109/TNSRE.2013.2268194

4. Sanei, S. and Chambers, J.A. (2007) EEG Signal Processing. John Wiley & Sons Ltd., 35-125. https://doi.org/10.1002/9780470511923

5. Assi, E.B. and Rihana, S. (2014) Kmeans-ICA Based Automatic Method for EOG Denoising in Multi-Channel EEG Recordings. _Proceedings of 11th IASTED International Conference on Biomedical Engineering_, Zurich, 23-25 June 2014, 233-238.

6. Kroupi, E., Yazdani, A., Vesin, J.M. and Ebrahimi, T. (2011) Ocular Artifact Removal from EEG: A Comparison of Subspace Projection and Adaptive Filtering Methods. 19th _European Signal Processing Conference,_ Barcelona, 2011, 1395-1399.

7. Assi, E.B., Rihana, S. and Sawan, M. (2014) Kmeans-ICA Based Automatic Method for Ocular Artifacts Removal in a Motorimagery Classification. _Engineering in Medicine and Biology Society (EMBC)_, 2014 36th _Annual International Conference of the IEEE_, Chicago, IL, USA, 26-30 August 2014, 6655-6658. https://doi.org/10.1109/EMBC.2014.6945154

8. Croft, R.J. and Barry, R.J. (2000) Removal of Ocular Artifact from the EEG: A Review. _Clinical Neurophysiology,_ **30**, 5-19. https://doi.org/10.1016/S0987-7053(00)00055-1

9. Gotman J., Skuce, D.R., Thompson, C.J., Gloor, P., Ives, J.R. and Ray, W.F. (1973) Clinical Applications of Spectral Analysis and Extraction of Features from Electroencephalograms with Slow Waves in Adult Patients. _Electroencephalography and Clinical Neurophysiology_, **35**, 225-235.

10. Lagerlund, T.D., Sharbrough F.W and Busacker, N.E. (1997) Spatial Filtering of Multichannel Electroencephalographic Recordings through Principal Component Analysis by Singular Value Decomposition. _Journal of Clinical Neurophysiology_, **14**, 73-82. https://doi.org/10.1097/00004691-199701000-00007

11. Romero, S., Mananas, M.A., Riba, J., Morte, A., Gimenez, S., Clos, S. and Barbanoj, M.J. (2004) Evaluation of an Automatic Ocular Filtering Method for Awake Spontaneous EEG Signals Based on Independent Component Analysis. 26_th Annual International Conference of the IEEE Engineering in Medicine and Biology Society_, San Francisco, CA, 1-5 September 2004, 925-928. https://doi.org/10.1109/iembs.2004.1403311

12. Joyce, C.A., Gorodnitsky, M., I.F. and Kutas, M. (2004) Automatic Removal of Eye Movement and Blink Artifacts from EEG Data Using Blind Component Separation. _Psychophysiology_, **41**, 313-325. https://doi.org/10.1111/j.1469-8986.2003.00141.x

13. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., _et al._ (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. _Circulation_, **101**, e215-e220. https://doi.org/10.1161/01.cir.101.23.e215

14. Nuwer, M.R. (1987) Recording Electrode Site Nomenclature. _Journal of Clinical Neurophysiology_, **4**, 121-133. https://doi.org/10.1097/00004691-198704000-00002

15. McFarland, D.J., McCane, L.M., David, S.V. and Wolpaw, J.R. (1997) Spatial Filter Selection for EEG-Based Communication. _Electroencephalography and Clinical Neurophysiology_, **103**, 386-394.

16. Brunner, C., Delorme, A. and Makeig, S. (2013) Eeglab—An Open Source Matlab Toolbox for Electrophysiological Research. _Biomedizinische Technik_, **58**. https://doi.org/10.1515/bmt-2013-4182

17. Reddy, N.P. (2002) Biomedical Signal Analysis: A Case-Study Approach, by Rangaraja M. Rangayyan. _Annals of Biomedical Engineering_, **30**, 983-983. https://doi.org/10.1114/1.1509766

18. Saker, M. and Rihana, S. (2013) Platform for EEG Signal Processing for Motor Imagery—Application Brain

Computer Interface. 2013 *2nd International Conference on Advances in Biomedical Engineering*, Tripoli, 11-13 September 2013, 30-33. https://doi.org/10.1109/icabme.2013.6648839

19. Hsu, W.Y. (2012) Application of Competitive Hopfield Neural Network to Brain-Computer Interface Systems. *International Journal of Neural Systems*, **22**, 51-62. https://doi.org/10.1142/S0129065712002979

20. Wang, Y., Hong, B., Gao, X. and Gao, S. (2006) Phase Synchrony Measurement in Motor Cortex for Classifying Single-Trial EEG during Motor Imagery. *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, 30 August-3 September 2006, 75-78. https://doi.org/10.1109/iembs.2006.259673

21. Varela, F., Lachaux, J.-P., Rodriguez, E. and Martinerie, J. (2001) The Brainweb: Phase Synchronization and Large-Scale Integration. *Nature Reviews Neuroscience*, **2**, 229-239. https://doi.org/10.1038/35067550

22. Rice, J. (2001) Mathematical Statistics and Data Analysis. Duxbury Press, Grove, CA.

23. Cohen, M.E. and Hudson, D.L. (1999) Neural Networks and Artificial Intelligence for Biomedical Engineering. Wiley-IEEE Press.

24. Rakotomamonjy, A. and Guigue, V. (2008) BCI Competition III: Dataset II—Ensemble of SVMs for BCI P300 Speller. *IEEE Transactions on Biomedical Engineering*, **55**, 1147-1154. https://doi.org/10.1109/TBME.2008.915728