CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva

Volume 10 Number 1
April 2019
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Chilean Journal of Statistics
Vol. 10, No. 1, April 2019, 3–20

Data Science
Research Paper

A combined Fourier analysis and support vector machine for EEG classification

Jhonnata B. de Carvalho[1,*], Murilo C. Silva[2], George F. von Borries[2], André L.S. de Pinho[1] and Ricardo F. von Borries[3]

[1]Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, Brazil,
[2]Department of Statistics, Universidade de Brasília, Brasília, Brazil,
[3]Department of Electrical and Computer Engineering, University of Texas at El Paso, El Paso, United States of America

(Received: 30 September 2018 · Accepted in final form: 10 March 2019)

Abstract

This paper introduces a method for the classification of electroencephalogram (EEG) data combining Fourier analysis, support vector machine (SVM) and a weighting system, called WFF-SVM, that provides high correct classification rates (accuracy) using a small training data set. Basically, an SVM classifier is calculated for each frequency in the periodogram and a proposed weighting system, based on the error rate of each SVM classifier, is used to obtain a final decision value. Also, it is shown that principal component analysis can be used to identify the best group of EEG channels to apply to the classification method, improving the correct classification rate. Two applications with real data are presented. The first application uses a public data set of epileptic patients and compares the proposed method with other methods presented in the literature. In this case, the correct classification rate obtained was 100%. The second application consists of EEG data collected from a subject submitted to 10 visual stimuli and the correct classification rate obtained was 95.31%. The classifier WFF-SVM combines multiple existing techniques, each one of them widely used in time series and dimensionality reduction problems. Our paper combines standard signal processing techniques to obtain high classification rates of EEG data.

Keywords: Epilepsy data · Periodogram · Principal components analysis · Simple moving averages · Supervised learning.

Mathematics Subject Classification: Primary 62H25 · Secondary 68Q32.

1. Introduction

Machine learning (ML) techniques have been gaining prominence due to real-world problems as well as large databases. Basically, one can divide ML methods into two classes, supervised learning and unsupervised learning. In unsupervised learning, the method has to recognize the groups by existing standards with a certain criterion. This type of learning tries to gain some understanding of the process that generated the data, e.g., the K-means method applied in DNA gene expression and Internet newsgroups (Ding and He, 2004),

*Corresponding author. Email: jhon_dbz@yahoo.com.br

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

clustering with hill-climbing optimization method applied to bee species (Friedman and Rubin, 1967), botanical data (Rubin, 1967) and in clustering of plants, wines and heart diseases (Souza et al., 2017). In supervised learning, groups (or classes) are known a priori and it is necessary to provide examples for method training. These methods are often used in classification and regression problems, e.g., logistic regression in the prediction of a financial crisis in Latin American companies (Giampaoli et al., 2016), in the fault diagnosis in chemical processes using Fisher discriminant analysis (Chiang et al., 2000), SVM classification in validation of cancer tissue samples (Furey et al., 2000). However, our interest is in the classification of electroencephalography signals.

An EEG are recordings of the electrical potentials produced by the brain (Bronzino, 1999; Buzsaki, 2006). Basically, the digital EEG is a time series containing information of the electrical activity generated by the brain. EEG has vast application in areas such as epilepsy detection (Andrzejak et al., 2001), emotion regulation using neurofeedback (Ruiz et al., 2014), affective neuroscience (Sitaram et al., 2011), and brain computer interface (Kübler et al., 2001; Wolpaw et al., 2002). For an efficient classification of EEG, an algorithm should address two main problems: feature extraction and classification method. Several methods have been used to extract features of EEG data, such as discrete wavelet transforms (DWT) (Jahankhani et al., 2006; Subasi, 2007; Subasi and Gursoy, 2010), amplitude values (Kaper et al., 2004), clustering techniques (Li and Wen, 2011), autoregressive and adaptive autoregressive parameters (Penny et al., 2000; Pfurtscheller et al., 1998), wavelet packet decomposition and extracted eigenvalues from the resultant wavelet coefficients using principal component analysis (PCA) (Acharya et al., 2012), continuous wavelet transform (CWT), higher order spectra (Acharya et al., 2013), approximate entropy and DWT (Ocak, 2009), analytic time-frequency flexible wavelet transform and fractal dimension (Sharma et al., 2017).

In order to classify a set of extracted features, several pattern recognition methods have been used, such as artificial neural network (Guo et al., 2009; Jahankhani et al., 2006; Nigam and Graupe, 2004; Subasi, 2007), mixture of expert model (Subasi, 2007), linear discriminant analysis (Subasi and Gursoy, 2010), SVM (Chandaka et al., 2009; Subasi and Gursoy, 2010), decision trees (Polat and Günes, 2007), least squares SVM (Li and Wen, 2011; Übeyli, 2010) and hidden markov models (Chiappa and Bengio, 2004). For a more complete review refer to Lotte et al. (2007).

Recently several algorithms have been developed to classify EEG in a variety of applications, such as in Zhang et al. (2016), which proposed a linear Bayesian discriminant with a Laplace prior, named sparse Bayesian method by exploiting a Laplace prior. A major advantage of this method is that it estimates automatically all the parameters of the classifier, without the need to use cross-validation. However, we point out that any Bayesian procedure needs a suitable prior distribution and although the Laplace distribution has been suggested it is conceivable that for a particular application a better prior distribution can be found. Wang et al. (2016) introduces a new approach that utilizes spatiotemporal feature extraction with multivariate linear regression (MLR) to learn discriminative of steady-state visual evoked potentials (SSVEP) features, for improving the detection accuracy. SSVEP are signals that are natural responses to visual stimulation at specific frequencies. MLR is implemented on dimensionality reduced EEG training data and a constructed label matrix to find optimally discriminative subspaces. Jiao et al. (2017) proposed a method that is an extension of multiset canonical correlation analysis (MsetCCA), called multilayer correlation maximization (MCM) model for further improving SSVEP recognition accuracy. MCM combines advantages of both Canonical Correlation Analysis and MsetCCA by carrying out three layers of correlation maximization processes. Zhang et al. (2018) introduced a new method, called multi-kernel extreme learning machine (MKELM) to EEG classification. Basically, this method transforms the EEG through the common spatial pattern (CSP) and inserts a kernel function in the extreme learning machine (ELM). The MKELM provides a way to circumvent calculation of the hidden layer outputs and inherently encode it in a kernel matrix.

The proposed WFF-SVM is a classifier based on the SVM and the Fourier transform, providing the periodogram as feature extraction. In addition, it uses a weighting system based on the error rate. Thus, we call this classifier weighted Fourier frequencies and SVM, WFF-SVM for short. The WFF-SVM classifier differs from the other methods because it requires just one data transformation (Fourier), which leads to a good capacity to discriminate among groups. The PCA is used to identify the most active regions of the brain, providing the use of fewer electrodes and reducing the complexity of the data, since some electrodes pick up only noises, whereas the other methods ended up losing information by reducing the dimension based on the application of CSP or PCA. In relation to the Fourier transform, we observed that analyzing the signals in the frequency domain (periodogram), as shown in Figure 1, allows us to discriminate the signals for some frequencies. Our classifier takes into account the most distinct frequencies for classification through the weighting system. However, we point out that the choice of the kernel function is not unique, but for our applications the results are virtually the same by considering different kernels, suggesting a robust procedure.

Visual stimuli are commonly used to understand different components, such as color, texture, motion, objects, readability (text versus nontext), and others (Thomas and Vinod, 2017). Moreover, visual stimuli are also used in biometric authentication (Zuquete et al., 2010), emotion classification (Wang et al., 2014), person identification (Das et al., 2009), and others. We tested our classification method using real-world EEG data of two main applications: epilepsy and vision. The first application (described in Subsection 4.1) uses a publicly available data set described in Andrzejak et al. (2001), already used in previous works on EEG classification, and it allows a direct comparison of our classification method to other methods presented in the literature. In this application, the proposed method achieved a correct classification rate of 100.00% under a relatively simple model, showing that the proposed method performs well compared to other methods in the literature. The second application (described in Subsection 4.2) uses a data set collected in an experiment conducted at the University of Texas at El Paso in which the EEG data are acquired while the subject is submitted to visual stimuli. The proposed method showed a high correct classification rate of 95.31% using only three signals from each class in the training phase.

This paper is organized as follows. Section 2 provides a brief review of the SVM classifier relevant for our work and presents the periodogram, which is used for feature extraction. Section 3 presents our classification method integrating Fourier data analysis, SVM and a weighting system. Section 4 reports the performance of our method using real-world data of two applications and compares it with concurrent methods found in the literature. Section 5 provides some discussions, conclusions and recommendations for future work.

2. Background

In this section, the methods used in the WFF-SVM classifier are described. The first method is the SVM and it includes three main blocks: the basic classifier, parameters estimation and SVM with nonlinear functions. The other methods are the Fourier analysis, periodogram, and the technique of simple moving averages.

2.1 Support vector machine

The SVM is a pattern recognition technique that has been widely used in problems like regression and classification (Hastie et al., 2008; Hornik et al., 2006; Theodoridis and Koutroumbas, 2008; Vapnik, 1996). In classification problems the SVM technique separates two classes (say W_1 and W_{-1}) by a hyperplane $\langle \boldsymbol{\beta}, \boldsymbol{x} \rangle + \beta_0 = 0$, where $\langle \cdot, \cdot \rangle$ is the inner product, $\boldsymbol{x}, \boldsymbol{\beta} \in \mathrm{R}^{\mathrm{D}}$ and $\beta_0 \in \mathrm{R}$, corresponding to the decision function

$$f(\boldsymbol{x}) = \mathrm{sign}(\langle \boldsymbol{\beta}, \boldsymbol{x} \rangle + \beta_0). \tag{1}$$

The optimal hyperplane is defined as the one maximizing the margin of separation between classes. Note that the optimal hyperplane does not necessarily guarantee a complete separation of points from the two classes. This hyperplane can be constructed using Lagrange multipliers and then solving a constrained convex optimization problem.

Consider a set of training samples $\boldsymbol{x}_i$ with $i = 1, 2, \ldots, N$, then the primal optimization problem along with the soft margin method (Cortes and Vapnik, 1995) is given by

$$\min_{\boldsymbol{\beta}, \beta_0, \xi_i} \frac{1}{2} \|\boldsymbol{\beta}\|^2 + c \sum_{i=1}^{N} \xi_i, \tag{2}$$

$$\text{subject to} \begin{cases} y_i \left(\langle \boldsymbol{\beta}, \boldsymbol{x}_i \rangle + \beta_0 \right) \geq 1 - \xi_i, \\ \xi_i \geq 0, \ \text{for } i = 1, \ldots, N, \end{cases}$$

where the constant c is previously chosen and determines the influence of the two terms in the minimization problem. The variables ξ_i are known as slack variables measuring the proportional amount of predictions that fall on the wrong side of the margin, and y_i is an indicator variable defined by

$$y_i = \begin{cases} +1, \text{if } \boldsymbol{x}_i \in W_1, \\ -1, \text{if } \boldsymbol{x}_i \in W_{-1}. \end{cases}$$

Using Lagrange multipliers (Hastie et al., 2008), one can obtain the Wolfe dual function given by

$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k \langle \boldsymbol{x}_i, \boldsymbol{x}_k \rangle. \tag{3}$$

The solution is obtained by maximizing L_D, a simple convex optimization problem which must satisfy the conditions $0 \leq \alpha_i \leq c$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$.

One can also generalize the SVM technique using a non-linear discriminant (unlike the hyperplane). In this case, a mapping is used in a larger number of dimensions. It can be shown (Theodoridis and Koutroumbas, 2008) that this mapping in a larger number of dimensions can be implemented without increasing the computational demand by replacing the inner product $\langle \boldsymbol{x}_i, \boldsymbol{x}_k \rangle$ in Equation (3) by a kernel $K(\boldsymbol{x}_i, \boldsymbol{x}_k)$ to compute the inner product in a higher dimensional space. In this study, we consider two popularly used kernels:

- Gaussian kernel: $K_1(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left\{-\sigma ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2\right\}$;
- Polynomial kernel: $K_2(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle^d$;

where σ and d are kernel width and polynomial degree, respectively. Note when $d = 1$, the polynomial kernel is called linear kernel.

2.2 Fourier analysis

Fourier frequency analysis is a very important tool in signal processing and the periodogram is one of its subproducts (Fuller, 1996). The periodogram shows how the covariance of a time series is distributed in frequency. Any stationary time series can be represented as a sum of sines and cosines (Fuller, 1996), that is, a discrete stationary time series $\{X_t\}$, where $t = 1, \ldots, n$, (n being odd) can be represented by

$$X_t = \frac{a_0}{2} + \sum_{j=1}^{\lfloor n/2 \rfloor} a_k \cos(\omega_k t) + b_k \sin(\omega_k t),$$

where $\lfloor n/2 \rfloor$ is the largest integer less than or equal to $n/2$, a_k and b_k are parameters to be estimated. Also, the Fourier frequencies are defined by

$$\omega_k = \frac{2\pi k}{n}, \qquad k = 0, \ldots, \left\lfloor \frac{n}{2} \right\rfloor.$$

The periodogram can be defined as the sequence $\{J_k\}$, where

$$J_k = \frac{n}{2}\left(a_k^2 + b_k^2\right), \tag{4}$$

and the sum of squares removed by $\cos(\omega_k t)$ and $\sin(\omega_k t)$ is

$$J_k = \frac{2}{n}\left[\left(\sum_{t=1}^{n} X_t \cos(\omega_k t)\right)^2 + \left(\sum_{t=1}^{n} X_t \sin(\omega_k t)\right)^2\right].$$

Thus, the value of the periodogram at frequency ω_k is the contribution from this frequency to the sum of squares of $\{X_t\}$ or, equivalently, its energy.

Some periodograms shown in this paper are smoothed using a moving average technique (Brockwell and Davis, 2002). Considering $\{J_k\}$ a sequence of points in the periodogram, for some $\alpha \in \mathbb{N}$, we define the smoothing by

$$J_k^{\alpha} = \frac{1}{\alpha}\sum_{j=1}^{\alpha} J_{k+j-1}, \quad k = 0, 1, \ldots, \left\lfloor \frac{n}{2} \right\rfloor + 1 - \alpha, \tag{5}$$

where J_k^{α} is the average of α terms in sequence starting at the point J_k, meaning that each point J_k^{α} is the average contribution of α frequencies for the total energy of the series.

Let $\boldsymbol{X}_{i,1}$ and $\boldsymbol{X}_{i,2} \in \mathrm{R}^{\mathrm{P}\times\mathrm{C}}$ EEG samples of two classes from the i-trial with C and P being the number of channels and samples, respectively. The application of the Fourier transform will be in each column (channel) of $\boldsymbol{X}_{i,1}$ and $\boldsymbol{X}_{i,2}$ from the i-trial, building a vector

$$\boldsymbol{J}_{\ell,k}^{\alpha} = (J_{\ell,k_{i,g}}^{\alpha})^{\top}, \tag{6}$$

with $i = 1, \ldots, N_g$, N_g being the number of trials belonging to class g ($g = 1, 2$) and $\ell = 1, \ldots, \mathrm{C}$. These vectors together with the vector of labels $\boldsymbol{y} = (y_1, y_2, \ldots, y_{N_1+N_2})^{\top}$ are the inputs of the classifier WFF-SVM.

3. New Method for EEG Classification

The classification of EEG data is a difficult task, with the analysis disturbed because most of the EEG channels may not be relevant to the classification at hand. Usually, traditional classification techniques alone do not provide good results when applied to EEG data. Therefore, it is important to construct a new method able to distinguish important brain regions and to capture the essential information contained in the data.

3.1 Motivation

The Fourier analysis, especially the periodogram, can reveal hidden patterns in signals. Figure 1 has a set of 4 plots, all of which represent the signals generated by two different stimuli and captured by a channel of the EEG data (red for W_1 class and black for W_{-1} class) for a visual stimuli study (see Section 4.2). The top-left graph represents the superimposed plots of the original EEG signals. Note that it is difficult to visually distinguish two different classes in the time-domain plots presented in this graph. The top-right and bottom-left graphs represent the periodogram and the smoothed periodogram (J_k^4 in Equation (5)) of the signals, respectively. Now, it is easier to notice hidden patterns revealed by the periodograms of the data.

The plots indicate that the periodograms of W_1 have higher values at central frequencies than the periodograms of W_{-1}. In fact, the bottom-right graph in Figure 1 shows a possible discriminant (the dashed line) for these periodograms. Note that the periodograms of W_1 always have values above this hypothetical discrimination line for the central frequencies of the periodogram. However, it should be noted that this type of pattern does not occur for all the channels nor in all regions of the brain. It is necessary to use methods that identify both the relevant channels and the relevant frequencies in a set of periodograms, so that in an application, such as epilepsy detection of signals can be automatically classified into one of the expected classes.

3.2 Calculating the discriminant

The graphs in Figure 1 are revealing. It is easy to discriminate the periodograms for certain frequencies, but this separation is not so clear for other frequencies. It is noticeable that each frequency has its own importance and, therefore, could be evaluated individually and not as a whole. Thus, this paper describes a method in which a different discriminant is calculated for each frequency using the SVM classifier.

Considering the set of training $\boldsymbol{J}_{\ell,k}^{\alpha}$ of Equation (6) and the label vector $\boldsymbol{y}$ with C channels, $\ell = 1, 2, \ldots,$ C and a set of F frequencies, $k = 0, 1, \ldots, F$ (k-th point of the smoothed periodogram and $F = \lfloor n/2 \rfloor$), define $\mathrm{SVM}_{\ell,k}[j_{\ell,k}^{\alpha}]$ as the discriminant function generated by SVM, given by Equation (1), that classifies a new value $j_{\ell,k}^{\alpha}$ of the periodogram for a test signal into one of two classes, W_1 or W_{-1}, according to

$$\mathrm{SVM}_{\ell,k}[j_{\ell,k}^{\alpha}] = \begin{cases} +1, & \text{if } j_{\ell,k}^{\alpha} \text{ is classified in } W_1, \\ -1, & \text{if } j_{\ell,k}^{\alpha} \text{ is classified in } W_{-1}. \end{cases} \tag{7}$$

Then, each discriminant will classify a new signal between two classes depending on whether the periodogram has higher or lower value at a particular frequency. Figure 2 shows an example of these discriminants. Note that each discriminant function $\mathrm{SVM}_{\ell,k}[.]$ could present a different decision. Thus, in order to unify these decisions, the next two sections present a weighting system that generates a single answer to the decision problem.

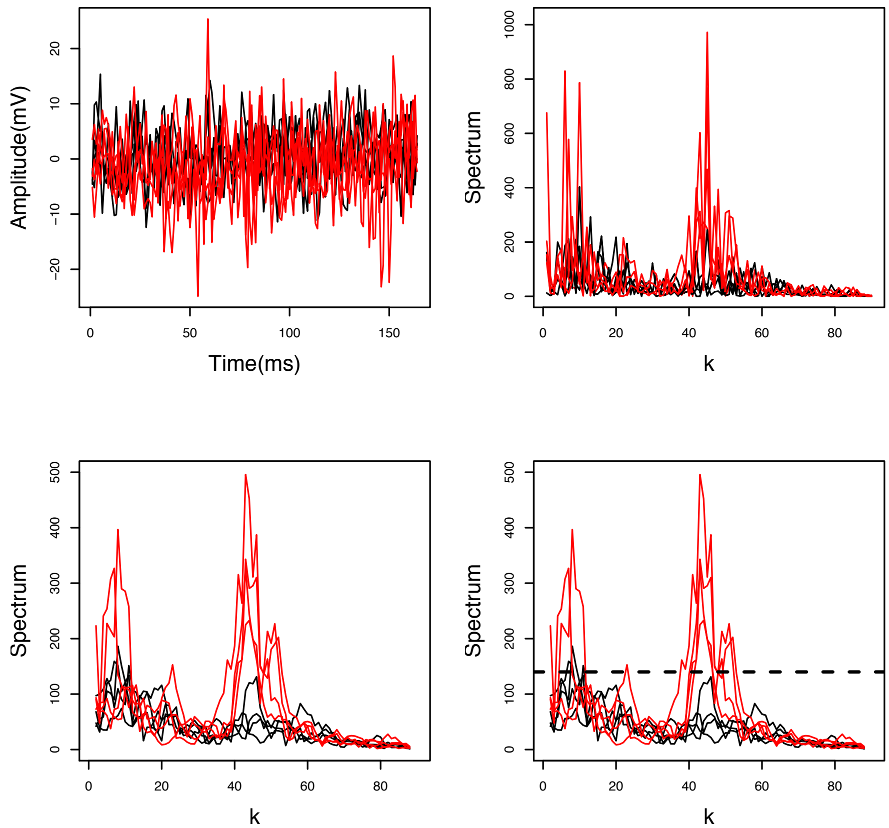

Figure 1. Representations of a set of signals generated by two stimuli. Each line is a signal from the W_1 class (red/lighter lines) or W_{-1} class (black/darker lines). Top-left: original signals. Top-right: periodogram of the signals. Bottom-left: smoothed periodogram of the signals. Bottom-right: smoothed periodogram of the signals with a possible naive discriminant (dashed line). These data are obtained at the Multi-Sensing-Processing and Learning Laboratory (MSPL) at the University of Texas at El Paso (UTEP).

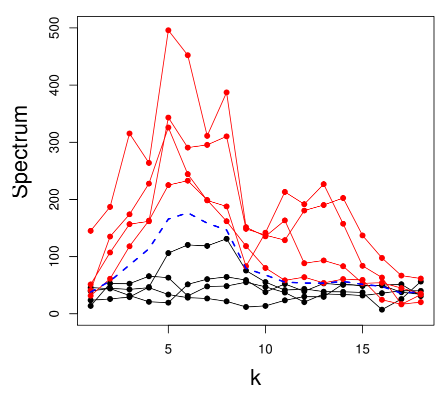

Figure 2. Some discriminating points (dashed line) for some Fourier frequencies ω_k for classes W_1 and W_{-1}. Red (lighter lines) represents class W_1 and black (darker lines) represents class W_{-1}.

3.3 Weighting system

Now, we have several discriminant functions, one for each EEG channel and each point in frequency, with discriminant functions producing different decisions. However, it is clear that there are some discriminants more reliable than others and this reliability is determined by the incorrect classification rate (or error rate) on the training phase of the classification problem. For example, if for some channel ℓ and frequency k the discriminant function $\text{SVM}_{\ell,k}[.]$ provides a low error rate on the training phase, then it is considered more reliable than another discriminant function with a higher error rate. Having this in mind, we introduce a weighting system based on the error rate for each discriminant.

The weight for channel ℓ and frequency k is defined as

$$\Psi_{\ell,k} = [1 - 2 \cdot \min(\text{Error Rate}, 0.5)]^{\hat{\rho}_{\ell,k}}, \tag{8}$$

where Error Rate $\in [0, 1]$ and $\hat{\rho}_{\ell,k} \geq 1$ is a constant given by

$$\hat{\rho}_{\ell,k} = \frac{\text{SS}_{\text{Total}}}{\text{SS}_{\text{Treatment}}}, \tag{9}$$

where $\text{SS}_{\text{Total}} = \sum_{i=1}^{n_c} \sum_{j=1}^{N_i} (J_{i,j}^{\alpha} - \overline{J})^2$ and $\text{SS}_{\text{Treatment}} = \sum_{i=1}^{n_c} \sum_{j=1}^{N_i} (\overline{J}_{i.} - \overline{J})^2$, with n_c representing the number of classes (in this case we have $n_c = 2$), N_i is the number of frequencies of the smoothed periodogram of the i-th class, $J_{i,j}^{\alpha}$ is the j-th smoothed periodogram of the i-th class, $\overline{J}_{i.}$ is the arithmetic mean of the i-th class and $\overline{J}$ is the mean of all smoothed periodograms. The basic concept of our truncated weighting system is to allocate 0 to the ones that have at least a 50% error rate, since $\min\{0, 0.5\} = 0$ implies zero weight. This is so because, based on our experience, it does not make sense to consider classifiers that provide over 50% error rate. On the other hand, the weighting system is an increasing function as the error rate tends to zero, achieving its maximum value when the error rate is zero. Finally, the power $\hat{\rho}_{\ell,k}$ is used to penalize the classifiers that have an error rate between 0 and 50%.

There are several advantages in the use of the exponent $\hat{\rho}_{\ell,k}$ in Equation (9) for the weighting system. It only involves sums, is easy to implement, does not involve optimization, has computational cost almost zero, it uses the data for calculation, it measures the distance between the groups taking into account the variability between and within the groups, and each frequency will have its own weight for SVM.

It is very important to use this kind of information to classify EEG data because much of the data contain non-relevant information of non-activated brain regions such as artifacts in EEG or noise. The next section will show how to use these weights to produce a single decision between one of the two classes W_1 or W_{-1} for new signals.

The implementation of the WFF-SVM method is presented in Algorithm *1*. In Figure 3 we display a flowchart of the SVM framework that summarizes all the steps proposed. This classifier is denominated weighted Fourier and support vector machine (WFF-SVM).

Algorithm *1* Training WFF-SVM algorithm.

1: Let $\boldsymbol{X}_{1,i} \in \text{R}^{\text{P}\times\text{C}}$ and $\boldsymbol{X}_{2,i} \in \text{R}^{\text{P}\times\text{C}}$ denote EEG samples of two classes recorded from the i-th trial. Choose the SVM kernel, the value of c and α smoothing parameter of Equation (5);

2: Apply the Fourier transform of Equation (4) in each column (channel) of $\boldsymbol{X}_{i,1}$ and $\boldsymbol{X}_{i,2}$ from the i-trial and use the moving average technique of Equation (5);

3: Use the SVM in the smoothed periodograms in step 2, totalizing $C \times F$ models;

4: Calculate the training error rate to each model in step 3 and the respective weight of Equation (8).

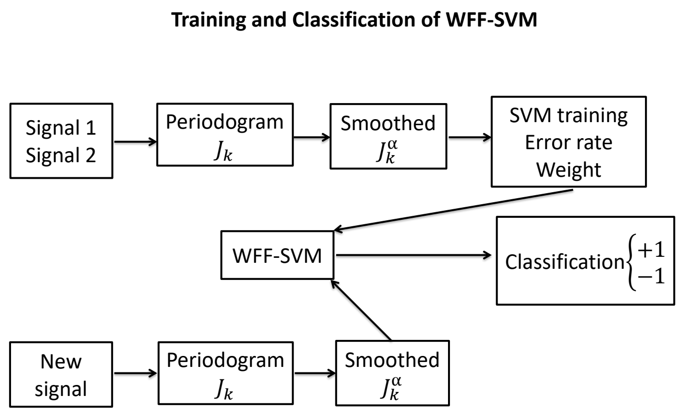

Figure 3. Flowchart for the training and classification phase of a new signal.

3.4 Test phase for practical application

On the test phase for a practical application, we have a new set of signals (one signal per channel) to be classified as class W_1 or W_{-1}. This is done in two different ways (which will be compared later in this paper) using the discriminant function of Equation (7) associated with the weight of Equation (8).

The proposed classification method comprises the following main steps: first, consider a new stimulus $\boldsymbol{X} \in \mathrm{R}^{\mathrm{P}\times\mathrm{C}}$ and for each channel $\ell(\ell = 1, \ldots, \mathrm{C})$ calculate the periodogram $\{J^{\alpha}_{\ell,k}\}$. Then, for each channel ℓ and frequency k of the periodogram use the discriminant function $\mathrm{SVM}_{\ell,k}[J^{\alpha}_{\ell,k}]$ given by Equation (7) to obtain a particular decision (+1 or -1). Finally, using the weights, two decision methods are devised to classify the EEG signals.

In the first decision method, which we label as D_1, each decision $\mathrm{SVM}_{\ell,k}[J^{\alpha}_{\ell,k}]$ is weighted by $\Psi_{\ell,k}$ and each channel has its own decision weighting as in

$$D_1 = \mathrm{sign}\left\{\sum_{\ell=1}^{\mathrm{C}} \mathrm{sign}\left\{\sum_{k=0}^{F} \Psi_{\ell,k} \times \mathrm{SVM}_{\ell,k}[J^{\alpha}_{\ell,k}]\right\}\right\}. \tag{10}$$

In the second decision method, which we label as D_2, each channel has its own decision weighting $\mathrm{SVM}_{\ell,k}[J^{\alpha}_{\ell,k}]$ by $\Psi_{\ell,k}$, and the final decision is a pool between channels. Thus, we define

$$D_2 = \mathrm{sign}\left\{\sum_{\ell=1}^{\mathrm{C}} \frac{\sum\limits_{k=0}^{F} \mathrm{SVM}_{\ell,k}[J^{\alpha}_{\ell,k}] \times \Psi_{l,k}}{\sum\limits_{k=0}^{F} \Psi_{\ell,k}}\right\}. \tag{11}$$

Basically, this decision system takes into account the performance of the channel in the training phase, because if there is a considerable disagreement regarding the classifiers in a given channel, the contribution of this channel to the final classification will not have a great influence. Then, for both decision methods, we apply the criteria

$$\text{Decision} = \begin{cases} W_1, & \text{if } D_j = +1, \\ W_{-1}, & \text{if } D_j = -1, \\ \text{None}, & \text{if } D_j = 0, \end{cases} \tag{12}$$

for $j = 1, 2$. The implementation of the classification of a new signal is presented in Algorithm *2*.

Algorithm *2* Classification of a new signal in WFF-SVM algorithm.

1: Let $\boldsymbol{X}_{\text{new}} \in \mathrm{R}^{\mathrm{P}\times\mathrm{C}}$ denote EEG sample of a new recorded;
2: Apply the Fourier transform of Equation (4) in each column (channel) of $\boldsymbol{X}_{\text{new}}$ and use the moving average technique of Equation (5);
3: Apply the $C \times F$ SVM models calculated by Algorithm *1* in the smoothed periodograms of step 2, totalizing $C \times F$ of values of Equation (7);
4: Use the $C \times F$ values calculated in the step 3 and use the decision weighting of Equations (10) or (11).

The following sections present two applications with real EEG data. First the proposed method is compared to other methods proposed in the literature, then we use it with a new data set.

4. Applications and Results

This section presents two applications of our classification method. The first application uses a publicly available data set described in Andrzejak et al. (2001) which is used in several papers and is very useful to compare the proposed classification method with other methods. The second application uses a data set collected in an experiment conducted by the MSPL at UTEP. The classifier is implemented in the `R` software and to have access to the respective code, visit `https://carvalhomysearches.weebly.com`; see R (2018).

4.1 Epilepsy data classification

The epilepsy data consists of five distinct sets each containing 100 single-channel EEG segments (Andrzejak et al., 2001). Two of these sets, denoted A and B, are obtained from EEG recordings from five healthy volunteers in an awake state with eyes open and eyes closed, respectively. Sets C, D, and E originated from an EEG archive of pre-surgical diagnosis. Segments in set D are recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity (for more details about these data sets see Andrzejak et al. (2001)). As in previous studies (Nigam and Graupe, 2004; Subasi, 2007; Subasi and Gursoy, 2010), we used only two datasets (A and E) to test the classifier.

Both sets A and E have 100 signals each, one signal for each channel and each signal corresponding to 4097 samples. To perform the classification it is cut out the beginning and the end of the signals and subsampled them into 20 signals (components) of 200 samples each. Then, for each set A and E, we randomly selected 10 of the corresponding 20 signals to use in the training phase. In the test phase we repeated this same subsampling process to all the signals in both sets A and E. Thus, it is generated 2000 signals to use in the test phase.

Many authors also proposed methods for the classification of EEG data using data sets A and E to test their classifiers. Table 1 has a summary of the overall results and also the result with the application of the proposed method, named WFF-SVM. In WFF-SVM is used the linear kernel, $c = 1$, $\alpha = 5$ and D_2 as described in Equations (2), (5) and (11), respectively.

According to Zhang et al. (2018), the MKELM is more efficient than the following methods: multilayer perceptron with a single hidden layer; the conventional SVM; SVM with Gaussian and polynomial kernel; multi-kernel SVM using both Gaussian and polynomial kernels; the conventional ELM; ELM with Gaussian kernel; ELM with polynomial kernel, and finally, the multi-kernel ELM using both Gaussian and polynomial kernels. Therefore, we also considered in the comparison the new classifier proposed by Zhang et al. (2018), called MKELM, in both applications.

Table 1. Comparison of results for epilepsy data.

Reference	% Accuracy	Method
Subasi (2007)	94.50	ME
	93.20	MLPNN
Subasi and Gursoy (2010)	98.75	DWT, PCA and SVM
	99.50	DWT, ICA and SVM
	100.00	DWT, LDA and SVM
Jahankhani et al. (2006)	98.00	NN
Guo et al. (2009)	95.00	RWE and NN
Nigam and Graupe (2004)	97.20	NN
Polat and Günes (2007)	98.72	TRF
Li and Wen (2011)	99.90	LS-SVM
Chandaka et al. (2009)	95.96	SVM
Übeyli (2010)	99.56	LS-SVM
Zhang et al. (2018)	100.00	MKELM
Proposed method	100.00	WFF-SVM

where ME is mixed of experts; MLPNN is multi-layer perceptron neural network; DWT is discrete wavelet transform; LDA is linear discriminant analysis; ICA is independent component analysis; NN is neural networks; RWE is relative wavelet energy; LS-SVM is least square support vector machine; MKELM is multi-kernel extreme learning machine using both Gaussian and polynomial kernels with CSP feature.

Note that the proposed method is as efficient as (or more efficient than) the other methods. A possible reason for this improvement is the weighting system capturing the most important regions for classification, strengthening the process.

Despite the greater efficiency of the proposed method, it can be noted that all methods are very efficient for this problem. The main reason for this result is that it is relatively easy to classify the epilepsy data; in fact, neurologists can visually distinguish the EEG patterns of epileptic patients and non-epileptics patients. For this reason, the following example presents a more complex application that uses EEG data collected in an experiment based on visual stimuli with a set of tasks to classify.

4.2 Classification of visual stimuli

In the visual stimuli application, the objective is to calculate the discriminant function so that, given a new visual stimulus event, our classification method is capable of identifying the slide presented to the subject from the EEG data recordings only. To do this, the proposed method is used after a selection of activated channels using PCA.

Experimental Design The data set used in this application is acquired at the MSPL at UTEP. The EEG data are recorded from a volunteer test subject using a Biosemi EEG acquisition system with 128 channels. The acquisition system recorded EEG signals corresponding to 10 different visual stimuli, each one presented multiple times in random order and during a regular interval of time. The visual stimuli used correspond to the slides shown in Figure 4. Each stimulus is shown on a computer monitor screen 4 times (in random order) with a five seconds break between each slide, corresponding to a blank screen. An audible tone alerted the subject each time a new slide is about to be displayed. Thus, the EEG data set of the second experiment comprised of 4 EEG signals for each one of the 10 visual stimuli, acquired by 128 channels.

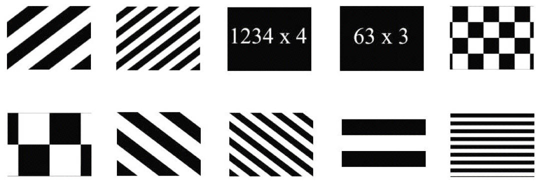

Figure 4. Ten visual stimuli shown to the test subject during EEG signal acquisition.

Using PCA for source localization The PCA is used to explain the variance-covariance structure of a set of variables by a smaller set of variables formed by linear combinations of the original ones (Johnson and Wichern, 2007). Generally, in databases that contain strongly correlated variables (as in EEG data) the PCA is very useful to reduce the dimensionality of the problem. In PCA, the first principal component is the linear combination with the highest possible variance. This means, in the case of EEG data, that the most important channels for the composition of the first principal component are the channels that capture signals with higher variance (the channels corresponding to the activated brain regions) as described in von Borries et al. (2013). Figure 5 shows contours obtained for the first principal component when PCA is applied to EEG signals from 128 channels of the visual stimuli experiment. One can observe that most of the variability in this experiment is present in the channels located on the brain's frontal lobe. The next sections show that, in fact, this region is the most important for classification and the other regions basically do not bring relevant information to the classification problem at hand. Actually, our results show that the correct classification rate increases when the signals from those regions are not included in classification.

Data analysis First, we train the classifier. Since the proposed method is a binary classifier and we have 10 apparently different visual stimuli, the classification process is implement sequentially by pairs of visual stimuli. Moreover, as many images are very similar, the classification is performed only with abstract images against images with arithmetic operations, making a total of 16 discriminants (or 16 pairs). Cross-validation is used to approximate the correct classification rate of this method, as follows: for each pair of images analyzed, the first repetition of each image (independent of the others) is excluded in the training phase to be used in the testing phase. Then, the second repetition of each image (independent of the others) is excluded in the training phase to be used in the testing phase, and so on. Thus, $4 \times 16 \times 2 = 128$ signals are used in the test phase. Note that the signals used in the test phase are not used to build the discriminant, resulting in a reliable analysis. The first test is done using the periodogram with the configurations $\alpha = 1$ and 4, linear kernel and using $c = 1$. Note in Table 2, the classification rates for each configuration. There is an increase of around 10% for all configurations when the smoothed periodogram ($\alpha = 4$) is used, indicating that smoothing is a good option to improve the classification rate. Furthermore, D_1 method is better than the D_2, but not having a very large difference between the rates. Figure 6 shows a contour plot of the accuracy of each brain region. It should be noted that the EEG signals located at the brain's frontal lobe had the best correct classification rates. The similarity between Figures 5 and 6 is remarkable, indicating that the regions identified using PCA actually correspond to the regions of higher correct classification rates. Therefore, one might think that the non-activated regions contain non-relevant information that actually disturbs the classification. Thus, the cross-validation process is repeated using 53 channels with the highest hit rates, where most are from the front of the brain, with parameters $c = 1, 10, 100$. The results presented in Table 3 indicate that the correct classification rates increase when using the smoothed periodograms and specially when selecting only the most relevant channels. Therefore, it appears to be extremely important, in a classification analysis of EEG data, to remove from the analysis the channels that appear basically to capture non-relevant information. However, the cost value does not seem to influence much on the results and the classification rates are very similar for all values of c, so, for the analyzes that will be done from now on, will be used $c = 1$.

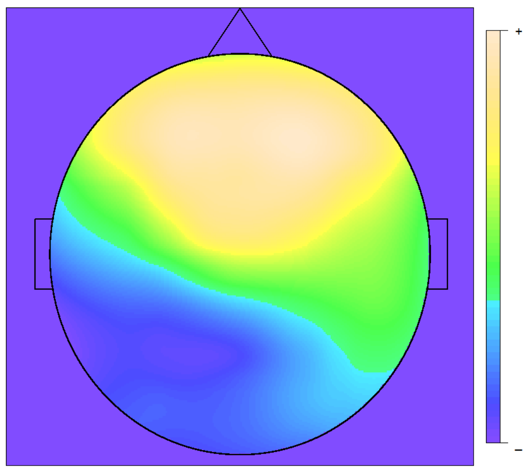

Figure 5. Variability of signals through the Brain. Contours for the first principal component when PCA is applied to EEG signals from 128 channels of the visual stimuli experiment. The front of the brain presents most of the signal variability.

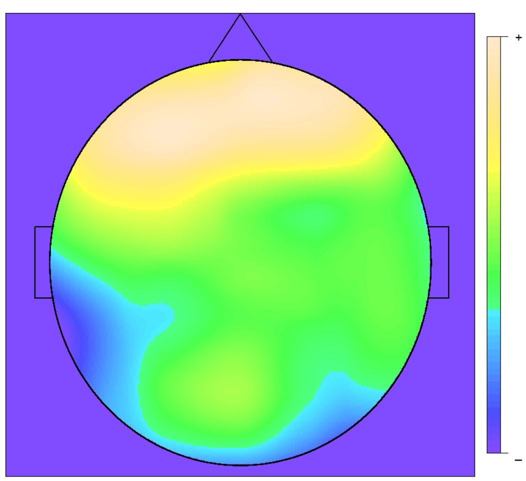

Figure 6. Contour lines for the correct classification rates by channel: new method with the smoothed periodogram, $\alpha = 4$ and $c = 1$.

Table 2. Accuracy for some settings of smoothing parameter in the WFF-SVM algorithm.

Classifier	α	Method	% accuracy
WFF-SVM	1	D_1	75.00
		D_2	73.44
	4	D_1	87.50
		D_2	84.37

Table 3. Results using $\alpha = 4$ for the accuracy using some values of cost (c), number of channels and type of decision.

Cost	Channel	Method	% accuracy
$c = 1$	128	D_1	87.50
		D_2	84.37
	53	D_1	92.97
		D_2	92.97
$c = 10$	128	D_1	85.94
		D_2	85.16
	53	D_1	92.97
		D_2	92.97
$c = 100$	128	D_1	85.95
		D_2	85.16
	53	D_1	92.97
		D_2	92.97

After some α variations, we obtained a classification rate of 95.31% with $c = 1$, $\alpha = 5$, using D_2 with 53 channels, and 73.44% to MKELM using all the channels with CSP feature. These are the best results found in this study. The non-requirement of an extensive training data set constitutes an important characteristic of the proposed classification method since in real-world applications the collection of signals available to train the classifier can be limited to only a few cases.

5. Discussion and Conclusions

EEG technique is employed to help in a variety of diagnosis, such as posttraumatic stress, human emotions and epilepsy. Regarding the latter one, there is a special interest to detect as early as possible epilepsy in order to initiate the proper treatment and mitigate this neurological disorder effects. Several studies were conducted with this objective, such as Fergus et al. (2015) who uses machine learning, whereas Thodoroff et al. (2016) and Acharya et al. (2018) have used the deep learning (DL) approach. The DL method has been used in several problems as in image recognition (Krizhevsky et al., 2012), diagnosis of Alzheimer's disease (Ortiz et al., 2016), prediction of sale prices of real estate units (Rafiei and Adeli, 2015) and in the estimation of concrete compressive strength Rafiei et al. (2017). There are examples in the literature that use SVM and DL, such as in Tang (2013), who developed an approach in DL replacing the softmax layer by a linear SVM. Erfani et al. (2016) used a hybrid model where an unsupervised deep belief networks is trained to extract generic underlying features, and one class SVM is trained from the features learned by the deep belief networks. Therefore, these works show that the use of SVM in DL is not new and suggests that in future works WFF-SVM in DL can also be contemplated in order to search for more efficient methods. The WFF-SVM can be used in any type of signal, EEG, electrocardiogram, electromyogram, etc. In order to accomplish that, it is sufficient to represent the data as a time series or in a certain proper order. This proposed paper in based on a broader study found in Carvalho (2016), in which electromyogram data were also considered. Furthermore, this classifier can be used in clinical application or any other application. Regarding the computational intensive aspect, with the rapidly increasing performance of new computers, including parallel programing and the promising quantum programming the tendency is to be feasible. The application using epilepsy data showed that the proposed method has no better competitor among other methods presented in the literature. This paper presents a second and more complete application. This application using EEG data captured during an experiment involving visual stimuli showed a number of specific features for the classification of EEG data. In particular, this application showed that the brain region identified using PCA was similar to the region where the channels had the best individual correct classification rates. In fact, the correct classification rate increased significantly by discarding the EEG channels that had non-relevant information. The proposed method of using smoothed periodograms and assigning weights to the channels based on their individual error rates resulted in higher correct classification rates than other methods reported in the literature. It should be noted that the proposed method showed a high correct classification rate of 95.31% using only three signals from each class in the training phase. Thus, a topic for future research is to extend the WFF-SVM to accept more than two groups for training and classification. In addition, it would be important to propose some sort of threshold for decision-making, in guiding the decision Equation (12) on how far it must be from zero to have a more objective classification.

This paper presented a new method for classification of EEG data that uses Fourier analysis and SVM. The proposed method employs a specific SVM decision value for each frequency of the periodogram. In addition, a simple weighting system based on the performance of the classifier, obtained in the training phase, is applied to the classification phase. We used two data sets to test the performance of the proposed classifier. The first data set referred to EEG of an epilepsy study and the second to EEG of a visual stimulation study. Finally, one point for improvement include the extension of our classification method to more than two classes and to expand the performance comparison with other methods.

Acknowledgment

The authors thank and acknowledge financial support from, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Universidade Federal do Rio Grande do Norte (UFRN), Universidade de Brasília (UnB) and University of Texas at El Paso (UTEP). We also thank the Department of Statistics at UFRN for inviting the fourth author to a visit that generate a partnership that led to this research. Furthermore, we thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

References

Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., and Adeli, H., 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers in Biology and Medicine, 100, 270-278.

Acharya, U.R., Sree, S.V., Alvin, A.P.C., and Suri, J.S., 2012. Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework. Expert Systems with Applications, 39, 9072-9078.

Acharya, U.R., Yanti, R., Zheng, J.W., Krishnan, M.M.R., Tan, J.H., Martis, R.J., and

Lim, C.M., 2013. Automated diagnosis of epilepsy using CWT, HOS and texture parameters. International Journal of Neural Systems, 23, 1350009.

Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C.E., 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Physical Review E, 64, 061907.

Brockwell, P.J. and Davis, R.A., 2002. Introduction to Time Series and Forecasting. Springer, New York.

Bronzino, J.D., 1999. Biomedical engineering handbook. CRC press, Boca Raton, US.

Buzsaki, G., 2006. Rhythms of the Brain. Oxford University Press, Oxford, UK.

Carvalho, J.B., 2016. Classificador Máquina de Suporte Vetorial com Análise de Fourier Aplicada em Dados de EEG e EMG. Master's thesis. Federal University of Rio Grande do Norte, available at https://repositorio.ufrn.br/jspui/handle/123456789/20964.

Chandaka, S., Chatterjee, A., and Munshi, S., 2009. Cross-correlation aided support vector machine classifier for classification of EEG signals. Expert Systems with Applications, 36, 1329-1336.

Chiang, L.H., Russell, E.L., and Braatz, R.D., 2000. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Chemometrics and Intelligent Laboratory Systems, 50, 243-252.

Chiappa, S., and Bengio, S., 2004. HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems. IDIAP.

Cortes, C., and Vapnik, V., 1995. Support-vector networks. Machine Learning, 20, 273-297.

Das, K., Zhang, S., Giesbrecht, B., and Eckstein, M.P., 2009. Using rapid visually evoked EEG activity for person identification. In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 2490-2493). IEEE.

Ding, C. and He, X., 2004. K-means clustering via principal component analysis. In Proceedings of the Twenty-first International Conference on Machine Learning (p. 29). ACM.

Erfani, S.M., Rajasegarar, S., Karunasekera, S., and Leckie, C., 2016. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recognition, 58, 121-134.

Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., and Hamdan, H., 2015. A machine learning system for automated whole-brain seizure detection. Applied Computing and Informatics, 12, 70-89.

Friedman, H.P., and Rubin, J., 1967. On some invariant criteria for grouping data. Journal of the American Statistical Association, 62, 1159-1178.

Fuller, W.A., 1996. Introduction to Statistical Time Series. Wiley, New York

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M., and Haussler, D., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics, 16, 906-914.

Giampaoli, V., Tamura, K.A., Caro, N.P., and de Araujo, L.J.S., 2016. Prediction of a financial crisis in Latin American companies using the mixed logistic regression model. Chilean Journal of Statistics, 7, 31-41.

Guo, L., Rivero, D., Seoane, J.A., and Pazos, A., 2009. Classification of EEG signals using relative wavelet energy and artificial neural networks. In Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation (pp. 177-184). ACM.

Hastie, T., Tibshirani, R, and Friedman, J., 2008. The Elements of Statistical Learning. Springer, New York.

Hornik, K., Meyer, D., and Karatzoglou, A., 2006. Support vector machines in R. Journal of statistical software, 15, 1-28.

Jahankhani, P., Kodogiannis, V., and Revett, K., 2006. EEG signal classification using wavelet feature extraction and neural networks. In Modern Computing, 2006. JVA'06. IEEE John Vincent Atanasoff 2006 International Symposium on (pp. 120-124). IEEE.

Jiao, Y., Zhang, Y., Wang, Y., Wang, B., Jin, J., and Wang, X., 2018. A novel multilayer correlation maximization model for improving CCA-based frequency recognition in SSVEP braincomputer interface. International Journal of Neural Systems, 28, 1750039.

Johnson, R.A. and Wichern D.W., 2007. Applied Multivariate Statistical Analysis. Prentice-Hall, New Jersey.

Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., and Ritter, H., 2004. BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. IEEE Transactions on Biomedical Engineering, 51, 1073-1076.

Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (pp. 1097-1105).

Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J.R., and Birbaumer, N., 2001. Brain-computer communication: Unlocking the locked in. Psychological Bulletin, 127, 358.

Li, Y. and Wen, P.P., 2011. Clustering technique-based least square support vector machine for EEG signal classification. Computer Methods and Programs in Biomedicine, 104, 358-372.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B., 2007. A review of classification algorithms for EEG-based braincomputer interfaces. Journal of Neural Engineering, 4, R1.

Nigam, V.P., and Graupe, D., 2004. A neural-network-based detection of epilepsy. Neurological Research, 26, 55-60.

Ocak, H., 2009. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. Expert Systems with Applications, 36, 2027-2036.

Ortiz, A., Munilla, J., Gorriz, J.M., and Ramirez, J., 2016. Ensembles of deep learning architectures for the early diagnosis of the Alzheimers disease. International Journal of Neural Systems, 26, 1650025.

Penny, W.D., Roberts, S.J., Curran, E.A., and Stokes, M.J., 2000. EEG-based communication: a pattern recognition approach. IEEE Transactions on Rehabilitation Engineering, 8, 214-215.

Pfurtscheller G., Neuper C., Schlogl A., and Lugger K., 1998. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. IEEE Transactions on Rehabilitation Engineering, 6, 316-325.

Polat, K. and Günes, S., 2007. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Applied Mathematics and Computation, 187, 1017-1026.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rafiei, M.H. and Adeli, H., 2015. A novel machine learning model for estimation of sale prices of real estate units. Journal of Construction Engineering and Management, 142, 04015066.

Rafiei, M.H., Khushefati, W.H., Demirboga, R., and Adeli, H., 2017. Supervised deep restricted Boltzmann machine for estimation of concrete. ACI Materials Journal, 114, 237-244.

Rubin, J., 1967. Optimal classification into groups: an approach for solving the taxonomy problem. Journal of Theoretical Biology, 15, 103-144.

Ruiz, S., Buyukturkoglu, K., Rana, M., Birbaumer, N., and Sitaram, R., 2014. Real-time fMRI brain computer interfaces: self-regulation of single brain regions to networks. Biological Psychology, 95, 4-20.

Sharma, M., Pachori, R.B., and Acharya, U.R., 2017. A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. Pattern Recognition Letters, 94, 172-179.

Sitaram, R., Lee, S., Ruiz, S., Rana, M., Veit, R., and Birbaumer, N., 2011. Real-time support vector classification and feedback of multiple emotional brain states. Neuroimage, 56, 753-765.

Souza, M.P.S., Filho, T.M.S., Amaral, G.J.A., and Souza, R.M.C.R., 2019. Investigating different fitness criteria for swarm-based clustering. International Journal of Business Intelligence and Data Mining, pages in press, available at http://dx.doi.org/10.1504/IJBIDM.2017.10007517.

Subasi, A., 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Systems with Applications, 32, 1084-1093.

Subasi, A. and Gursoy, M.I., 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications, 37, 8659-8666.

Tang, Y., 2013. Deep learning using linear support vector machines. Working paper, available at https://arxiv.org/abs/1306.0239.

Theodoridis, S. and Koutroumbas, K., 2008. Pattern Recognition. Academic Press, New York.

Thodoroff, P., Pineau, J., and Lim, A., 2016. Learning robust features using deep learning for automatic seizure detection. In Machine Learning for Healthcare Conference (pp. 178-190).

Thomas, K.P. and Vinod, A.P., 2017. Toward EEG-based biometric systems: the great potential of brain-wave-based biometrics. IEEE Systems, Man, and Cybernetics Magazine, 3, 6-15.

Übeyli, E.D., 2010. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. Expert Systems with Applications, 37, 233-239.

Vapnik, V., 1996. Statistical Learning Theory. Wiley, New York.

von Borries, G.F., Rocha, L.C.S., Coutinho, M.S., and von Borries, R.F., 2013. Identification of brain-generated activity with EEG using visual stimuli. Revista Brasileira de Biometria, 31, 28-42.

Wang, H., Zhang, Y., Waytowich, N.R., Krusienski, D.J., Zhou, G., Jin, J., Wang, X., and Cichocki, A., 2016. Discriminative feature extraction via multivariate linear regression for SSVEP-based BCI. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24, 532-541.

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

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M., 2002. Brain-computer interfaces for communication and control. Clinical Neurophysiology, 113, 767-791.

Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., and Cichocki, A., 2018. Multikernel extreme learning machine for EEG classification in brain-computer interfaces. Expert Systems with Applications, 96, 302-310.

Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., and Cichocki, A., 2016. Sparse Bayesian classification of EEG for braincomputer interface. IEEE Transactions on Neural Networks and Learning Systems, 27, 2256-2267.

Zuquete, A., Quintela, B., and Cunha, J.P.S., 2010. Biometric authentication using electroencephalograms: a practical study using visual evoked potentials. Electrónica e Telecomunicações, 5, 185-194.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to Víctor Leiva, Editor-in-Chief of the ChJS (E-mail: chilean.journal.of.statistics@gmail.com). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at http://chjs.mat.utfsm.cl/files/ChJS.zip. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at http://www.ams.org/mathscinet/msc/. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., (1994). Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., (1997). Local influence. In Kotz, S., Read, C.B. and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Update, Vol. 1, Wiley, New York, pp. 380-385.

Rukhin, A.L., (2009). Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters 79, 1004-1007.

Stein, M.L., (1999). Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Pena, D., and Pankratz, A.E., (2000). Outliers in multivariate time series. Biometrika 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright