Computers in Biology and Medicine 40 (2010) 705–714

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Tackling EEG signal classification with least squares support vector machines: A sensitivity analysis study

Clodoaldo A.M. Lima [a,1], André L.V. Coelho [b,*], Marcio Eisencraft [c]

[a] Graduate Program in Electrical Engineering, School of Engineering, Mackenzie Presbyterian University, Rua da Consolação, 896, 01302-907 São Paulo, SP, Brazil
[b] Graduate Program in Applied Informatics, Center of Technological Sciences, University of Fortaleza, Av. Washington Soares, 1321/J30, 60811-905 Fortaleza, CE, Brazil
[c] Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas, Universidade Federal do ABC, Rua Santa Adélia, 166, 09210-170 Santo André, SP, Brazil

ARTICLE INFO

Article history:
Received 1 September 2009
Accepted 11 June 2010

Keywords:
Least squares support vector machines
Epilepsy
EEG signal classification
Sensitivity analysis
Kernel functions

ABSTRACT

The electroencephalogram (EEG) signal captures the electrical activity of the brain and is an important source of information for studying neurological disorders. The proper analysis of this biological signal plays an important role in the domain of brain–computer interface, which aims at the construction of communication channels between human brain and computers. In this paper, we investigate the application of least squares support vector machines (LS-SVM) to the task of epilepsy diagnosis through automatic EEG signal classification. More specifically, we present a sensitivity analysis study by means of which the performance levels exhibited by standard and least squares SVM classifiers are contrasted, taking into account the setting of the kernel function and of its parameter value. Results of experiments conducted over different types of features extracted from a benchmark EEG signal dataset evidence that the sensitivity profiles of the kernel machines are qualitatively similar, both showing notable performance in terms of accuracy and generalization. In addition, the performance accomplished by optimally configured LS-SVM models is also quantitatively contrasted with that obtained by related approaches for the same dataset.

1. Introduction

In the last decades, the electroencephalogram (EEG) signal, which represents the electrical activity of the brain, has been intensively studied. This is because it can convey valuable clinical information about the current neurological conditions of patients, being widely used in the study of the nervous system properties, for monitoring sleep stages, and for the diagnosis of many disorders such as epilepsy, sleep disorders, and dementia [1,2]. Moreover, the analysis and processing of this type of biological signal has played an important role in the domain of brain–computer interface [3], which aims at setting up communication channels between human brain and computers.

Temporary electrical disturbance of the brain can cause epileptic seizures. Sometimes seizures may go unnoticed, depending on their strength, and sometimes may be confused with other events, such as strokes, which can also cause falls or migraines. Unfortunately, the occurrence of an epileptic seizure seems unpredictable and its course of action is still very little understood [4]. So, more research is needed for a better understanding of the mechanisms causing epileptic disorders.

Despite rapid advances of neuroimaging techniques, EEG recordings continue to play an important role in both the diagnosis of neurological diseases and the understanding of psychophysiological processes. In order to extract relevant information from recordings of brain electrical activity, a variety of computerized-analysis methods have been developed. Most of them assume that the EEG signal is generated by a highly complex linear system, which results in characteristic properties like non-stationarity and difficulty of prediction [5].

Recently, there has been a growing interest in applying techniques from the domains of nonlinear analysis and chaos theory for studying the behavior of experimental time series such as EEG signals [5–9]. Moreover, many nonlinear classification methods have been proposed. Among them, we can mention artificial neural networks [4,10–19] and support vector machines (SVM), either for two-class [3] or multiclass [20,21] EEG signal discrimination.

In particular, the application of SVM is justified for this type of machine learning (ML) technique has shown to be quite successful in coping with a number of complex data analysis problems. The SVM approach is based on the structural risk minimization principle, which asserts that the generalization error is delimited by the sum of the training error and a parcel that depends on the Vapnik–Chervonenkis dimension [22,23]. By

* Corresponding author. Tel.: +55 11 2114 8833; fax: +55 11 2114 8600.
E-mail addresses: moraes@mackenzie.br (C.A.M. Lima), acoelho@unifor.br (A.L.V. Coelho), marcio.eisencraft@ufabc.edu.br (M. Eisencraft).

[1] Post-publication contact author.

0010-4825/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2010.06.005

minimizing this summation, high generalization performance can be obtained. Besides, the number of free parameters in SVM does not explicitly depend upon the input dimensionality of the problem at hand. Another important feature of the support vector learning approach is that the underlying optimization problems are inherently convex and have no local minima, which comes as the result of applying Mercer's conditions on the characterization of kernels [24].

In the literature, several derived formulations have been proposed for SVM, seeking for advantages in terms of effectiveness or efficiency criteria. In particular, Suykens and Vandewalle [25], and Suykens et al. [26] have introduced the least squares SVM (LS-SVM) classifier by modifying the standard formulation so as to obtain a system of linear equations in the dual space. This is done by taking a least squares cost function, with equality instead of inequality constraints. Despite the fact that LS-SVM have gained increased attention recently [27], our perception is that their application to the automatic analysis of nonlinear biomedical signals has not been systematically investigated yet, with the work of Kemal et al. [28] over electrocardiogram (ECG) data being one of the first in this context.

It is important to stress that, although considered as high-performance models, the efficiency and effectiveness underlying the induction of SVM and LS-SVM depend critically on the appropriate selection of values for some important hyperparameters [29]. The fact is that a bad specification of such parameters may jeopardize the applicability of these kernel machines. Some works have already provided solutions to this problem, ranging from those based on cross-validated model selection to those based on extensive grid search or heuristic optimization rules [30,31].

In this paper, we present a thorough analysis regarding the impact of the choice of the kernel parameter value on the performance exhibited by LS-SVM and SVM classifiers induced for EEG signal discrimination. Initially, we preprocess the dataset of EEG signals [6,7] by extracting wavelet coefficients [10,13,16] and then compute statistical metrics over the resultant data to create the feature vectors. Such vectors serve as input to the kernel machine (either SVM or LS-SVM), which provides the final epilepsy/non-epilepsy decision. Several detailed graphs (referred to as sensitivity profiles) are presented here enabling one to visually inspect, for each combination of kernel type, kernel parameter value and type of feature extracted, the accuracy/generalization levels achieved by the machines in terms of misclassification rate as well as sensitivity and specificity indices [13]. In addition, given that the EEG dataset used in this paper has also been explored by many researchers working in the biomedical signal processing field [4,12–21], the classification performance achieved by optimally configured LS-SVM models is also quantitatively contrasted with that produced by related approaches as reported in the literature.

The remaining parts of the paper are organized as follows. In the next section, we describe the EEG benchmark data analyzed and the techniques used to preprocess it. Moreover, we present the mathematical formulations behind SVM and LS-SVM as well as comment upon the importance underlying the choice of kernel functions and parameters. In Section 3, we discuss the empirical results achieved, showing the sensitivity profiles exhibited by the machines and also providing a quantitative contrast, in terms of accuracy, with related work. Finally, Section 4 concludes the paper.

2. Materials and methods

In this section, the sets of EEG signals used in the experiments are described. Also, spectral analysis of the EEG signals using discrete wavelet transform (DWT) is explained and the statistical features effectively extracted are presented. In a third part, the mathematical formulations underlying the kernel machines considered are given.

2.1. Dataset characterization

In this work, we have used the EEG data made publicly available by Andrzejak et al. [6,7]. The complete dataset involves five sets (denoted A–E), each containing 100 single-channel EEG segments. All EEG signals from this dataset were recorded with the same amplifier system, using an average common reference. The data were digitized at 173.61 samples/s using 12 bit resolution. Bandpass filter settings were 0.53–40 Hz (12 dB/oct). Since in the experiments reported in Section 3 the kernel machines investigated only discriminate between samples from sets A and E, we focus on the description of these two sets in the sequel. The reader should refer to Refs. [6,7] for further details on the data acquisition process.

Set A consists of segments taken from signals recorded extracranially during the relaxed state of healthy subjects with eyes open. That is, surface EEG recordings were carried out on five healthy volunteers using a standardized electrode placement scheme, namely the International 10–20 system [2]. Then, 100 segments were selected and cut out from continuous multi-channel EEG recordings (i.e., from all 20 electrodes used—refer to Fig. 1 of Ref. [6] for the anatomical disposition of these electrodes over the scalp) after visual inspection for artifacts, due, for example, to muscle activity or eye movements.

In contrast, set E originated from an EEG archive of pre-surgical diagnosis. That is, EEG time series recorded intracranially from five patients were selected, all of whom had achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the seizure generating area. In this context, the implantation of electrodes was carried out to exactly localize this area, termed as the epileptogenic zone. The 100 specific segments that compose set E were selected from all recording sites exhibiting ictal activity (i.e., actual epileptic seizures) [6].

2.2. Data pre-processing

Wavelet transform is a spectral estimation technique in which any general function can be expressed as an infinite series of wavelets [4,10]. The basic idea underlying wavelet analysis consists of expressing a signal as a linear combination of a particular set of functions (wavelet transform), obtained by shifting and dilating one single function called a mother wavelet. The decomposition of the signal leads to a set of coefficients called wavelet coefficients. Therefore, the signal can be reconstructed as a linear combination of the wavelet functions weighted by the wavelet coefficients. In order to obtain an exact reconstruction of the signal, an adequate number of coefficients must be computed [13,16].

The key feature of wavelets is the time–frequency localization. It means that most of the energy of the wavelet is restricted to a finite time interval. Frequency localization means that the Fourier transform is band limited. When compared to short-time Fourier transform, the advantage of time–frequency localization is that wavelet analysis varies the time–frequency aspect ratio, producing good frequency localization at low frequencies (long time windows), and good time localization at high frequencies (short time windows). This produces a segmentation or tiling of the time–frequency plane that is appropriate for most physical signals, especially those of a transient nature. The wavelet

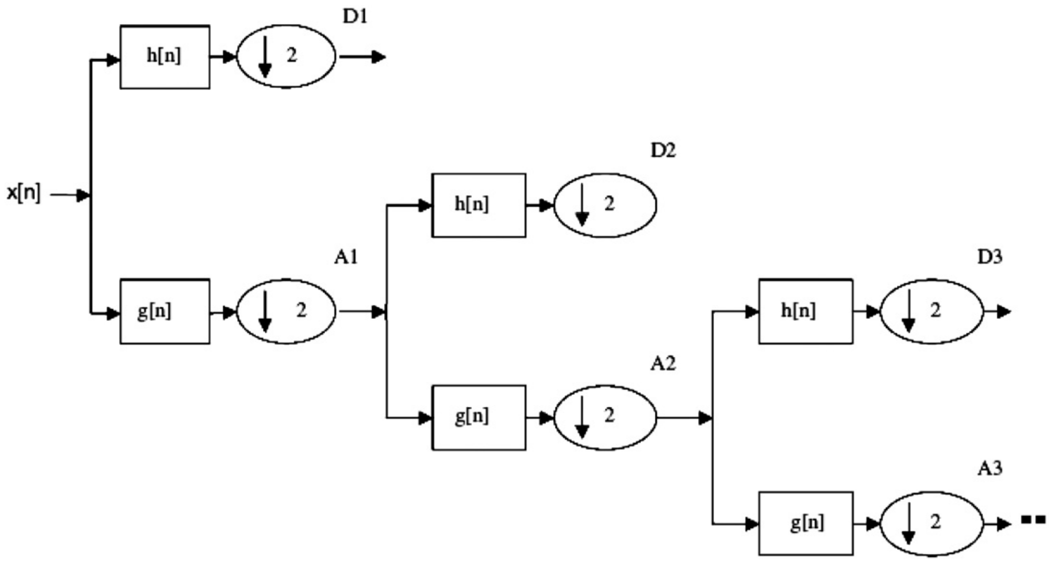

Fig. 1. Sub-band decomposition of DWT implementation; $h[n]$ is the high-pass filter, $g[n]$ the low-pass filter (taken from [12]).

technique applied to the EEG signal will reveal features related to the transient nature of the signal which are not obvious by the Fourier transform [12].

The selection of a suitable wavelet and the number of decomposition levels is very important in the analysis of signals using the DWT. The number of decomposition levels is usually chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies necessary for the classification of the signal are retained in the wavelet coefficients. In the present study, since the EEG signals do not have any useful frequency components above 30 Hz, the number of decomposition levels was chosen to be 5. Thus, the EEG signals were decomposed into details D1–D5 and one final approximation, A5, as in [12]. Fig. 1 shows the procedure of multi-resolution decomposition of a given signal $x[n]$.

In the experiments reported in Section 4, we have made use of Daubechies of order 4 (db4) as wavelet basis, as its smoothing feature made it suitable to detect changes of EEG signal [12]. In order to reduce the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients were used to generate the input to the SVM and LS-SVM [9,16,19]:

- Average of wavelet coefficients in each sub-band (W_Avg).
- Standard deviation of wavelet coefficients in each sub-band (W_Std).
- Maximum of wavelet coefficients in each sub-band (W_Max).

2.3. Standard SVM and LS-SVM classifiers

Let the EEG training set in hand be $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, with input $\mathbf{x}_i \in \Re^m$ and $y_i \in \{\pm 1\}$. Either SVM or LS-SVM first accomplishes a mapping $\phi: \Re^m \to \Re^n$. Usually, n is much higher than m in such a way that the input vector is mapped into a high-dimensional space [24]. When data are linearly separable, the machine builds a hyperplane $\mathbf{w}^T\phi(\mathbf{x})+b$ in $\Re^n$, by means of which the margin between positive and negative samples is maximized. It can be shown that $\mathbf{w}$, for this optimal hyperplane, can be defined as a linear combination of the set of nonlinear data transformations, that is $\mathbf{w}=\sum_{i=1}^{N}\alpha_i y_i \phi(\mathbf{x}_i)$ [22].

In the standard SVM formulation [23], the generalized optimal separating hyperplane $\mathbf{w}$ is determined by minimizing the functional:

$$\min_{w,b,\xi_i} J(w, b, \xi_i) = \frac{1}{2}(\mathbf{w}^T\mathbf{w}) + C\sum_{i=1}^{N}\xi_i, \quad (1)$$

where C is a regularization hyperparameter and ξ_i, $i=1,\ldots,N$, are slack variables measuring the difference (error) between y_i and the actual SVM output. The optimization of (1) is subject to the constraints

$$y_i[\mathbf{w}^T\phi(\mathbf{x}_i)+b] \geq 1-\xi_i, \quad i=1,\ldots,N. \quad (2)$$

From (1) and (2), the resulting quadratic programming (QP) problem in the dual space may be written as [23]

$$\max_{\alpha} J(\alpha) = \max_{\alpha}\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \quad (3)$$

subject to $\sum_{i=1}^{N}\alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$, for $i=1,\ldots,N$. To obtain $\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ in (3), we do not need to calculate either $\phi(\mathrm{x}_i)$ or $\phi(\mathrm{x}_j)$ explicitly. Instead, for some ϕ, we can design a kernel matrix K such that $K(\mathbf{x}_i,\mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ [24].

Kernels are exploited for the purpose of (non)linearly mapping input data into high-dimensional feature spaces in a computationally efficient manner. It is within this novel, hidden feature space that the linear decision surface can be readily designed. Different kernel functions give origin to different feature spaces and thus also to different generalization capabilities of the resulting classifier. So, one important issue to be considered in the design of SVMs in general is how to choose the best kernel function for dealing with the nuances of the given problem.

Among the several types of kernels one can experiment with [24], the results shown in the next section were obtained with either RBF (radial basis function) or ERBF (exponential RBF) kernels, whose mathematical expressions are shown below:

$$K_{\mathrm{RBF}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\sigma^2}\right), \quad (4)$$

$$K_{\mathrm{ERBF}}(\mathbf{x}_i,\mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|}{2\sigma^2}\right), \quad (5)$$

where σ denotes a radius parameter to be adjusted previously.

By resorting to kernels, (3) can be rewritten as

$$\max_{\alpha} J(\alpha) = \max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j). \quad (6)$$

For the training samples along the decision boundary, the corresponding α_i's are greater than zero, as ascertained by the Kuhn–Tucker theorem [22]. These samples are known as support vectors. The number of support vectors is generally much smaller than N, being proportional to the generalization error of the classifier [23]. A test vector $\mathbf{x} \in \Re^m$ is then assigned to a given class according to $f(\mathbf{x}) = \mathrm{sign}[\mathbf{w}^T \phi(\mathbf{x}) + b] = \mathrm{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b\right)$.

In [25,26], a least squares type of SVM was introduced by modifying the problem formulation so as to obtain a system of linear equations in the dual space. This is done by taking a least squares cost function, with equality instead of inequality constraints. Hence, the parameters $\mathbf{w}$ and b of the hyperplane can be obtained by solving the following alternative formulation:

$$\min_{\mathbf{w},b,\xi_i} \Phi(\mathbf{w}, b, \xi_i) = \frac{1}{2}(\mathbf{w}^T\mathbf{w}) + \frac{C}{2}\sum_{i=1}^{N}(\xi_i)^2 \quad (7)$$

subject to equality constraints $y_i[\mathbf{w}^T\phi(\mathbf{x}_i) + b] = 1 - \xi_i, \quad i = 1,\ldots,N$.

The Lagrangian defined in the LS-SVM dual space is written as follows:

$$L(\mathbf{w}, b, \xi_i; \alpha_i) = \Phi(\mathbf{w}, b, \xi_i) - \sum_{i=1}^{N} \alpha_i \{ y_i [\mathbf{w}^T \phi(\mathbf{x}_i) + b] - 1 + \xi_i \}, \quad (8)$$

where $\alpha_i \in \Re$, $i = 1,\ldots,N$, are Lagrange multipliers, which can be positive or negative due to the equality constraints, as follows from the Karush–Kuhn–Tucker (KKT) conditions [25,26].

The conditions for optimality:

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \phi(\mathbf{x}_i), \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0, \\ \dfrac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = C\xi_i, \quad \forall i = 1,\ldots,N, \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i \left[\mathbf{w}^T \phi(\mathbf{x}_i) + b\right] - 1 + \xi_i = 0, \quad \forall i = 1,\ldots,N \end{cases} \quad (9)$$

can be written as the linear system

$$\begin{bmatrix} 0 & \mathbf{y}^T \\ \mathbf{y} & \Omega + C^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \quad (10)$$

with $\mathbf{y} = [y_1,\ldots,y_N]^T$, $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_N]^T$, $\mathbf{1} = [1,\ldots,1]^T$ and I refers to the identity matrix. Mercer's conditions for kernels are embedded within the matrix Ω, whose elements read simply as $\Omega_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$.

3. Computational experiments

In what follows, we provide details on how the sensitivity analysis experiments have been conducted. Then, we present comparative charts displaying, for each configuration of kernel type, kernel parameter value, and derived feature, the performance levels achieved by the two kernel machines in terms of misclassification rate as well as sensitivity and specificity indices [13]. Finally, we situate the performance levels achieved by optimally configured LS-SVM in the relevant literature.

3.1. Configuration of the sensitivity analysis experiments

In the sensitivity analysis experiments accomplished, we have assessed the performance of the standard and least squares SVM models with regard to the variation of the radius parameter alone, keeping the value of the regularization parameter C constant in 100. This value was achieved after some preliminary experiments and agrees with the fact that SVM models with low values of C tend in general to achieve better performance than those with high values of this parameter. Although we know that there are several rules-of-thumb to select the values of the radius parameter [30], we have opted to set the values of σ as 2^i, $i = -10, -9, \ldots, 14, 15$. For each of the 26 values in this range, a 10-fold cross-validation process was performed to better gauge the average performance of the models.

3.2. Sensitivity analysis results

In Table 1, we provide the value(s) of the radius parameter that has(ve) yielded the best average result, in terms of cross-validation error, for each triad ⟨feature vector, model type, kernel type⟩. In this table, the accuracy results are given in terms of average and standard deviation. Besides the feature vectors achieved through DWT, simulations have also been conducted with the EEG series without pre-processing. On the other hand, Figs. 2–9 depict the sensitivity profiles exhibited by the two types of kernel machine for the feature vectors considered. The sensitivity profiles relate to the following three criteria [12,13,18]: misclassification rate (i.e., the ratio between the number of misclassified EEG segments, either healthy or unhealthy, and the total number of segments available in the test partition), sensitivity, also called true positive rate (i.e., the ratio between the number of correctly classified healthy EEG segments and the total number of healthy EEG

Table 1
Comparative analysis: for each triad ⟨features vector, model type, kernel type⟩, we show the value of σ that gives the best average accuracy result.

Feature vector	Model type	Kernel type			
		RBF		ERBF	
		RBF error ± std. dev.	σ	ERBF error ± std. dev.	σ
EEG raw data	SVM	0.000 ± 0.000	16	0.000 ± 0.0000	4
	LS-SVM	0.000 ± 0.000	16	0.000 ± 0.0000	4
W_Avg	SVM	0.100 ± 0.0197	0.5	0.105 ± 0.0273	0.5
	LS-SVM	0.095 ± 0.0229	0.5	0.100 ± 0.0289	0.5
W_Std	SVM	0.000 ± 0.0000	1,2,4,8	0.000 ± 0.0000	1,2,4,8,16
	LS-SVM	0.000 ± 0.0000	0.5,1,2	0.000 ± 0.0000	1,2,4,8,16
W_Max	SVM	0.000 ± 0.0000	0.5	0.000 ± 0.0000	1,2,4
	LS-SVM	0.0050 ± 0.005	0.25,1,2	0.000 ± 0.0000	0.5,1,2,4

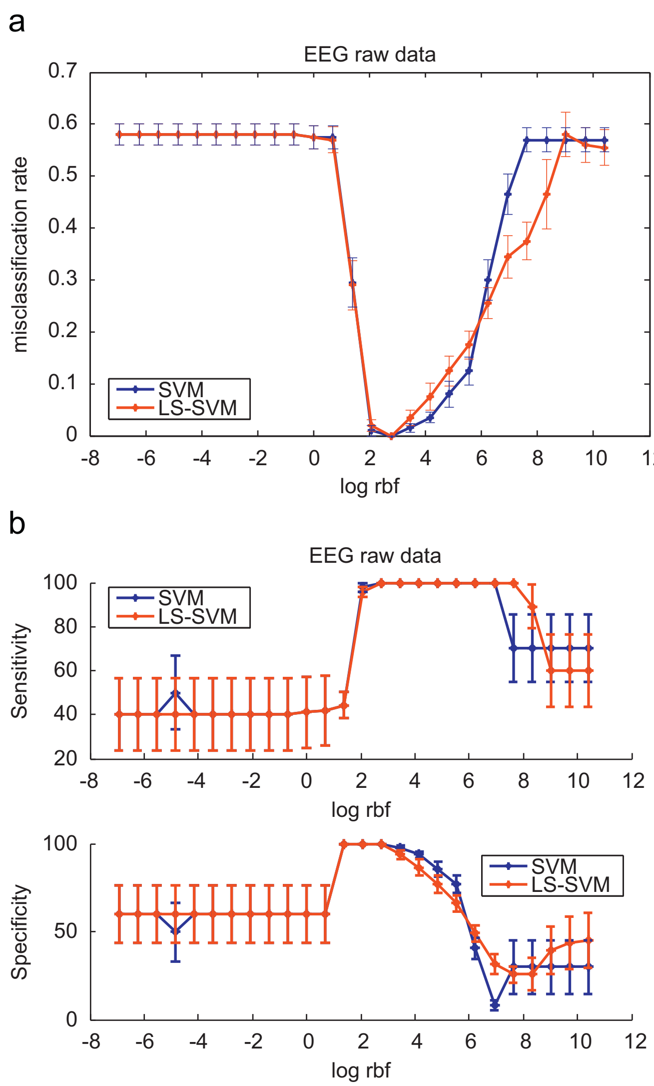

Fig. 2. Sensitivity profiles exhibited by SVM and LS-SVM models taking EEG raw data as input and RBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

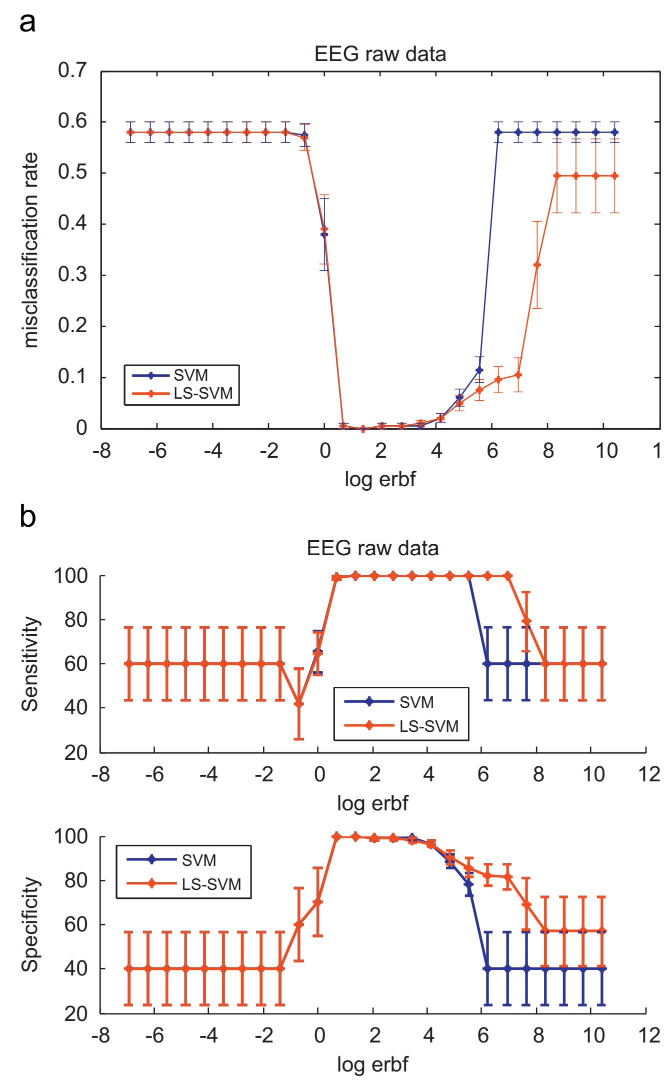

Fig. 3. Sensitivity profiles exhibited by SVM and LS-SVM models taking EEG raw data as input and ERBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

segments available in the test partition), and specificity, also called true negative rate (i.e., the ratio between the number of correctly classified epileptic seizure EEG segments and the total number of epileptic seizure EEG segments available in the test partition). The bars in these graphs represent the variance of each measure (one standard deviation from the mean) for each value of σ considered.

3.3. Discussion

Considering the results presented in Table 1, and Figs. 2–9, one can observe that, in most of the cases, the average performance indices (i.e., misclassification rate as well as sensitivity and specificity values) showed by the best models produced for each type of kernel machine were quite similar, reaching a remarkable level. From these results, it is possible to conclude that different combinations of the radius parameter value and the kernel function may yield best error rates exhibited by both machines. This means that the task of fine-tuning the SVM and LS-SVM models is instrumental for one to achieve optimal performance, even though this task does not seem to be so hard in nature in the present context of EEG signal classification. Also interesting is the fact that the best models of both types of machine could do very well with the healthy/unhealthy discrimination even when taking as input the original, raw time series.

Regarding the performance profiles depicted in Figs. 2–9, it is possible to assert that, in general, the two types of machines display the same patterns of sensitivity to the choice of the kernel parameter value—for all three measures, there are some regions of stability separated by zones of abrupt variation. It is worth noting that the zones of stability usually correlate with larger absolute values of σ; this is because the kernel functions are of a symmetric nature and their outcomes tend to 1 when the absolute value of the radius parameter increases. That is, in these cases, the output of the machines will always be the same (always one of the two classes), irrespective of the given input data.

Except given to the feature vector produced via the average of the wavelet coefficients, when comparing the performance profiles, one can observe that the sensitivity behavior is quite similar in general. This indicates that the selection of parameters to the kernel appears to be more important than the technique used for feature extraction. Anyway, the standard deviation of the

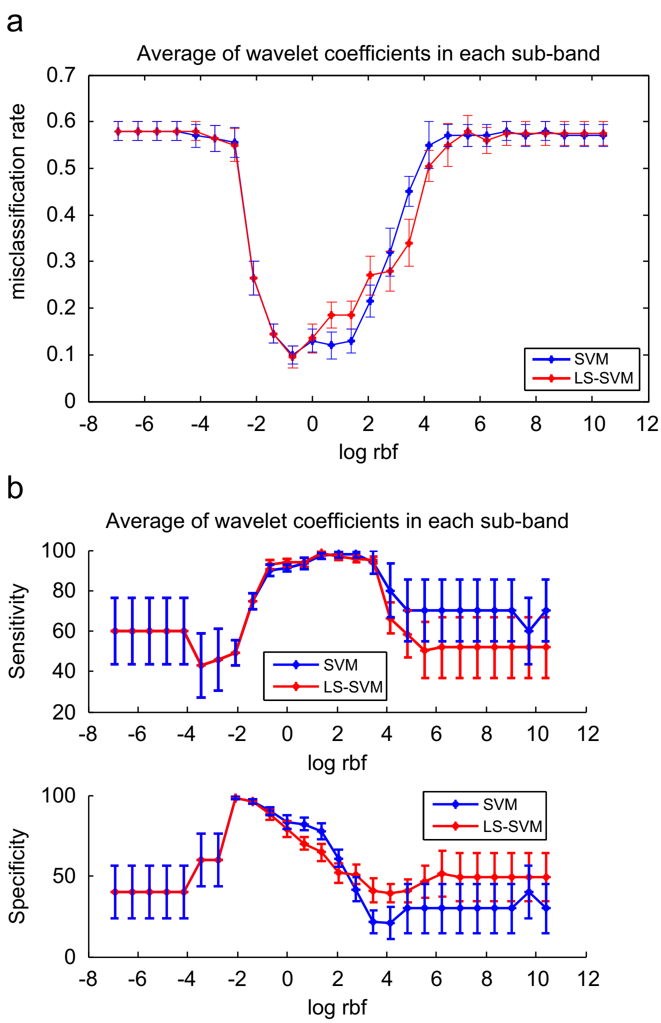

Fig. 4. Sensitivity profiles exhibited by SVM and LS-SVM models taking the average of wavelet coefficients in each sub-band as input and RBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

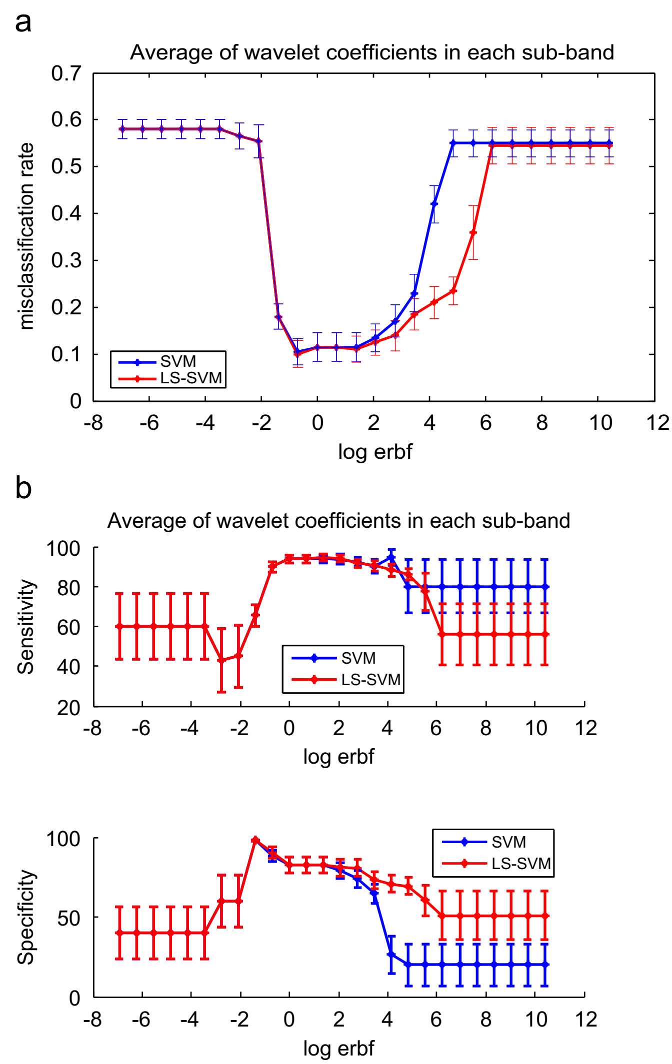

Fig. 5. Sensitivity profiles exhibited by SVM and LS-SVM models taking the average of wavelet coefficients in each sub-band as input and ERBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

wavelet coefficients in each sub-band seems to be the best type of feature in the contest (see Figs. 6 and 7).

Considering the features extracted by taking the standard deviation and the maximum of the wavelet coefficients in particular, there is a noticeable range of values for the kernel parameter that yields better performance. For example, for the ERBF kernel with feature vector extracted through standard deviation, regardless of the type of machine considered, a good value of the radius parameter usually lies in the range [1,16]. On the other hand, for the RBF kernel, this range turns out to be more restricted—namely [1, 4], for the SVM, and [0.5, 2], for the LS-SVM.

3.4. Contrast with related work

As mentioned before, several prominent works have already made use of the EEG dataset adopted in this paper (originally published in [6,7]) in order to assess the pros and cons of different machine learning approaches to cope with the epilepsy diagnosis problem. A great parcel of these works has reported their results in terms of overall classification accuracy and sensitivity/specificity values; a few have also reported confusion matrices as well as receiver operating characteristic (ROC) curves. Each column of a confusion matrix represents the instances in a predicted class, while each row represents the instances in the actual class. One benefit of a confusion matrix is that it allows one to see if a classifier is confusing two classes (i.e., commonly mislabeling one as another). So, one could tell the frequency at which the classifier is misclassifying healthy EEG segments as unhealthy, and vice versa. On the other hand, ROC curves provide a way to visually inspect the whole spectrum of sensitivity-specificity values a binary classifier produces for a given test dataset. It is a plot of the sensitivity vs. (1 − specificity) for different discrimination thresholds: The best classifier for a problem is the one for which the area under the ROC curve (AUC) equals to one. In what follows, we provide a brief comment on the accuracy results achieved by a sample of related works which have employed neural networks or kernel machines as part of their diagnosis system.

Nigam and Graupe [4] employed a multistage nonlinear filter in combination with a LAMSTAR neural network to cope with the automatic detection of epileptic seizures. As in the present work, the authors made use of only two sets of the EEG dataset, namely sets A and E, to assess their methodology. The overall success percentage achieved by the system, considering both the false positive and false negative rates, was of 97.2%. On the other hand, in the work of Subasi [12], who made use of mixture-of-experts (ME) models to cope with the same dichotomization problem, the author reports 94.5% as total accuracy rate, which was better than

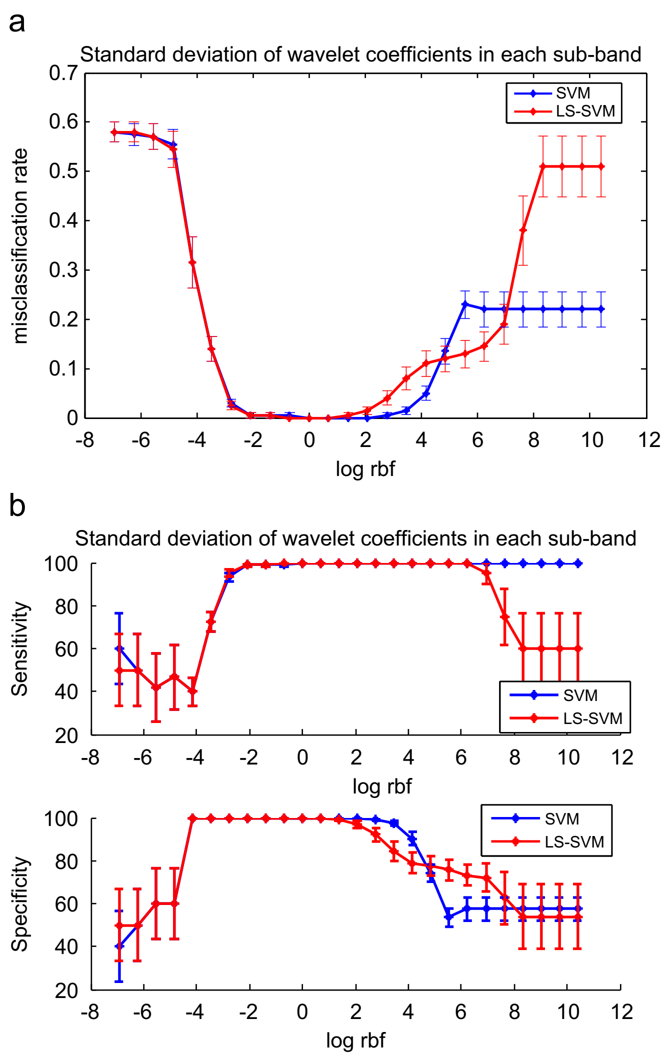

Fig. 6. Sensitivity profiles exhibited by SVM and LS-SVM models taking the standard deviation of wavelet coefficients in each sub-band as input and RBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

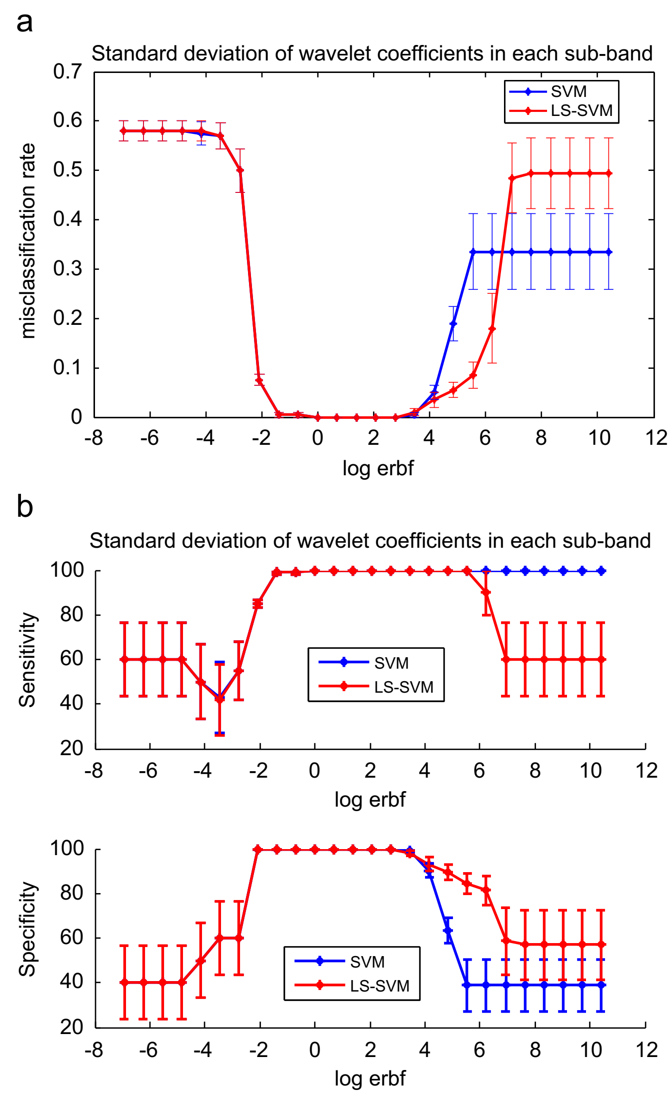

Fig. 7. Sensitivity profiles exhibited by SVM and LS-SVM models taking the standard deviation of wavelet coefficients in each sub-band as input and ERBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

the score achieved by single multilayer perceptron (MLP) neural networks (93.2%). The specificity and sensitivity values reported for the ME and MLP models were, respectively, 94%/92.6% and 95%/93.6%. ME models induced with wavelet coefficients have also been considered by Übeyli [13], even though, in that work, the performance of the models was measured over three sets of the EEG dataset (namely, sets A, D, and E). The total classification accuracy obtained by the ME network structures reported in [13] was 93.17% and ROC curves for single MLP and ME classifiers were also provided.

The paper of Tzallas et al. [14] presents a methodology of analysis of EEG signals that is based on time–frequency analysis. Initially, selected segments of the EEG signals are analyzed using time–frequency methods and several features are extracted for each segment, representing the energy distribution in the time–frequency plane. Then, those features are used as input to a feedforward neural network, which provides the final classification. In order to evaluate the methodology, the authors have generated four different classification problems, and the results achieved in terms of overall accuracy varied from 97.72% to 100%. By other means, Kocyigit et al. [15] have designed an MLP classifier based on the Fast independent component analysis (ICA) feature extraction technique in order to discriminate between normal and epileptic patients. The resulting system achieved a sensitivity rate of 98% and specificity rate of 90.5%.

Other types of neural network models than feedforward networks have also been investigated to tackle the epileptic/non-epileptic EEG signal discrimination. For instance, the works of Güler and Übeyli [16] and of Kannathal et al. [17] have both considered the application of a well-known class of neurofuzzy models, namely, the adaptive neurofuzzy inference system (ANFIS); the main difference between them lies in the type of feature extracted, either via wavelet transform or entropy measures, respectively. While, the classification accuracy reported by the second work was typically above 90% for different entropy measures, that achieved by the former was of 98.68% (with Daubechies of order 2 adopted as wavelet basis). On the other hand, Güler et al. [18] have investigated the diagnostic accuracy of recurrent neural networks (RNN) employing Lyapunov exponents trained with the Levenberg–Marquardt algorithm. The results were obtained for sets A, D, and E of the EEG dataset, and the values of specificity, sensitivity, and total classification accuracy of the produced RNN models were 97.38%, 96.88%/96.13%, and 96.79%, respectively. Finally, it is worth commenting about the

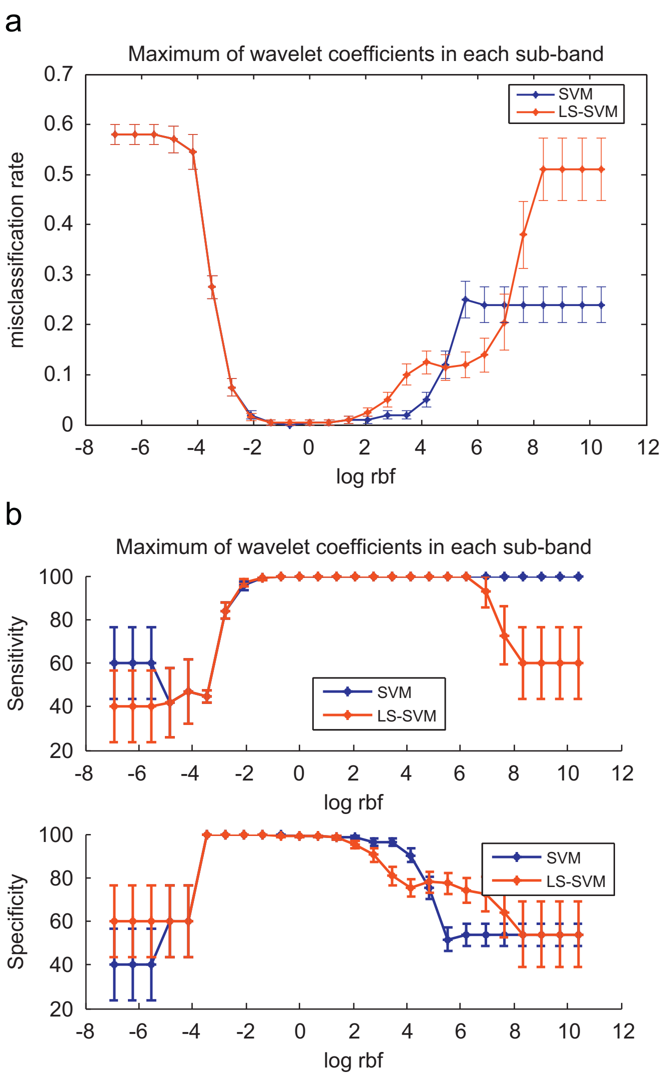

Fig. 8. Sensitivity profiles exhibited by SVM and LS-SVM models taking the maximum of wavelet coefficients in each sub-band as input and RBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

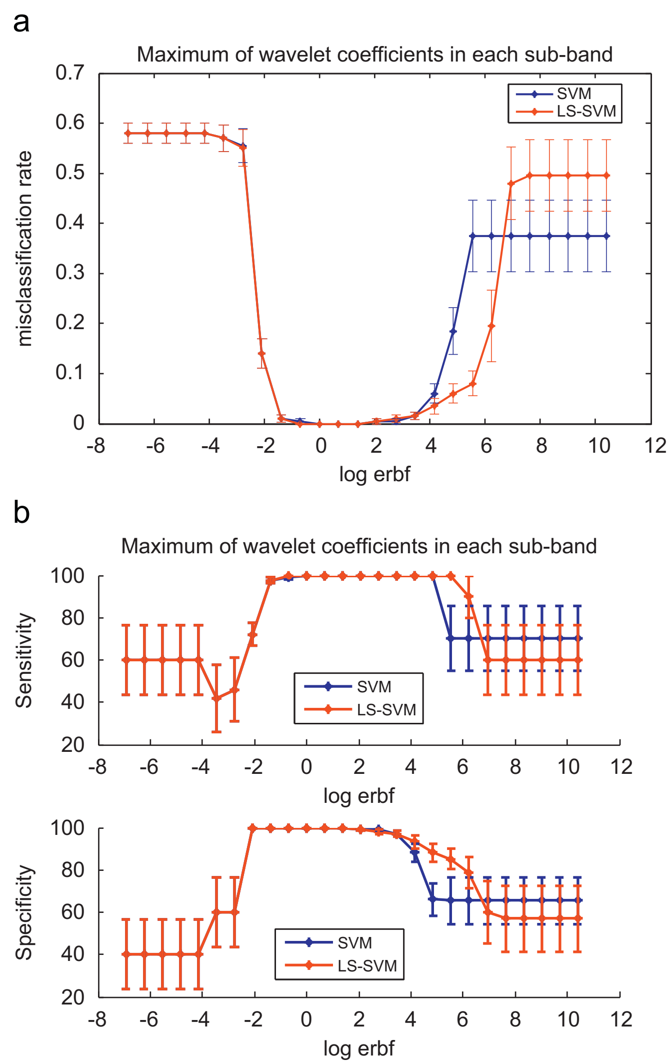

Fig. 9. Sensitivity profiles exhibited by SVM and LS-SVM models taking the maximum of wavelet coefficients in each sub-band as input and ERBF as kernel function: (a) misclassification rate and (b) sensitivity and specificity values.

results recently achieved by Übeyli with combined models of neural networks [19] and by Übeyli and Güler with multiclass SVM as EEG signal classifiers [20,21]. In [19], a two-level network architecture is employed: the first-level networks were implemented for the EEG signals classification using the statistical features as inputs, while the second-level networks were trained using the outputs of the first-level networks as input data. Three types of EEG signals were classified by the system, which achieved 96%, 94.5%/94%, and 94.83% as specificity, sensitivity, and total accuracy rate, respectively. Conversely, in [21], Übeyli investigated the behavior of multiclass SVM configured with error correcting output codes (ECOC) for classification of the signals available in all five sets of the EEG dataset (sets A–E). The features were extracted by means of eigenvector methods. The total classification accuracy obtained by the resulting SVM was very high, namely, 99.30%, and specificity/sensitivity values as well as the ROC curve are also provided in the paper.

As a means to contrast the performance of LS-SVM models with the results reported in the aforementioned papers, we have followed the same experimental methodology adopted by Übeyli [21] (considering, however, two instead of five classes) and performed additional experiments involving a different manipulation of the EEG dataset. In this new format, the 100 time series available for each class were windowed by a rectangular window composed of 256 discrete data and then the training and test sets were derived. In total, 3200 samples (1600 for each class) were produced, of which 1600 samples were randomly selected to compose the training partition (800 for each class) and the remaining 1600 vectors (800 for each class) were used for test. Then, for each combination of kernel type, kernel parameter value, and derived feature, we have optimally calibrated the LS-SVM models by following the sensitivity analysis process described previously. As result, we have produced the confusion matrices and ROC curves described in Tables 2–5 and Fig. 10, respectively. By taking into account these results, one can notice

Table 2
Confusion matrices produced by LS-SVM taking EEG raw data as input and RBF (a) and ERBF (b) as kernel function.

	Predicted class			
	(a)		(b)	
Real class	800	0	800	0
	11	789	2	798

Table 3
Confusion matrices produced by LS-SVM taking the average of wavelet coefficients in each sub-band as input and RBF (a) and ERBF (b) as kernel function.

	Predicted class			
	(a)		(b)	
Real class	760	40	764	36
	96	704	71	729

Table 4
Confusion matrices produced by LS-SVM taking the standard deviation of wavelet coefficients in each sub-band as input and RBF (a) and ERBF (b) as kernel function.

	Predicted class			
	(a)		(b)	
Real class	800	0	800	0
	0	800	0	800

Table 5
Confusion matrices produced by LS-SVM taking the maximum of wavelet coefficients in each sub-band as input and RBF (a) and ERBF (b) as kernel function.

	Predicted class			
	(a)		(b)	
Real class	797	3	797	3
	4	796	6	794

that it is possible to achieve the remarkable score of 100% for all criteria (sensitivity, specificity, and total classification accuracy) by making use of the LS-SVM model configured with either RBF or ERBF kernels and taking the standard deviation of wavelet coefficients in each sub-band as derived feature.

4. Summary

In this paper, least squares SVM classifiers were designed to deal with the epilepsy diagnosis task through EEG signal classification. In particular, a sensitivity analysis contrasting the performance profiles exhibited by LS-SVM and standard SVM with respect to the choice of the kernel function and its parameter value was performed. In this analysis, different types of features extracted from the raw EEG signal were considered. Such study is interesting as it can provide hints on how these kernel machines are affected by the hyper-parameter tuning process [29] as well the derived features from complex signals such as EEG. In such regard, the results presented here suggest that the dependence profiles exhibited by the two machines tend to be qualitatively similar, and their performance is not too much influenced by the choice of the derived EEG feature vector. In addition, we have contrasted the performance achieved by optimally configured LS-SVM with the accuracy rates reported by the models investigated in recent related work, showing that, by means of the sensitivity analysis process, it is possible to produce an LS-SVM that can achieve the remarkable score of 100% as total accuracy rate for the EEG dataset considered.

As ongoing work, we are currently extending the scope of investigation in SVM and LS-SVM sensitivity analysis by considering other types of biomedical signal processing problems, like those involving ECG [28] and Doppler ultrasound signals [32–34]. Moreover, we plan to investigate how the combination of models coming from different types of kernel machines [35,36] can improve the levels of performance, in terms of accuracy and generalization, from that achieved by each machine type alone. In this regard, Übeyli [19] has recently shown the potentials of combining models while tackling the multiclass version of the EEG signal classification problem, even though the classifiers considered in that work were restricted to be feedforward neural networks.

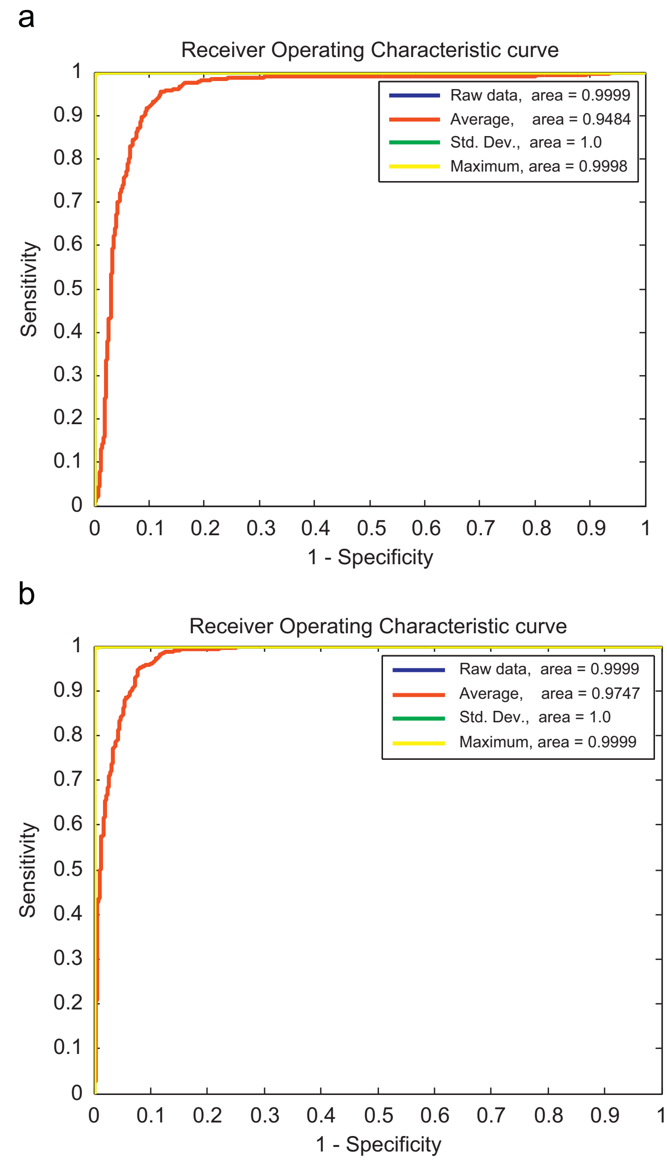

Fig. 10. ROC curves produced by LS-SVM taking EEG raw data, W_Avg, W_Std, and W_Max as input, and RBF (a) and ERBF (b) as kernel function. One should notice that some of the curves overlap in the two plots.

Conflict of interest statement

None declared.

Acknowledgements

This work was financially sponsored by CNPq/Brazil, via Grants #474843/2008-4 and #312934/2009-2, and MackPesquisa, via Grant #1349.

References

[1] L. Sörnmo, P. Laguna, Bioelectrical Signal Processing in Cardiac and Neurological Applications, Elsevier, Amsterdam, 2005.
[2] E. Niedermeyer, F.L. da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Lippincott Williams & Wilkins, 2004.
[3] X. Liao, Y. Yin, C. Li, D. Yao, Application of SVM framework for classification of single trial EEG, in: J. Wang et al.(Ed.), Advances in Neural Networks—ISNN 2006, Lecture Notes in Computer Science, vol. 3973, Springer 2006, pp. 548–553.
[4] V.P. Nigam, D. Graupe, A neural-network-based detection of epilepsy, Neurol. Res. 26 (2004) 55–60.
[5] K. Lehnertz, Non-linear time series analysis of intracranial EEG recordings in patients with epilepsy—an overview, Int. J. Psychophysiol. 34 (1999) 45–52.
[6] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, Phys. Rev. E 64 (2001) article ID 061907.
[7] R.G. Andrzejak, G Widman, K. Lehnertz, C. Rieke, P. David, C.E. Elger, The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy, Epilepsy Res. 44 (2001) 129–140.
[8] F.M. Alicata, C. Stefanini, M. Elia, R. Ferri, S. Del Gracco, S.A. Musumeci, Chaotic behavior of EEG slow-wave activity during sleep, Electron. Clin. Neurophysiol. 99 (1996) 539–543.
[9] E.D. Übeyli, Statistics over features: EEG signals analysis, Comput. Biol. Med. 39 (2009) 733–741.
[10] A. Subasi, Epileptic seizure detection using dynamic wavelet network, Expert Syst. Appl. 28 (2005) 701–711.
[11] A. Subasi, E. Ercelebi, Classification of EEG signals using neural network and logistic regression, Comput. Methods Prog. Biomed. 78 (2005) 87–99.
[12] A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Syst. Appl. 32 (2007) 1084–1093.
[13] E.D. Übeyli, Wavelet/mixture of experts network structure for EEG signals classification, Expert Syst. Appl. 34 (2008) 1954–1962.
[14] A.T. Tzallas, M.G. Tsipouras, D.I. Fotiadis, Automatic seizure detection based on time–frequency analysis and artificial neural networks, Comput. Intel. Neurosci. 2007 (2007) article ID 80510.
[15] Y. Kocyigit, A. Alkan, H. Erol, Classification of EEG recordings by using fast independent component analysis and artificial neural network, J. Med. Syst. 32 (2008) 17–20.
[16] İ. Güler, E.D. Übeyli, Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, J. Neurosci. Methods 148 (2005) 113–121.
[17] N. Kannathal, L.C. Min, U.R. Acharya, P.K. Sadasivan, Entropies for detection of epilepsy in EEG, Comput. Methods Prog. Biomed. 80 (2005) 187–194.
[18] N.F. Güler, E.D. Übeyli, İ. Güler, Recurrent neural networks employing Lyapunov exponents for EEG signals classification, Expert Syst. Appl. 29 (2005) 506–514.
[19] E.D. Übeyli, Combined neural network model employing wavelet coefficients for EEG signals classification, Digital Signal Process 19 (2009) 297–308.
[20] I. Güler, E.D. Übeyli, Multiclass support vector machines for EEG-signals classification, IEEE Trans. Inf. Technol. Biomed. 11 (2007) 117–126.
[21] E.D. Übeyli, Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines, Comput. Biol. Med. 38 (2008) 14–22.
[22] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, London, 2000.
[23] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[24] B. Schölkopf, A. Smola, Learning with Kernels, The MIT Press, Cambridge, 2002.
[25] J.A.K. Suykens, J. Vandewalle, Least squares support machine classifiers, Neural Proc. Lett. 9 (1999) 293–300.
[26] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, in: Least Squares Support Vector Machines, World Scientific Publishers, Singapore, 2002.
[27] T. Van Gestel, J.A.K. Suykens, B. Baesens, S. Viaene, J. Vanthienen, G. Dedene, B. De Moor, J. Vandewalle, Benchmarking least squares support machine classifiers, Mach. Learn. 54 (2004) 5–32.
[28] P. Kemal, A. Bayram, S. Güneş, Computer aided diagnosis of ECG data on the least square support vector machine, Digital Signal Process. 18 (2008) 25–32.
[29] O. Chapelle, V. Vapnik, O. Bousquet, S. Mukherjee, Choosing multiple parameters for support vector machines, Mach. Learn. 46 (2002) 131–159.
[30] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (2004) 113–126.
[31] F. Friedrichs, C. Igel, Evolutionary tuning of multiple SVM parameters, Neurocomputing 64 (2005) 107–117.
[32] E.D. Übeyli, Probabilistic neural networks employing Lyapunov exponents for analysis of Doppler ultrasound signals, Comput. Biol. Med. 38 (2008) 82–89.
[33] E.D. Übeyli, İ. Güler, Feature extraction from Doppler ultrasound signals for automated diagnostic systems, Comput. Biol. Med. 35 (2005) 735–764.
[34] E.D. Übeyli, Usage of eigenvector methods to improve reliable classifier for Doppler ultrasound signals, Comput. Biol. Med. 38 (2008) 563–573.
[35] C.A.M. Lima, A.L.V. Coelho, F.J. Von Zuben, Model selection based on VC-dimension for heterogeneous ensembles of support vector machines, Proceedings of the Fourth International Conference on Recent Advances in Soft Computing, Nottingham University Press, Nottingham, 2002 pp. 459–464.
[36] C.A.M. Lima, W.J.P. Villanueva, E.P. dos Santos, F.J. Von Zuben, A multistage ensemble of support vector machine variants, Proceedings of the Fifth International Conference on Recent Advances in Soft Computing, Nottingham, University Press, Nottingham, 2004, pp. 670–675.

Clodoaldo A.M. Lima received the B.Sc. degree (1997) in Electrical Engineering from Federal University of Juiz de Fora (UFJF), Brazil, and M.Sc. (2000) and Ph.D. (2004) degrees in Computer Engineering from the State University of Campinas (Unicamp), Brazil. He was a post-doctoral fellow (from 2004 to 2006) and then a research collaborator (2007–2008) at the same university. Currently, he is with the Graduate Program in Electrical Engineering at Mackenzie Presbyterian University, São Paulo, Brazil. His current research interests include multivariate analysis, machine learning, time series forecasting, and digital signal processing.

André L.V. Coelho received the B.Sc. degree in Computer Engineering in 1996, and earned the M.Sc. and Ph.D. degrees in Electrical Engineering in 1998 and 2004, respectively, all from the State University of Campinas (Unicamp), Brazil. He has a record of publications related to the themes of machine learning, data mining, computational intelligence, metaheuristics, and multiagent systems. He is a member of ACM and has served as a reviewer for a number of scientific conferences and journals. Currently, he is an adjunct professor affiliated with the Graduate Program in Applied Informatics at the University of Fortaleza, Ceará, Brazil.

Marcio Eisencraft received the M.Sc. (2001) and Ph.D. (2006) degrees in Electrical Engineering from the University of São Paulo (USP), Brazil. From 2006 to 2010, he was with the Graduate Program in Electrical Engineering at Mackenzie Presbyterian University. Currently, he is affiliated with the Federal University of ABC, São Paulo, Brazil. His current research interests include digital signal processing, intermittent time series, and chaotic and nonlinear systems.