Outliers resistant methods for Motor Imagery classification

Ana Julia Villar
Computer Engineering Department
Instituto Tecnológico de Buenos Aires
C1437FBH Lavarden 315, Buenos Aires, Argentina
+54-9-11-4193-9382
jvillar@itba.edu.ar

ABSTRACT

Common spatial patterns analysis (CSP) and linear discriminant analysis (LDA) are widely used techniques for spatial filtering and classifying in motor imagery (MI). However, CSP is very sensitive to noise and artifacts.

A method to detect and eliminate anomalous electroencephalogram (EGG) signals before applying CSP is presented. An outlier score of the signal is obtained by calculating the similarities with the other signals of the sample through the Bounded Coordinate System (BCS). Besides, it is proposed to replace the usual estimators of mean, covariance and scale, used in the algorithms, by Olive and Hawkins estimators to get robust versions of BCS and CSP.

The assumption done in LDA that the covariance of each of the classes in MI are identical may not be true; if it is not satisfied, it is better to use quadratic discrimination. Tests to verify this hypothesis and decide which discriminant function must be used are considered.

The performances of the methods are evaluated and compared on EGG data from BCI competition datasets; results show that robust methods outperformed classical techniques, especially for subjects with poor classification accuracy.

CCS Concepts

• **Applied computing → Bioinformatics**

Keywords

Brain computer interface; motor imagery; robustness; bounded coordinate system; common spatial patterns; discriminant analysis.

1. INTRODUCTION

Motor imagery is a neural activity produced when a subject voluntarily imagines making a movement, eg. right hand movement. Imagining a movement produces neural activity that is spatio-temporally similar to the activity generated during real movement. It is used as a technique to enhance motor learning and to improve neurological rehabilitation in patients after stroke. Repeated practice of MI can induce plasticy changes in the brain [1].

In MI based on Brain-Computer Interface (BCI), the electro-encephalogram signals are recorded while the patient imagines various types of movements. The characteristics used to quantify the EEG activity are extracted; subsequently classifiers are applied to discriminate between two or more movements imagined; finally each type of activity is assigned to a particular control signal.

EEG signals are widely affected by a variety of contaminations or artifacts, such as loose electrodes, eye movement and blink, heart and muscle activity, head and body movement as well as external interference due to power sources [2]. Since the shape of the neuro-logical phenomenon is affected, artifacts can reduce the performance of BCI-based systems. Noise reduction methods and outlier detection aims to find anomalies in the data and remove or downweigh their influence. The use of methods that consider these facts will provide more robust systems.

The multichannel EEG signals can be seen as multivariate time series (MTS). The distance-based algorithm proposed by Wang [3] for detecting outlying samples in MTS datasets can be applied to remove signals with artifacts. It is based on the Bounded Coordinate System technique, introduced by Huang et al. [4], which is used to compute the distance between two MTS samples. Then, the outlier score of a signal is calculated to find the top outlying MTS samples and later to eliminate them.

Common spatial patterns analysis [5] is a supervised method which is successful calculating spatial filters that extract discriminative activity and reduce feature dimensions in motor imagery BCI. It projects multichannel signals into a subspace, where the differences between classes are emphasized and the similarities are minimized. CSP is sensitive to outliers because it involves sample covariance estimates. Yong et al [6] proposed a robust version of the CSP algorithm using the minimum covariance determinant to estimate the covariance matrices and the median absolute deviation to estimate the variance of the projected EEG signals. Another robust estimators of covariance, such as Stahel-Donoho and MM- estimators were used in [7] to robustify CSP.

Linear Discriminant Analysis [8], a simple classifier that provides acceptable accuracy without high computation requirements, has been used successfully in numerous BCI systems. However, it assumes the equality of the covariance matrices of the classes, which is not always a true hypothesis; in that case it should be used Quadratic Discriminant Analysis (QDA) [9].

This paper presents algorithms to process and classify EEG signals that combine several robust methods. First the anomalous signals of the sample are detected and eliminated, later with the resulting signals, a robust CSP algorithm is applied using the

Olive and Hawkins estimator [10] for the covariance matrix. Once the features of each class have been extracted, a test is applied to determine if their covariance matrices are equal or not, and a classifier is constructed using LDA or QDA as appropriate.

Section 2 will provide the methods to detect and eliminate the anomalous signals and to classify the non-discarded ones. In Section 3 the dataset used to compare the performance of the proposed techniques is presented and the way in which it is processed is explained. Results will be shown in Section 4 and the conclusions will be established in Section 5.

2. METHODOLOGY

In BCI design the goal is to process the EEG signal to translate into the mental state of the user. Lotte [11] presented the general architecture of an EEG signal processing system for BCI, whose two main steps are feature extraction and classification.

Let $\mathcal{X} = \{X_1, \cdots, X_{\mathrm{M}}\}$ be a sample of M training EEG trials corresponding to two different mental states (classes 1 and 2), where $X_{\mathrm{j}} \in \mathbb{R}^{N\times C}$ is the data matrix which corresponds to a trial j ($j = 1, \cdots, M$) of imaginary movement, with N the number of observations in each trial and C the number of channels; each X_{j} can be considered as a MTS.

2.1 MTS Outlier Detection

The outlier detection in MTS dataset is to find top outlying MTS samples that have the largest outlier scores. Wang [3] proposed a distance-based algorithm where the Bounded Coordinate System is used to compute the similarity between two MTS samples.

The BCS method characterizes a multidimensional point set by the mean value of elements and a few top ranked principal components bounded by two furthermost projections. In this coordinate system, each of its coordinate axes is identified by the Principal Components [12] and bounded by the range of data projections along the axis. Principal Components only indicate the directions of coordinate axes; its corresponding Bounded Principle Component (BPC) identifies a line segment bounded by two furthermost projections on it. To capture the data information more accurately and avoid the negative effect of noises, the length of BPC can be redefined by the standard deviation of data projections.

Given a matrix $X \in \mathbb{R}^{N\times C}$ such that $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_C)$ with $\boldsymbol{x}_i = (\boldsymbol{x}_{1i}, \cdots, \boldsymbol{x}_{Ni})^t$ ($i = 1, \cdots, C$), its representation is $\mathrm{BCS}(X) = \left(O^X, \ddot{\Phi}_1^X, \cdots, \ddot{\Phi}_C^X\right)$, where O^X is the vector of means for all $\boldsymbol{x}_i$ and $\ddot{\Phi}_i$ is the BPC for $\boldsymbol{x}_i$. The length of the bounded principal component $\|\ddot{\Phi}_i\|$ is $2\sigma_i$, where σ_i is the average distance from the origin of coordinate system (the mean O^X) to a projection on the principal component Φ_i (see details in [3], [4]).

2.1.1 Robust BCS

Principal component analysis (PCA) [12] uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. PCA can be done by eigenvalue decomposition of the data covariance matrix after mean centering.

The sample mean and sample covariance are sensitive to outliers and therefore highly non-robust; they can be replaced by the highly outlier-resistant OH estimators of multivariate location and dispersion proposed by Olive and Hawkins [10]. These OH estimators are obtained by first generating trial estimates and then using a concentration technique from each trial fit to obtain attractors (see details in [10], [13]).

Standard deviation is a measure of the average distance of data points from their mean; it is also sensitive to outliers. The median absolute deviation (MAD) [14] is a robust measure of the variability of a univariate sample of quantitative data.

A robust version of BCS ($\widetilde{BCS}$) can be constructed by obtaining the principal components from the OH covariance estimator, and then by replacing the vector of means with the OH location estimator and using the MAD of the principal component projections to get the length of the components.

2.1.2 Outlier Score

The similarity between two MTS can be measured by comparing their corresponding BCS. A modified version of the distance between two MTS, defined by Wang, can be obtained by replacing their corresponding BCS with their robust versions. Given two MTS samples $X \in \mathbb{R}^{N\times C}$ and $Y \in \mathbb{R}^{N\times C}$ and their corresponding robust BCS representations $\widetilde{BCS}(X) = \left(\tilde{O}^X, \widetilde{\Phi}_1^X, \cdots, \widetilde{\Phi}_C^X\right)$ and $\widetilde{BCS}(Y) = \left(O^Y, \widetilde{\Phi}_1^Y, \cdots, \widetilde{\Phi}_C^Y\right)$, the distance between them is computed by:

$$D_{\mathrm{BCS}}\left(\widetilde{BCS}(X), \widetilde{BCS}(Y)\right) = \|\tilde{O}^X - \tilde{O}^Y\| + \sum_{i=1}^{C} \|\widetilde{\Phi}_i^X - \widetilde{\Phi}_i^Y\|,$$

where $\|\cdot\|$ is the Euclidean norm.

Let $\mathfrak{D}$ be a MTS dataset whose cardinality is $|\mathfrak{D}|$, $k \in \mathbb{N}$ a parameter such that $1 \leq k \leq |\mathfrak{D}| - 1$. The outlier score for $X \in \mathfrak{D}$ is defined as:

$$D_k(X) = \frac{1}{k} \sum_{\substack{l=1 \\ X_l \in NN(X)}}^{k} D_{\mathrm{BCS}}(X, X_l)$$

where $NN(X)$ denotes the set of the k nearest neighbors of X in $\mathfrak{D}$ according to the D_{BCS} distance. This score represents the degree of outlierness for each sample.

2.1.3 Identification of Outliers

A way for labeling outliers is the interquartile range method developed by Tukey [15]. In this context, it only interests to eliminate the signals with large outlier scores. Thus, the anomalous signals will be those whose scores exceed in value the upper cut-off point C^{Tu}, given by:

$$C^{Tu} = q_3 + 1.5\ (q_3 - q_1),$$

where q_1 and q_3 are the sample quartiles of the signals' outlier scores.

A modification of the previous is the median rule ([16], [17]), which upper cut-off point C^{Med} is given by:

$$C^{Med} = q_2 + 2.3\ (q_3 - q_1),$$

where q_2 is the sample median of the signals' outlier scores.

Let $\mathcal{X}_e$ the set of trials in $\mathcal{X}$ belonging to the class e ($e = 1,2$). For each class e, let $\mathfrak{D} = \mathcal{X}_e$ and k be a given number such that $1 \leq k \leq |\mathfrak{D}| - 1$. For each $X_{\mathrm{j}} \in \mathfrak{D}$, the outlier score $D_k(X_{\mathrm{j}})$ is calculated; then the sample X_{j} is considered an outlier if $D_k(X_{\mathrm{j}}) \geq C^P$, where C^P is the upper cut-off point C^{Tu} or C^{Med}. The set $\widetilde{\mathcal{X}}_e \subseteq \mathcal{X}_e$ is constructed with the no outliers samples in $\mathfrak{D}$.

2.2 Feature Extraction

One way to extract features from multiple EEG channels is to use spatial filtering, for example CSP [5], which performs a linear combination of EEG signals for enhancing signals coming from a particular area of the brain.

2.2.1 Robust CPS

CSP yields a decomposition of the signal parameterized by a matrix $W \in \mathbb{R}^{m\times C}$. Let $\Sigma_e \in \mathbb{R}^{C\times C}$ the covariance matrix of the band-pass filtered signals in $\tilde{\mathcal{X}}_e$ $(e = 1, 2)$. W projects signal $X \in \mathbb{R}^{N\times C}$ in the original sensor space $X_{\text{CSP}} = XW^t$, where the m spatial filters (rows of W) $\boldsymbol{w}_h \in \mathbb{R}^C$ $(h = 1, \cdots, m)$ maximize the Rayleigh quotient $\text{J}(\boldsymbol{w}) = \boldsymbol{w}^t\Sigma_1\boldsymbol{w}/w^t\Sigma_2\boldsymbol{w}$.

The vector $\boldsymbol{w}$ in $\text{J}(\boldsymbol{w})$ can be achieved by solving the generalized eigenvalue problem $\Sigma_1\boldsymbol{w} = \lambda\Sigma_2\boldsymbol{w}$. Then the matrix W in X_{CSP} consists of m generalized eigenvectors $\boldsymbol{w}_h$, which have associated eigenvalues λ_h $(h = 1, \cdots, m)$.

Let $\tilde{\mathcal{X}} = \tilde{\mathcal{X}}_1 \cup \tilde{\mathcal{X}}_2$ and $\tilde{M} = |\tilde{\mathcal{X}}|$ $(\tilde{M} \le M)$. For a given trial matrix $X_{\text{j}} \in \tilde{\mathcal{X}}$, the normalized sample covariance matrix is obtained as:

$$S_j = Y_{\text{j}}{}^tY_{\text{j}}/\text{tr}(Y_{\text{j}}{}^tY_{\text{j}}) ,$$

where Y_{j} is the matrix X_{j} after centering and scaling it $(j = 1, \cdots, \tilde{M})$. For each class e $(e = 1, 2)$ an estimator of Σ_e is computed by averaging the covariances matrices of each trial as:

$$\hat{\Sigma}_e = \frac{1}{|\tilde{\mathcal{X}}_e|}\sum_{j\in\tilde{\mathcal{X}}_e} S_j.$$

Such computation of covariance matrices assumes that the signals have zero mean, which is true in practice for band-pass filtered signals.

One way to make the CSP robust is to replace the covariance estimators with a robust version, such as the OH estimators. Then, replacing $\hat{\Sigma}_e$ by the robust $\tilde{\Sigma}_e$, a matrix $\tilde{W}$ will be generated as output.

2.2.2 Feature Selection

The classical measure for the selection of CSP filters is based on the previous eigenvalues; the m filters $(m = 2p)$, corresponding to the p largest and the p lowest eigenvalues are used.

Each eigenvalue is the relative variance of the signal filtered with the corresponding spatial filter. This measure is not robust to outliers if it is based on simply pooling the sample covariance matrices in each class. A simple way to fix this issue was proposed in [18] using a score, where the ratio of medians of the variance of the filtered signal within each trial is calculated as:

$$\text{score}(\boldsymbol{w}) = med_1/(med_1 + med_2),$$

with $med_e = \text{median}_{X_j\in\tilde{\mathcal{X}}_e}[\text{var}(\boldsymbol{w}X_j)]$ $(e = 1, 2)$. Then, the m filters corresponding to the p largest and the p lowest scores are used.

Let $\mathcal{W} = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_p, \boldsymbol{w}_{C-p+1}, \cdots, \boldsymbol{w}_C\}$ the set of such filters. A CSP feature f_h for a signal X is defined as:

$$f_h = \ln(\boldsymbol{w}_h{}^tX^tX\boldsymbol{w}_h) = \ln[\text{var}(\boldsymbol{w}_hX)],$$

for $\boldsymbol{w}_h \in \mathcal{W}$ $(h = 1, \cdots, m)$. Using any of the previous techniques, the vector $\boldsymbol{F}_j = [f_{j1} \cdots f_{jm}]$ of features is found for each signal $X_{\text{j}} \in \tilde{\mathcal{X}}$, generating the set $\mathcal{F} = \{\boldsymbol{F}_1, \cdots, \boldsymbol{F}_{\tilde{M}}\}$.

2.3 Classification

A preliminary test of equality of covariance matrices is often used in discriminant analysis to decide whether linear (LDA) or quadratic discriminant analysis (QDA) should be applied in a given problem. Unlike LDA, in QDA there is no assumption that the covariance of each of the classes is identical.

Box [19] proposed a likelihood-ratio test (LRT) statistic for testing the hypothesis of equal covariance matrices. However, in the high-dimensional setting, where the dimension can be much larger than the sample size, the conventional testing procedures such as the LRT either perform poorly. Cai et al [20] developed a test that is powerful against sparse alternatives and robust with respect to the population distributions.

For each X_{j} in the sample $\tilde{\mathcal{X}}$ $(j = 1, \cdots, \tilde{M})$, the features vector $\boldsymbol{F}_j \in \mathcal{F}$ is obtained. Let Φ_e the set of these vectors in $\mathcal{F}$ whose signals are in class $(e = 1, 2)$, and let $\breve{\boldsymbol{\mu}}_e$ and $\breve{\Sigma}_e$ be the sample mean and covariance obtained with the data in Φ_e.

For a new signal $X \notin \mathcal{X}$ $(X \in \mathbb{R}^{N\times C})$, let π_e be the probability that the signal X to classify belongs to class $\boldsymbol{e}$ and $\boldsymbol{F}$ be the features vector for X. Let $\breve{\Sigma}$ be the within groups covariance matrix given by the pooled version of the different scatter matrices $\breve{\Sigma} = \sum_{e=1}^{2}\pi_e\breve{\Sigma}_e$.

X is assigned to class l for which $D_l(\boldsymbol{F}) = \min_{e=1,2} D_e(\boldsymbol{F})$, where $D_e^2(\boldsymbol{F}) = (\boldsymbol{F} - \breve{\boldsymbol{\mu}}_e)^t\breve{\Sigma}^{-1}(\boldsymbol{F} - \breve{\boldsymbol{\mu}}_e) - 2\ln\pi_e$ in the case of LDA, or $D_e^2(\boldsymbol{F}) = (\boldsymbol{F} - \breve{\boldsymbol{\mu}}_e)^t\breve{\Sigma}_e{}^{-1}(\boldsymbol{F} - \breve{\boldsymbol{\mu}}_e) + \ln[\det(\breve{\Sigma}_e)] - 2\ln\pi_e$ in the case of QDA.

3. NUMERICAL EXPERIMENTS

3.1 Dataset

In order to compare the performance of the different techniques presented, a dataset of motor imagery EEG signals of BCI competition was used. The dataset was collected in a multiclass setting with the subjects performing more than two different MI tasks, however, the algorithms were evaluated on two-class problems by selecting only signals of left and right hands MI trials.

The data [21] were recorded using 22 electrodes from 9 subjects who performed left-hand, right-hand, foot and tongue MI. The training and testing sets contain 72 trials for each class.

3.2 Data Processing

The EEG signals were band-pass filtered in 8-30 Hz, using a 5th order Butterworth filter. For each trial, the features were extracted from the time segment located from 0.5s to 2.5s after the cue instructing the subject to perform MI. The Regularized Common Spatial Pattern (RCSP) toolbox [22] was adapted to process the signals.

Let $\mathcal{X}$ be the set of the filtered signals of the training set and $\mathcal{Y}$ be the set of the filtered signals of the testing set. The anomalous signals in $\mathcal{X}$ were detected and eliminated (step 1) and the subset $\tilde{\mathcal{X}} \subseteq \mathcal{X}$ of the no outliers signals were used to extract the set of features $\mathcal{F}$ (step 2). Each signal in $\mathcal{Y}$ is classified (step 3) and the subjects' performances were calculated.

4. RESULTS

The dataset was processed in the way explained in the previous section. The performances for the classic CSP using all the signals in the training set with $p = 3$ pairs of spatial filters combined with LDA (CSP+LDA) (e.g. [18]), denoted by *CSP*, are reported to compare them with those of the others methods.

In step 1, the number k nearest neighbors was varied between 1 and 71. In step 2, although in the BCI literature the usual number of features considered is between 2 and 6, the number of pairs p was varied between 1 and 11. In step 3, the classification of the testing signals was done using LDA (Table 1), QDA (Table 2) or choosing the discriminant function through the tests of Box (Table 3) or Cai et al (Table 4) with significant level 0.05 for both tests.

Tables 1 to 4 show only the best performance for the method studied for each of the subjects S_i $(i = 1,\cdots,9)$ (i.e. the performance for values of p and k which give the best accuracy when classifying the signals of the testing set).

The notation $x_1x_2x_3x_4$ used is the following: $x_1 = W$ if Wang's technique was used to obtain the outlier scores or $x_1 = O$ if the outlier scores were obtained as in Section 2.1.2; $x_2 = t$ if the identification of outliers were done using the upper cut-off point given by the method of Tukey or $x_2 = m$ if the median rule is applied; $x_3 = C$ if CSP was based on sample covariance matrix or $x_3 = R$ if the robust CSP of Section 2.2.1 is used; $x_4 = e$ if the features were selected using extreme eigenvalues or $x_4 = s$ if scores were considered.

The results show that the elimination of the anomalous signals before the extraction of features during the learning leads to a better classification for all subjects.

The use of the upper cut-off point obtained with the median rule produces better accuracies than those with Tukey's in most cases.

For almost all subjects, the performances are better if the filters, obtained with classic CSP, are selected using the extreme scores and the linear discrimination function is considered.

The feature extraction using robust CPS only produces better results for some subjects. This may be due to the fact that the samples used at this stage are non-outliers seen as MTS, and therefore, the possibility of anomalous observations within such signals is reduced; if the data has Gaussian distribution, the sample mean and covariance are the most efficient estimators [23].

The discriminant function that gives the greatest accuracy depends on the subject, as it can see by comparing the Tables 1 and 2. Tables 3 and 4 show that the application of test of Cai et al gives performances similar or superior to those obtained with Box's test, although, in some cases, the decision of the test does not determine the most convenient discriminate function.

5. CONCLUSIONS

In this paper are presented outliers resistant methods to classify EEG signals from motor imagery experiments. The methods are evaluated on MI data by comparing their accuracies when classifying with those obtained with the standard method of CSP combined with LDA.

The elimination of anomalous signals before the extraction of characteristics, the use of robust estimators of mean, covariance and scale, together with the selection of the appropriate discriminant function, produces better results than with CSP+LDA.

Future work could be the development of techniques to automatically determine the values of the parameters k (number of neighbors used when calculating the outlier scores) and p (number of pairs of spatial filters).

Table 1. Best classification accuracies using LDA (for each subject, the best result is displayed in bold characters)

	S1	S2	S3	S4	S5	S6	S7	S8	S9
CSP	*88,9*	*51,4*	*96,5*	*70,1*	*54,9*	*71,5*	*81,3*	*93,8*	*93,8*
WtCe	91,0	57,6	96,5	78,5	64,6	71,5	**84,0**	97,2	93,8
WmCe	91,7	56,3	96,5	77,1	66,0	71,5	**84,0**	97,2	93,8
OtCe	91,7	58,3	**97,9**	79,2	64,6	70,1	82,6	97,2	93,8
OmCe	92,4	58,3	**97,9**	79,2	64,6	70,1	83,3	96,5	93,8
OtCs	**93,8**	**60,4**	**97,9**	79,2	63,2	70,1	81,9	96,5	**94,4**
OmCs	**93,8**	**60,4**	**97,9**	**80,6**	63,9	70,1	**84,0**	96,5	**94,4**
OtRe	88,9	**60,4**	97,2	72,2	**68,8**	**74,3**	75,7	**97,9**	92,4
OmRe	88,9	59,0	97,2	72,2	63,9	**74,3**	75,7	**97,9**	92,4

Table 2. Best classification accuracies using QDA (for each subject, the best result is displayed in bold characters)

	S1	S2	S3	S4	S5	S6	S7	S8	S9
CSP	*88,9*	*51,4*	***96,5***	*70,1*	*54,9*	*71,5*	*81,3*	*93,8*	***93,8***
WtCe	**91,0**	56,3	94,4	**79,9**	65,3	68,1	84,0	95,8	**93,8**
WmCe	**91,0**	56,3	94,4	77,1	65,3	68,1	84,0	95,8	**93,8**
OtCe	**91,0**	56,3	**96,5**	79,2	64,6	69,4	82,6	95,1	93,1
OmCe	88,9	56,3	**96,5**	79,2	65,3	69,4	**84,7**	95,8	93,1
OtCs	88,2	**59,7**	95,8	75,0	56,9	68,8	80,6	94,4	**93,8**
OmCs	88,9	56,3	95,1	75,0	56,9	**73,6**	81,9	95,8	**93,8**
OtRe	89,6	57,6	**96,5**	72,2	65,3	71,5	76,4	**97,9**	90,3
OmRe	88,9	57,6	95,1	70,1	**66,0**	71,5	75,7	**97,9**	89,6

Table 3. Best classification accuracies using the classifier determined by Box's test (for each subject, the best result is displayed in bold characters)

	S1	S2	S3	S4	S5	S6	S7	S8	S9
WtCe	89,6	57,6	96,5	77,8	65,3	68,8	83,3	96,5	93,8
WmCe	91,0	56,3	95,8	77,1	65,3	70,1	**84,0**	95,8	93,8
OtCe	91,0	56,3	**97,9**	79,2	64,6	69,4	82,6	96,5	93,1
OmCe	88,9	56,3	96,5	79,2	65,3	68,8	83,3	96,5	93,1
OtCs	88,2	**60,4**	96,5	79,2	59,7	70,1	81,9	95,8	**94,4**
OmCs	**93,8**	56,3	96,5	**80,6**	63,9	70,1	**84,0**	95,8	**94,4**
OtRe	89,6	57,6	96,5	72,2	**68,8**	**74,3**	75,7	**97,9**	91,0
OmRe	88,9	57,6	96,5	70,8	63,9	**74,3**	75,7	**97,9**	91,0

Table 4. Best classification accuracies using the classifier determined by Cai's test (for each subject, the best result is displayed in bold characters)

	S1	S2	S3	S4	S5	S6	S7	S8	S9
WtCe	91,0	57,6	96,5	78,5	64,6	71,5	**84,0**	97,2	93,8
WmCe	91,0	56,3	96,5	77,1	66,0	71,5	**84,0**	97,2	93,8
OtCe	91,0	56,3	**97,9**	79,2	64,6	70,1	82,6	97,2	93,8
OmCe	88,9	56,3	**97,9**	79,2	64,6	68,8	83,3	96,5	93,8
OtCs	**93,8**	**60,4**	**97,9**	79,2	63,2	70,1	81,9	96,5	**94,4**
OmCs	**93,8**	59,0	**97,9**	**80,6**	63,9	70,1	**84,0**	96,5	**94,4**
OtRe	89,6	60,4	97,2	72,2	**68,8**	**74,3**	75,7	**97,9**	92,4
OmRe	88,9	59,0	97,2	72,2	63,9	**74,3**	75,7	**97,9**	92,4

The proposed methods show more accuracy in the classification which leads to obtain neurophysiologically more relevant spatial filters. This would allow to extend the use of the BCI, increasing its reliability and reducing the recognition of erroneous mental commands from the user.

6. REFERENCES

[1] Jackson, P., Lafleur, M., Malouin, F., Richards, C. and Doyon, J. 2001. Potential role of mental practice using motor imagery in neurologic rehabilitation. Archives of physical medicine and rehabilitation *Archives of physical medicine and rehabilitation. 82. 1133-1141*.

[2] Fatourechi, M., Bashashati, A., Ward, R.K., and Birch, G.E. 2007. EMG and EOG artifacts in brain computer interface systems: A survey. *Clin. Neurophysiol.* 2007, *118*, 480–494.

[3] Wang, X. 2011. Two-phase Outlier Detection in Multivariate Time Series *Eighth International Conference on Fuzzy Systems and Knowledge Discovery* 2011 (FSKD) IEEE.

[4] Huang, Z., Shen, H. T., Shao, J., Zhou, X. and Cui, B. 2009. Bounded Coordinate System Indexing for Real-Time Video Clip Search. *ACM Transactions on Information Systems*, Volume 27 Issue 3, May 2009 Article No. 17.

[5] Hoffman, I. Farkaš, A. 2013. Using common spatial patterns for EEG feature selection. *Technical report TR-2013-040. Comenius University in Bratislava*, 2013.

[6] Yong, X., Ward, R. K. and Birch, G. E. 2008. Robust Common Spatial Patterns for EEG Signal Preprocessing. *30th Annual International IEEE EMBS Conference,* Vancouver, British Columbia, Canada, August 20-24, 2008.

[7] Villar, A. J. 2016. Comparative Study of Robust Methods for Motor Imagery Classification based on CSP and LDA. *IFMBE Proceedings Volume 60, 2017, Pages 126-129 7th Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia*.

[8] Fisher, R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics **7** (2): p179–188, 1936.*

[9] Tharwat, A. 2016. Linear vs. quadratic discriminant analysis classifier: a tutorial. *International Journal of Applied Pattern Recognition. 3. 145.*

[10] Olive, D. J. and Hawkins, D. M. 2010. Robust Multivariate Location and Dispersion. Preprint at: http://www.math.siu.edu/olive/preprints.htm

[11] Lotte, F. 2014. A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces. *Guide to Brain-Computer Music Interfacing*, Springer, 2014.

[12] Pearson, K. 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine. **2** (11): 559–572.*

[13] Olive, D. J. 2004. A resistant estimator of multivariate location and dispersion. *Computational Statistics and Data Analysis, Vol. 46, pp. 99–102*.

[14] Gauss, C. F. 1816. Bestimmung der Genauigkeit der Beobachtungen. *Zeitschrift für Astronomie und verwandte Wissenschaften. **1**: 187–197.*

[15] Tukey, J. W. 1977. *Exploratory Data Analysis*. Addison-Wesley Reading, Massachusetts.

[16] Carling, K. 2000. Resistant outlier rules and the nongaussian case. *Computational Statistics & Data Analysis 33 (2000): 249- 258.*

[17] Olewuezi, N.P. 2011. Note on the Comparison of Some Outlier Labeling Techniques. *Journal of Mathematics and Statistics 7 (4): 353-355*, 2011.

[18] Blankertz, B., Tomioka, R., Lemm, S., Kawancabe, M., and Müller, K. R. 2008. Optimizing Spatial Filters for Robust EEG Single-Trial Analysis. *IEEE Signal Processing Magazine, Vol. XX, 2008.*

[19] Box, G. E. P. 1949. A General Distribution Theory for a Class of Likelihood Criteria *Biometrika* Vol. 36, No. 3/4 (Dec., 1949), pp. 317-346.

[20] Cai, T., Liu, W. and Xia, Y. 2013. Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings. *Journal of the American Statistical Association Vol. 108, No. 501, 2013.*

[21] Naeem, M., Brunner, C., Leeb, R., Graimann, B. and Pfurtscheller, G. 2016. Separability of four-class motor imagery data using independent components analysis. *Journal of Neural Engineering, Vol. 3, no. 3: 208-216, 2016.*

[22] RCSP toolbox at: sites.google.com/site/fabienlotte/code-and-softwares

[23] Lehmann, E.L. and Casella, G. 1998. *Theory of Point Estimation* (2nd ed.). Springer.