Expert Systems with Applications 39 (2012) 1956–1967

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection

Felipe Alonso-Atienza [a,*], José Luis Rojo-Álvarez [a], Alfredo Rosado-Muñoz [b], Juan J. Vinagre [a], Arcadi García-Alberola [c], Gustavo Camps-Valls [b]

[a] Departamento de Teoría de la Señal y Comunicaciones, Universidad Rey Juan Carlos, Camino del Molino s/n, 28943 Fuenlabrada, Madrid, Spain
[b] Departament de Enginyeria Electrónica, Universitat de Valéncia, Doctor Moliner 50, 46100 Burjassot, Valéncia, Spain
[c] Unidad de Arritmias, Hospital Universitario Virgen de la Arrixaca, Ct. Madrid-Cartagena s/n, 30120 El Palmar, Murcia, Spain

ARTICLE INFO

Keywords:
Feature selection
Support vector machines
Bootstrap
Arrhythmia classification
Ventricular fibrillation detection

ABSTRACT

Early detection of ventricular fibrillation (VF) is crucial for the success of the defibrillation therapy in automatic devices. A high number of detectors have been proposed based on temporal, spectral, and time–frequency parameters extracted from the surface electrocardiogram (ECG), showing always a limited performance. The combination ECG parameters on different domain (time, frequency, and time–frequency) using machine learning algorithms has been used to improve detection efficiency. However, the potential utilization of a wide number of parameters benefiting machine learning schemes has raised the need of efficient feature selection (FS) procedures. In this study, we propose a novel FS algorithm based on support vector machines (SVM) classifiers and bootstrap resampling (BR) techniques. We define a backward FS procedure that relies on evaluating changes in SVM performance when removing features from the input space. This evaluation is achieved according to a nonparametric statistic based on BR. After simulation studies, we benchmark the performance of our FS algorithm in AHA and MIT-BIH ECG databases. Our results show that the proposed FS algorithm outperforms the recursive feature elimination method in synthetic examples, and that the VF detector performance improves with the reduced feature set.

1. Introduction

Ventricular fibrillation (VF) is a life-threatening cardiac arrhythmia caused by a disorganized electrical activity of the heart (Moe, Abildskov, & Han, 1964). During VF, ventricles contract in an unsynchronized way (Baykal, Ranjan, & Thakor, 1997), failing the heart pumping of blood. Sudden cardiac death will follow in a matter of minutes unless medical care is provided immediately. The only effective treatment to revert VF is the electrical defibrillation of the heart (Beck, Pritchard, Giles, & Mensah, 1947), which consists of delivering a high energy electrical stimulus to the heart with a so-called defibrillator device (Mirowski, Mower, & Reid, 1980; Thakor, 1984). Clinical and experimental studies have demonstrated that the success of defibrillation is inversely related to the time interval between the beginning of the VF episode and the application of the electrical discharge (White, Asplin, Bugliosi, & Hankins, 1996; Yakaitis, Ewy, & Otto, 1980). This has impelled the development of VF detection algorithms for monitoring systems and automatic external defibrillators (AED). These algorithms analyze the surface electrocardiogram (ECG), providing an accurate fast diagnosis of VF, in order to reduce the reaction time of the health care personnel in monitory systems, and to supply the appropriate therapy without the need of qualified personnel in AEDs (Faddy, 2006).

A high number of VF detection schemes based on parameters extracted from the ECG have been proposed in the literature. These parameters are usually obtained from different ECG representations, such as time, frequency and time–frequency domains. Time-domain methods analyze the morphology of the ECG to discriminate VF rhythms (Aubert, Denys, Ector, & Geest, 1982; Chen, Thakor, & Mower, 1987; Chen, Clarkson, & Fan, 1996; Clayton, Murray, & Campbell, 1993; Jack et al., 1986; Thakor, Zhu, & Pan, 1990; Zhang, Zhu, Thakor, & Wang, 1999). Frequency-domain measurements are motivated by experimental studies supporting that VF is not a chaotic and disorganized pathology, but instead a certain degree of spatio-temporal organization exists (Clayton, Murray, & Campbell, 1995; Davidenko, Pertsov, Salomonsz, Baxter, & Jalife, 1992; Jalife, Gray, Morley, & Davidenko, 1998). Spectral description of the ECG has revealed important differences between normal and fibrillatory rhythms (Clayton et al., 1995; Forster &

* Corresponding author. Address: Escuela Técnica Superior de Ingeniería de Telecomunicación, Dept. Teoría de la Señal y Comunicaciones, Universidad Rey Juan Carlos, Camino del molino s/n. 28943, Fuenlabrada, Madrid, Spain. Tel.: +34 914888702; fax: +34 914887500.

E-mail address: felipe.alonso@urjc.es (F. Alonso-Atienza).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2011.08.051

Weaver, 1982; Herschleb, Heethaar, de Tweel, Zimmerman, & Meijler, 1979; Murray, Campbell, & Julian, 1985), and in this context, relevant parameters of the ECG spectrum have been used for developing VF detectors (Barro, Ruiz, Cabello, & Mira, 1989; Kuo & Dillman, 1978; Forster & Weaver, 1982; Nolle et al., 1989; Nygards & Hulting, 1978). On the other hand, given the non-stationary nature of the VF signal, algorithms based on time–frequency distributions have been also proposed to detect VF episodes (Afonso & Tompkins, 1995; Rosado et al., 1999; Clayton & Murray, 1998).

Though many VF detectors based on temporal, spectral, or time–frequency parameters have been disclosed, comparative studies have shown that these algorithms are not optimal when considered separately (Amann, Tratnig, & Unterkofler, 2005; Clayton, Murray, & Campbell, 1994). The combination of ECG parameters have been suggested as a useful approach to improve detection efficiency. In Clayton et al. (1994), Neurauter et al. (2007) and Pardey (2007), a set of temporal and spectral features were used as input variables to a neural network, exhibiting better performance than previously proposed methods. Following this approach, other statistical learning algorithms such as clustering methods (Jekova & Mitev, 2002), support vector machines (SVM) (Ubeyli, 2008), or data mining general procedures (classification trees, self-organizing maps) (Rosado-Muñoz et al., 2002), have been explored aiming to enhance VF detection capabilities. However, this has increased the number of ECG parameters used to detect VF, which in turn has raised the need of efficient feature selection (FS) techniques for assessing the discriminatory properties of the selected variables (Ribeiro, Marques, Henriques, & Antunes, 2007; Zhang, Lee, & Lim, 2008). Besides of improving the accuracy of VF detectors, the use of FS techniques might help researchers to provide a better understanding of the unresolved mechanisms responsible for the initiation and perpetuation of VF.

In this paper, we present a novel FS algorithm to reduce the size of the input feature space while providing an accurate detection of VF episodes. We use a set of temporal, spectral, and time–frequency parameters extracted from the AHA and MIT-BIH ECG signal databases as the input space to nonlinear SVM. We choose SVM as detection algorithm for VF since they have shown an excellent performance in arrhythmia discrimination applications (Osowski, Hoai, & Markiewicz, 2004; Ubeyli, 2008), and it has been demonstrated that FS methods can further improve SVM performance (Guyon, Weston, Barnhill, & Vapnik, 2002). The relevance of input variables is evaluated by comparing the detection performance of the complete set of input variables and a reduced subset of them. This comparison is achieved according to a nonparametric statistical test, based on bootstrap resampling (BR) (Efron & Tibshirani, 1994). Starting with the whole set of input variables, we progressively eliminate the most irrelevant feature, until a subset of significant variables is identified. This ensures that the performance of the final VF detector will not be significantly different worse from the initial one containing all features. The aim of this study is, therefore, to develop an accurate VF detector using the smallest yet representative set of ECG parameters. We compare this novel method to the most commonly used FS algorithm in the SVM literature, the so-called SVM recursive feature elimination (SVM-RFE) (Guyon et al., 2002; Rakotomamonjy, 2003), by means of a toy example. Then, we apply the proposed FS algorithm to the ECG signal databases.

The paper is organized as follows. Section 2 provides a brief background on SVM and FS techniques. Section 3 describes the ECG database used in this study. In Section 4, the proposed FS algorithm is presented. Section 5 is dedicated to analyze the performance of our novel FS method by means of a toy example. Then, in Section 6, results over the ECG signal databases are presented and finally, in Section 7, we discuss the scope and limitations of our approach along with future extensions.

2. Background

This section reviews the SVM formulation and the field of FS.

2.1. SVM classifiers

In recent years, SVM classification algorithms have been used in a wide number of practical applications (Camps-Valls, Rojo-Álvarez, & Martínez-Ramón, 2007). Their success is due to the SVM good properties of regularization, maximum margin, and robustness with data distribution and with input space dimensionality (Vapnik, 1995). SVM binary classifiers are sampled-based statistical learning algorithms which construct a maximum margin separating hyperplane in a reproducing kernel Hilbert space.

Let $\mathbf{V}$ be a set of N observed and labeled data, $\mathbf{V} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$. Be $\phi(\mathbf{x}_i)$ a nonlinear transformation to a (generally unknown) higher dimensional space $\mathbb{R}^l$, called Reproducing Hilbert Kernel Space (RKHS) in which a separating hyperplane is given by

$$\langle \phi(\mathbf{x}_i), \mathbf{w} \rangle + b = 0 \tag{1}$$

where $\langle \cdot, \cdot \rangle$ expresses the vector dot product operation. We know that $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ is a Mercer's kernel, which allows us to calculate the dot product of pairs of vectors transformed by $\phi(\cdot)$ without explicitly knowing neither the nonlinear mapping nor the RKHS. Two often used kernels are the linear, given by $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$, and the Gaussian, given by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right) \tag{2}$$

With these conditions, the problem is to solve

$$\min_{\mathbf{x}, b, \xi_i} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \right\} \tag{3}$$

constrained to $y_i(\langle \phi(\mathbf{x}_i), \mathbf{w} \rangle + b) - 1 + \xi_i \geqslant 0$ and to $\xi_i \geqslant 0$, for $i = 1, \ldots, N$, where ξ_i represent the losses, and C is a regularization parameter that represents a trade-off between margin and losses. By using Lagrange multipliers, (3) can be rewritten into its dual form, and then, the problem consists of solving

$$\max_{\alpha_i} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i y_i \alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \right\} \tag{4}$$

constrained to $0 \leqslant \alpha_i \leqslant C$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$, where α_i are the Lagrange multipliers corresponding to primal constraints. After obtaining the Lagrange multipliers, the SVM classification for a new sample $\mathbf{x}$ is simply given by

$$y = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{5}$$

Gaussian kernel width σ, and parameter C, are free parameters that have to be settled, and methods such as cross-validation or bootstrap resampling can be used for this purpose.

2.2. Feature selection techniques

Performance of supervised learning algorithms can be strongly affected by the number and relevance of input variables. FS techniques emerge to cope with this problem, aiming to find a subset of the input variables that best describes the underlying structure of the data as well or better than the original features

(Salcedo-Sanz, Camps-Valls, Pérez-Cruz, Sepulveda-Sanchís, & Bousoño-Calzón, 2004). FS techniques can be divided into three major categories (Saeys, Inza, & Larrañaga, 2007): filter methods, wrapper methods, and embedded methods.

Filter methods (Blum & Langley, 1997) evaluate the relevance of each variable by individually examining the intrinsic properties of the data. Variables are ranked according to a predefined relevance score, so that low-scored variables are removed. Those selected variables constitute then the input space of the classifier. Examples of filter methods (Salcedo-Sanz et al., 2004) are χ^2-test, Wilks's lambda criterion, principal/independent component analysis, mutual information techniques, correlation criteria, Fisher's discriminant scores, classification trees, self-organization maps, or fuzzy clustering. Filter methods are computationally easy and fast. However, they do not usually take into account the existence of nonlinear relationships among features, and the classification performance of a detector can be reduced in this previous step.

Wrapper methods (Kohavi & John, 1997) use the performance of a (possibly nonlinear) classification algorithm as quality criterion for evaluating the relevant information conveyed by a subset of features, i.e., a search procedure in the whole feature space is defined, and different candidate subsets are scored according to their classification performance. The subset of features which yields the lowest classification error is selected. Using a wrapper method often requires to define a classification algorithm, a relevance criterion to assess the prediction capacity of a given subset of features, and a searching procedure in the space of all possible subsets of features. The (usually heuristic) searching procedures can be divided into two types, namely, randomized and deterministic search methods. Examples of randomized methods are genetic algorithms or simulated annealing (Kohavi & John, 1997). On the other hand, deterministic methods, also called greedy strategies, perform a local search in the feature space and are computationally advantageous and robust against overfitting. The most common deterministic algorithms are forward and backward selection methods. Starting with an empty set of features, forward selection methods progressively add those variables that lead to the lowest classification error until the prediction performance is not longer improved. Backward selection methods start with the full set of features, and progressively eliminate those variables with the lowest discrimination capacity. Wrapper methods usually outperform filter strategies in terms of classification error, however, they are computationally intense and can suffer from overfitting if working with reduced data sets.

Finally, embedded methods combine the training process with the search in the feature space. For the particular case of the so-called nested methods (Guyon & Elisseeff, 2003), the search procedure is guided by estimating changes in the objective function (e.g., classifier performance) for different subsets of features. Together with backward and forward selection techniques, nested methods constitute very efficient schemes for FS (Guyon & Elisseeff, 2003).

An example of such nested method is the SVM-RFE algorithm which is a SVM weight-based method proposed by Guyon et al. for selecting relevant genes in a cancer classification problem (Guyon et al., 2002), and it was subsequently extended by Rakotomamonjy for its application in nonlinear classification problems (Rakotomamonjy, 2003). The SVM-RFE algorithm analyzes the relevance of input variables by estimating changes in the cost function

$$\Delta J_u = \|\mathbf{w}\|^2 - \|\mathbf{w}_u\|^2 \tag{6}$$

where $\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \phi(\mathbf{x}_i)$ represents the SVM weight vector in the RKHS for the complete set of input variables and $\mathbf{w}_u = \sum_{i=1}^{N} \alpha_i^{(u)} y_i \phi\left(\mathbf{x}_i^{(u)}\right)$ denotes the SVM weight vector when variable u is removed. It is assumed that $\alpha_i^{(u)} = \alpha_i$ to compute changes in ΔJ_u. A detailed description of the algorithm formulation can be found in Guyon et al. (2002) and Rakotomamonjy (2003).

In this study, we develop an embedded method based on the SVM formulation. Previously proposed embedded methods Rakotomamonjy (2003), Neumann, Schnörr, and Steidl (2005) and Bi et al. (2003) are based on scores which may have significant variations with small variations on the input data. Therefore, a robust statistical criterion would be desirable to evaluate the relevance of a set of variables. We propose the use of BR for this purpose, as presented in Section 4.

3. ECG parameters database

This section details the characteristics of the datasets used in this study and the features extracted.

3.1. Data collection and pre-processing

ECG signals from the AHA Arrhythmia Database (8200 series) (AHA, 2010) and the MIT-BIH Malignant Ventricular Arrhythmia Database (MIT, 2010) were considered. No preselection of ECG episodes was made. A total of 29 patient recordings were analyzed, each containing an average of 30 min of continuous ECG, from which approximately 100 min corresponded to VF. For each record, segments of 128 samples and 125 Hz sampling frequency were used, giving a 1.024 s window for the analysis. This segment length was chosen to contain at least one QRS complex (if existing in the analyzed signal). A general signal pre-processing was done, firstly subtracting the mean ECG signal value, and secondly, low-pass filtering at 40 Hz to remove the 50 Hz or 60 Hz power line interference and other high frequency components that were not relevant for the analysis.

3.2. Time–frequency parametrization

Each window segment was processed to obtain a set of temporal (t), spectral (f), and time–frequency (tf) parameters (see Table 1). The first two parameters were extracted in the time domain, due to their simplicity and their ability to reject non-VF rhythms (Rosado et al., 2000). Let $x[n]$ be the sampled ECG signal. Then, the following temporal parameters were used:

- `VR`: Variance of the $x^2[n]$ signal, normalized by its maximum. `VR` is closely related to peak presence. Since VF signal lacks of prominent peaks, a high value of VR is considered as corresponding to a non-VF episode.
- `RatioVar`: Ratio of the variance of $x[n] - x[n-1]$ to the variance of its absolute value. This parameter accounts for the symmetry between positive and negative values of $x[n]$. Due to the oscillatory nature of FV episodes, high values of `RatioVar` were observed during FV.

Next, a total of 25 parameters were obtained from the Pseudo Wigner–Ville (PWV) distribution (Claasen & Mecklenbrauker, 1980). The time–frequency distribution of a time-dependent signal represents the evolution of its spectral components along time, providing with joint information of both time and frequency domains. Therefore, based on this time–frequency analysis, temporal, spectral, or time-frequency parameters can be defined. For each ECG segment, we calculated the absolute value of its PWV distribution. Then, components falling below 10% of the maximum were set to zero to eliminate noise and interference, while keeping the major informative content. In order to characterize VF episodes, two spectral bands of interest were defined (Herschleb et al., 1979; Macfarlane & Veitch, 1989). Since most of the energy

Table 1
Statistics of the temporal (t), spectral (f) and time–frequency (tf) ECG extracted parameters (mean ± std), for the different pathologies under consideration.

#	Variable	Domain	NORMAL	OTHER	VT	VF-FLUTTER
1	VR	t	$(8.2 \pm 6.7) \cdot 10^{+0}$	$(6.0 \pm 5.0) \cdot 10^{+0}$	$(1.6 \pm 3.4) \cdot 10^{+0}$	$(1.5 \pm 1.1) \cdot 10^{+0}$
2	RatioVar	t	$(1.6 \pm 0.5) \cdot 10^{+0}$	$(1.8 \pm 0.5) \cdot 10^{+0}$	$(2.5 \pm 0.6) \cdot 10^{+0}$	$(2.7 \pm 0.4) \cdot 10^{+0}$
3	PmxFreq	f	$(5.5 \pm 3.2) \cdot 10^{+0}$	$(4.0 \pm 2.5) \cdot 10^{+0}$	$(2.8 \pm 2.0) \cdot 10^{+0}$	$(2.6 \pm 1.2) \cdot 10^{+0}$
4	MaximFreq	f	$(2.2 \pm 0.8) \cdot 10^{+1}$	$(2.0 \pm 0.7) \cdot 10^{+1}$	$(1.5 \pm 0.8) \cdot 10^{+1}$	$(1.4 \pm 0.5) \cdot 10^{+1}$
5	MinimFreq	f	$(7.3 \pm 4.9) \cdot 10^{-1}$	$(6.3 \pm 3.8) \cdot 10^{-1}$	$(6.4 \pm 3.5) \cdot 10^{-1}$	$(6.9 \pm 3.6) \cdot 10^{-1}$
6	TSNZ	tf	$(1.1 \pm 0.6) \cdot 10^{+3}$	$(1.1 \pm 0.6) \cdot 10^{+3}$	$(1.6 \pm 0.5) \cdot 10^{+3}$	$(1.5 \pm 0.4) \cdot 10^{+3}$
7	TSNZL	f	$(6.4 \pm 3.1) \cdot 10^{+2}$	$(6.8 \pm 3.0) \cdot 10^{+2}$	$(1.2 \pm 3.1) \cdot 10^{+2}$	$(1.2 \pm 3.0) \cdot 10^{+2}$
8	TSNZH	f	$(2.0 \pm 2.3) \cdot 10^{+2}$	$(1.8 \pm 2.2) \cdot 10^{+2}$	$(1.5 \pm 2.1) \cdot 10^{+2}$	$(1.2 \pm 1.7) \cdot 10^{+2}$
9	QTL	f	$(0.6 \pm 1.0) \cdot 10^{-1}$	$(6.5 \pm 1.0) \cdot 10^{-1}$	$(7.7 \pm 1.1) \cdot 10^{-1}$	$(8.1 \pm 1.1) \cdot 10^{-1}$
10	QTH	f	$(1.8 \pm 1.0) \cdot 10^{-1}$	$(1.5 \pm 0.9) \cdot 10^{-1}$	$(0.8 \pm 0.9) \cdot 10^{-1}$	$(0.6 \pm 0.7) \cdot 10^{-1}$
11	QTEL	f	$(7.1 \pm 1.1) \cdot 10^{-1}$	$(7.3 \pm 1.1) \cdot 10^{-1}$	$(8.3 \pm 1.0) \cdot 10^{-1}$	$(0.9 \pm 1.0) \cdot 10^{-1}$
12	QTEH	f	$(1.7 \pm 1.2) \cdot 10^{-1}$	$(1.1 \pm 0.8) \cdot 10^{-1}$	$(0.5 \pm 0.7) \cdot 10^{-1}$	$(0.3 \pm 0.5) \cdot 10^{-1}$
13	te	tf	$(0.6 \pm 1.0) \cdot 10^{+9}$	$(0.2 \pm 5.1) \cdot 10^{+10}$	$(0.1 \pm 2.0) \cdot 10^{+11}$	$(1.2 \pm 1.9) \cdot 10^{+9}$
14	teh	f	$(0.8 \pm 1.2) \cdot 10^{+8}$	$(0.4 \pm 18.) \cdot 10^{+9}$	$(0.3 \pm 7.3) \cdot 10^{+10}$	$(0.3 \pm 1.2) \cdot 10^{+8}$
15	tel	f	$(4.8 \pm 7.0) \cdot 10^{+8}$	$(0.1 \pm 2.6) \cdot 10^{+10}$	$(0.7 \pm 9.3) \cdot 10^{+10}$	$(1.1 \pm 1.5) \cdot 10^{+9}$
16	CT8	t	$(3.7 \pm 1.6) \cdot 10^{+0}$	$(3.9 \pm 1.5) \cdot 10^{+0}$	$(6.3 \pm 1.3) \cdot 10^{+0}$	$(6.2 \pm 1.3) \cdot 10^{+0}$
17	MDL8	t	$(9.1 \pm 4.1) \cdot 10^{+1}$	$(8.6 \pm 3.8) \cdot 10^{+1}$	$(6.8 \pm 3.5) \cdot 10^{+1}$	$(6.1 \pm 2.4) \cdot 10^{+1}$
18	VDL8	t	$(9.7 \pm 4.2) \cdot 10^{+1}$	$(8.7 \pm 3.8) \cdot 10^{+1}$	$(4.9 \pm 2.8) \cdot 10^{+1}$	$(4.5 \pm 2.0) \cdot 10^{+1}$
19	Curve	f	$(1.4 \pm 1.7) \cdot 10^{-1}$	$(1.7 \pm 1.7) \cdot 10^{-1}$	$(-1.0 \pm 2.8) \cdot 10^{-1}$	$(-1.8 \pm 3.0) \cdot 10^{-1}$
20	Lfreq	f	$(9.9 \pm 4.5) \cdot 10^{+0}$	$(8.0 \pm 3.1) \cdot 10^{+0}$	$(6.1 \pm 4.2) \cdot 10^{+0}$	$(5.0 \pm 1.5) \cdot 10^{+0}$
21	Ltmp	t	$(1.5 \pm 1.1) \cdot 10^{+1}$	$(1.7 \pm 1.3) \cdot 10^{+1}$	$(3.4 \pm 2.1) \cdot 10^{+1}$	$(3.5 \pm 2.2) \cdot 10^{+1}$
22	MaxFreq	f	$(1.3 \pm 0.5) \cdot 10^{+1}$	$(1.0 \pm 0.4) \cdot 10^{+1}$	$(0.8 \pm 0.5) \cdot 10^{+1}$	$(0.7 \pm 0.2) \cdot 10^{+1}$
23	MimFreq	f	$(2.6 \pm 1.6) \cdot 10^{+0}$	$(2.2 \pm 1.4) \cdot 10^{+0}$	$(1.9 \pm 0.9) \cdot 10^{+0}$	$(2.0 \pm 0.8) \cdot 10^{+0}$
24	Area	tf	$(1.3 \pm 1.1) \cdot 10^{+2}$	$(1.3 \pm 1.0) \cdot 10^{+2}$	$(1.9 \pm 1.4) \cdot 10^{+2}$	$(1.7 \pm 1.1) \cdot 10^{+2}$
25	Nareas	tf	$(1.4 \pm 0.7) \cdot 10^{+0}$	$(1.4 \pm 0.9) \cdot 10^{+0}$	$(2.0 \pm 0.9) \cdot 10^{+0}$	$(1.8 \pm 0.8) \cdot 10^{+0}$
26	Tmy	tf	$(1.5 \pm 0.7) \cdot 10^{+2}$	$(1.5 \pm 0.6) \cdot 10^{+2}$	$(2.9 \pm 1.2) \cdot 10^{+2}$	$(2.7 \pm 1.3) \cdot 10^{+3}$
27	Dispersion	tf	$(2.1 \pm 4.6) \cdot 10^{+0}$	$(1.9 \pm 4.6) \cdot 10^{+0}$	$(5.9 \pm 7.7) \cdot 10^{+0}$	$(5.8 \pm 7.8) \cdot 10^{+0}$
28	DF	f	$(4.4 \pm 3.0) \cdot 10^{+0}$	$(4.0 \pm 3.6) \cdot 10^{+0}$	$(3.6 \pm 1.0) \cdot 10^{+0}$	$(3.9 \pm 1.2) \cdot 10^{+0}$
29	DFBW	f	$(1.5 \pm 1.3) \cdot 10^{+0}$	$(1.3 \pm 1.0) \cdot 10^{+0}$	$(0.9 \pm 0.8) \cdot 10^{+0}$	$(1.0 \pm 0.2) \cdot 10^{+0}$
30	FF	f	$(3.6 \pm 1.0) \cdot 10^{+0}$	$(3.7 \pm 1.2) \cdot 10^{+0}$	$(4.4 \pm 1.2) \cdot 10^{+0}$	$(4.5 \pm 1.3) \cdot 10^{+0}$
31	OI	f	$(4.7 \pm 1.5) \cdot 10^{-1}$	$(4.9 \pm 1.6) \cdot 10^{-1}$	$(5.1 \pm 1.8) \cdot 10^{-1}$	$(5.3 \pm 1.8) \cdot 10^{-1}$
32	RI	f	$(2.9 \pm 2.2) \cdot 10^{-1}$	$(3.3 \pm 2.3) \cdot 10^{-1}$	$(5.6 \pm 1.8) \cdot 10^{-1}$	$(5.3 \pm 1.6) \cdot 10^{-1}$
33	PF0	f	$(4.0 \pm 3.3) \cdot 10^{-3}$	$(4.3 \pm 3.3) \cdot 10^{-3}$	$(7.5 \pm 6.0) \cdot 10^{-3}$	$(7.3 \pm 6.5) \cdot 10^{-3}$
34	PF2	f	$(3.2 \pm 2.0) \cdot 10^{-3}$	$(3.3 \pm 2.1) \cdot 10^{-3}$	$(2.2 \pm 3.3) \cdot 10^{-3}$	$(2.5 \pm 4.0) \cdot 10^{-3}$
35	PF3	f	$(1.7 \pm 1.1) \cdot 10^{-3}$	$(1.7 \pm 1.3) \cdot 10^{-3}$	$(0.6 \pm 1.1) \cdot 10^{-3}$	$(0.5 \pm 1.2) \cdot 10^{-3}$
36	PF4	f	$(1.0 \pm 0.8) \cdot 10^{-3}$	$(8.8 \pm 8.7) \cdot 10^{-4}$	$(2.4 \pm 5.0) \cdot 10^{-4}$	$(1.6 \pm 4.0) \cdot 10^{-4}$
37	PF5	f	$(6.6 \pm 6.4) \cdot 10^{-4}$	$(5.2 \pm 7.2) \cdot 10^{-4}$	$(1.4 \pm 3.0) \cdot 10^{-4}$	$(0.9 \pm 2.1) \cdot 10^{-4}$

components of VF episodes reside in the low frequencies band, we defined a low frequency band (2 – 14 Hz) called BALO. A high frequency band (BAHI, 14 – 28 Hz) was also considered, which contained energy components of non-VF rhythms. Based on the PWV distribution, a number of temporal, spectral, and time–frequency parameters have been obtained (see Table 1, parameters from 3 to 27):

- `Pmxfreq`: Frequency where the maximum energy of the PWV occurs.
- `MaximFreq`, `MinimFreq`: Frequencies with the highest and lowest frequency content, respectively.
- `TSNZ`, `TSNZH`, `TSNZL`: Total sum of non-zero terms contained in the PWV distribution, in the BAHI and the BALO bands, respectively.
- `QTL`, `QTH`: Percentage of the total number of non-zero terms existing in the BALO and BAHI bands, respectively.
- `QTEL`, `QTEH`: Percentage of the total energy contained in the BALO and BAHI bands, respectively.
- `TE`, `TEH`, `TEL`: Total energy of the PWV distribution, in the BAHI band, and in the BALO band, respectively.
- `CT8`: The time axis of the PWV distribution is divided into eight window segments. Then, for every segment, the energy in the BALO band is measured. The `CT8` corresponds to the number of window segment that contain at least half of the energy if the total energy of the band would be equally distributed along the time axis.
- `MDL8`: Number of non-zero terms contained in the BALO band when measured at the eight window segments defined for `CT8`.
- `VDL8`: Standard deviation of the first-order derivative of `MDL8`.
- `Curve`: Curvature of the parabolic approximation performed over the number of non-zero terms at every frequency bin of spectral resolution in the BALO band.
- `Lfreq`, `Ltmp`, `MaxFreq`, `MimFreq`: These parameters quantify the components, so-called half-energy region, of the PWV distribution whose energy values fall below 50% of the maximum peak energy value. `Lfreq` and `Ltmp` represent the frequency length and the temporal length of this half-energy region, respectively. `MaxFreq` and `MimFreq` indicate the maximum and minimum frequencies that limit the half-energy region.
- `Area`, `Nareas`: `Area` gives the total number of points contained in a certain extracted half-energy region, and `Nareas` provides with the number of half-energy regions extracted in a single time–frequency representation.
- `Tmy`: Number of points between the 50% and 100% of the maximum energy value existing in the PWV.
- `Dispersion`: Difference between the maximum and the mean values of `Ltmp`.

A full detailed description of the first 27 parameters can be found in Rosado et al. (1999) and Rosado, Guerrero, Bataller, and Chorro (2001). This set of parameters was extended to include a number of spectral indices which have recently grown up in both the experimental and the clinical environments to target fibrillatory rhythms (Atienza et al., 2006; Everett, Kok, Vaughn, Moorman, & Haines, 2001; Everett, Moorman, Kok, Akar, & Haines, 2001; Sanders et al., 2005). For each window segment, the power density spectrum $P_n(f)$ (normalized by its total power) was estimated using the squared module of the Fast Fourier Transform (FFT) with a 128 samples Hamming window. Based on $P_n(f)$, the following spectral parameters have been considered (Table 1, parameters from 28 to 37):

- DF: Dominant frequency (f_d). Frequency where the maximum of $P_n(f)$ occurs.
- DFBW: Dominant frequency bandwidth ($bw(f_d)$). Difference between the upper and lower frequencies for which f_d falls to 75% of its power value.
- FF: Fundamental frequency (f_0). It is sometimes assumed that a VF episode is a near-periodic process, showing a fundamental signal period T_0. Thus, f_0 is defined as the inverse of T_0.
- PF0, PF2, PF3, PF4, PF5: Normalized power at harmonics frequency peaks. Harmonics are the frequencies corresponding to the integer multiples of f_0. Here, we consider up to the 5th harmonic, from $f_2 = 2 \times f_0$ to $f_5 = 5 \times f_0$. Then, we measure the normalized power at f_0 (1st harmonic), f_2, f_3, f_4 and f_5, which we denote by PF0, PF2, PF3, PF4 and PF5, respectively.
- OI: Organization index. Ratio of the power under harmonic peaks (up to f_4) to the total power in the BALO band.
- RI: Regularity index. Ratio of the power under $bw(f_d)$ to the total power in the BALO band.

The parameterization of ECG signal segments finally resulted in an input dataset consisting of N = 57,908 observations and 37 features. For each observation, four different groups have been considered according to different pathologies, which appeared with different prior probabilities: Normal (p_1 = 40.25%), for normal sinus rhythm; vt (p_2 = 8.84%), for ventricular tachycardia (VT) including their variants (regular VT, polymorphic VT or "torsades de pointes"); vf-Flutter (p_3 = 10.66%), for VF signal and flutter, both having the same application therapy (electric shock); and Others (p_4 = 40.25%), comprising the rest of arrhythmias. It is essential to remark that polymorphic VT is hardly distinguished of VF by means of the ECG, and for this reason the automatic discrimination between VF and VT (specially polymorphic) is a complex issue.

4. FS algorithm

In this section, we present our method for FS in SVM classifiers using BR techniques, which we call SVM-BR.

4.1. BR for SVM

BR is a computer-based method introduced by Efron in 1979 (Efron & Tibshirani, 1994), which constitutes a useful approach for nonparametric estimation of the distribution of statistical magnitudes, even when the observation set is small. We propose the use of BR to estimate the performance of SVM classifiers. This procedure can be also used to estimate SVM performance when a subset of the input data is considered, thus allowing us to compare the performance of the complete set of input variables and a reduced subset of them.

Let $\mathbf{V}$ be a set of pairs of data in a classification problem, which we call *complete model*. The dependence process between pairs of data in $\mathbf{V}$ can be estimated by using SVM, whose coefficients are

$$\alpha = [\alpha_1, \ldots, \alpha_N] = s(\mathbf{V}, C, \sigma) \tag{7}$$

where $s(\cdot)$ is the SVM optimization operator, depending on data $\mathbf{V}$ and on free parameters C and σ. The empirical risk for these coefficients is defined as the training error fraction of the set of pairs used to build the machine,

$$R_{emp} = t(\alpha, \mathbf{V}) \tag{8}$$

where $t(\cdot)$ is the empirical risk estimation operator.

A *bootstrap resample* $\mathbf{V}^* = \{(\mathbf{x}_1^*, y_1^*), \ldots, (\mathbf{x}_N^*, y_N^*)\}$ is a new data set drawn at random with replacement from sample $\mathbf{V}$. Let consider a partition of $\mathbf{V}$ in terms of the resample

$$\mathbf{V} = (\mathbf{V}_{in}^*, \mathbf{V}_{out}^*) \tag{9}$$

being $\mathbf{V}_{in}^*$ and $\mathbf{V}_{out}^*$ the subsets of samples included and excluded in the resample, respectively. Then, SVM coefficients for the resample are

$$\alpha^* = s(\mathbf{V}_{in}^*, C, \sigma) \tag{10}$$

The actual risk estimation for the resample can be obtained by taking

$$R^* = t(\alpha^*, \mathbf{V}_{out}^*) \tag{11}$$

Then, given a collection of B independent resamples, $\{\mathbf{V}^*(1), \mathbf{V}^*(2), \ldots, \mathbf{V}^*(B)\}$, the actual risk density function can be estimated by the histogram built from replicates $R^*(b)$, where $b = 1, \ldots, B$. A typical choice for B is from 100 to 500 resamples.

We now consider a reduced version of the observed data $\mathbf{W}_u$ (*incomplete model* in the following), in which the uth feature is removed from all the available observations, $\mathbf{W}_u = \left\{\left(\mathbf{x}_1^{(u)}, y_1\right), \ldots, \left(\mathbf{x}_N^{(u)}, y_N\right)\right\}$, being $\mathbf{x}_i^{(u)} \in \mathbb{R}^{d-1}$. A *paired resampling* procedure is carried out by using the same resampling set as the complete model $\mathbf{W}_u^* = \left\{\left(\mathbf{x}_1^{*,(u)}, y_1^*\right), \ldots, \left(\mathbf{x}_N^{*,(u)}, y_N^*\right)\right\}$, then yielding a bootstrap replication of the actual risk in the incomplete model

$$R_u^* = t\left(\alpha^*, \mathbf{W}_{u,out}^*\right) \tag{12}$$

Based on the aforementioned considerations, we use BR to quantify changes in the SVM performance due to the elimination of variable u. Let $\triangle R_u$ define the SVM performance difference (in terms of actual risk) between the complete model and the incomplete model when variable u is removed. Then, the statistic

$$\Delta R_u^*(b) = R_u^*(b) - R^*(b) \tag{13}$$

can be replicated at each resample $b = 1, \ldots, B$, and it represents the estimated loss due to the information in the removed variable. Accordingly, the statistic $\Delta R_u^*(b)$ can be used to evaluate the relevance (in terms of SVM performance) of variable u, as shown next.

4.2. SVM-BR algorithm

An adequate risk measurement in a classification task is the classification error probability, denoted by P_e. As stated before, the relevance of variable u can be evaluated by comparing the error probability between the complete feature dataset (denoted as $P_{e,c}$) and the incomplete model (denoted as $P_{e,u}$). To compare both magnitudes we propose the use of the statistic $\Delta P_e = P_{e,u} - P_{e,c}$ and the following hypothesis test:

- H_0: $\Delta P_e = 0$, hence variable u is not relevant;
- H_1: $\Delta P_e \neq 0$, hence variable u is relevant.

However, the distribution of ΔP_e is generally unknown, since the dependence process between pairs of data $p(\mathbf{x}_i, y_i)$ is not available. Therefore, we redefine the statistic as

$$\Delta P_e^*(b) = P_{e,u}^*(b) - P_{e,c}^*(b), \quad b = 1, \ldots, B \tag{14}$$

allowing us to estimate the distribution of test statistic $\mathbf{\Delta P}_e^*$ and compute its confidence interval, which we call *paired confidence interval* $(z_{\Delta \mathbf{P}_e^*})$. Then, for a given significance level, H_0 is fulfilled if $z_{\Delta \mathbf{P}_e^*}$ has negative values $(z_{\Delta \mathbf{P}_e^*} < 0)$ or it does contain the zero point

($z_{\Delta \mathbf{P}_e^*} \subset 0$), otherwise, the alternative hypothesis is accepted. These conditions imply that relevant variables emerge whenever their elimination results in a significant decrease in the error probability $P_{e,u}$ compared to the error probability of the complete model $P_{e,c}$, hence producing a significant increase of the statistic $\Delta \mathbf{P}_e^*$. Our proposed SVM-BR algorithm for FS is defined in Algorithm 1.

Algorithm 1: SVM-BR backward selection algorithm

1. Start with all features of the input space **V**.
2. Built B paired bootstrap resamples of the complete $\mathbf{V}^*$ and the incomplete model $\mathbf{W}_u^*$.
3. For each bootstrap sample b, and for each feature u compute the bootstrap statistic

$$\Delta P_e^*(b) = P_{e,u}^*(b) - P_{e,c}^*(b), \forall u, b = 1, \ldots, B.$$

   and calculate the 95% $z_{\Delta \mathbf{P}_e^*}$.
4. If $z_{\Delta \mathbf{P}_e^*} < 0$ for any feature u:
   - eliminate variable u.
   Otherwise, if $z_{\Delta \mathbf{P}_e^*} \subset 0$ for any feature u, then:
   - remove u with highest PCI, or
   - remove u with smallest PCI.
5. If there is any feature u for which $\mathbf{P}_{e,u}^* < \mathbf{P}_{e,c}^*$, then error probability of the complete model is redefined as:

$$\mathbf{P}_{e,c}^* = \mathbf{P}_{e,u}^*$$

6. Finish whenever every feature fulfills $z_{\Delta \mathbf{P}_e^*} > 0$. Otherwise, go to step (3).

It is worth noting that complex interactions among the input variables can be expected whenever nonlinear SVM models are built, such as collinearity (for the nonlinear case, co-information or redundant information), irrelevant or noisy variables, and subsets of variables being relevant only when interacting among them. Under these situations, $z_{\Delta \mathbf{P}_e^*}$ associated to relevant variables may also contain the zero point ($z_{\Delta \mathbf{P}_e^*} \subset 0$). For this reason, and since it has not been defined a statistic associated to the confidence interval of a statistic, our proposed backward selection procedure is based on two criteria. On the one hand, we consider u as the most irrelevant feature if it has the highest $z_{\Delta \mathbf{P}_e^*}$, H-PCI in the following. On the other hand, u is considered the most irrelevant feature if it has the smallest $z_{\Delta \mathbf{P}_e^*}$ (S-PCI). Evaluation of both criteria is achieved by means of toy examples, which are presented in the next section. Note also that the backward selection procedure defined in Algorithm 1 can be applied to the SVM-RFE algorithm by bootstraping the cost function (6).

5. Toy examples

The objective of this section is twofold. Firstly, to validate the proposed relevance criteria based on the width of the PCI, and secondly, to examine the performance of our SVM-BR algorithm by comparing it to the SVM-RFE method. We analyzed both SVM-BR and SVM-RFE algorithms by using a synthetic set of data in two different scenarios, namely, a linear and a nonlinear classification problem. Experiments consisted in selecting the most relevant features according to a predefined set of variables. FS algorithms were run for 10 random trials to avoid skewed results. In those cases where results were not reproduced in all trials, we present the variables that were selected in the higher number of trials, indicating also the number of times that those features were selected. In all simulations, we used $N = 1000$ training samples and $B = 500$ bootstrap resamples. All variables were standardized to have zero mean and standard deviation one.

5.1. Notation

Let $(\mathbf{x}_i, y_i)$ be a set of N observations and labeled data, $i = 1, \ldots, N$, where $\mathbf{x}_i \in \mathbb{R}^d$ consist of d variables or features and $y_i \in \{-1, +1\}$. In a convenient abuse of notation, we will denote the row vector $\mathbf{x}_j$ as the set of observations relative to variable j, such us $\mathbf{x}_j = \{x_{j,1}, x_{j,2}, \ldots, x_{j,N}\}$. Under these assumptions, $x_{j,i}$ refers to the jth variable of the ith observation. We denote $\mathcal{N}(\mu, \sigma)$ to be a Normal distribution with mean μ and standard deviation σ. We also denote $\mathcal{U}(a, b)$ to be a Uniform distribution in the interval (a,b), and $\mathcal{R}(r)$ a Rayleigh distribution with $r_{rms} = \sqrt{2}\sigma$.

5.2. Linear classification problem

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_5\}$ be a set of random variables, where $\mathbf{x}_1$ defines a linearly separable problem: $x_{1,i} = z + \mathcal{N}(0, \sigma_1)$, being z a random variable such as $z \in \{-2, +2\}$ and the probability of $z = 2$ or $z = -2$ is equal, for $i = 1, 2, \ldots, N$. Variables $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$ are noisy features defined as $x_{2,i} = \mathcal{N}(0, 3.5), x_{3,i} = \mathcal{U}(-0.5, 0.5)$, and $x_{4,i} = \mathcal{R}(1) - 1$, respectively. Finally, $\mathbf{x}_5$ represents a redundant variable $x_{5,i} = \mathcal{N}(0, \sigma_5) - 3x_{1,i}$. Note that the optimal separating hyperplane is $\mathbf{x}_1 = 0$, such that $y_i = +1$ if $x_{1,i} > 0$, resulting in a theoretical error probability given by Proakis (2001).

$$P_{e,t} = \frac{1}{2} \text{erfc}\left(\frac{\sqrt{2}}{\sigma_1}\right) \tag{15}$$

where erfc $(\cdot)$ represents the complementary error function. We analyzed the performance of both SVM-BR and SVM-RFE algorithms for different values of parameter $\sigma_1 = \{0.5, 1, 2.5, 5\}$, allowing us to evaluate the accuracy of both methods for different error probability working scenarios. For each value of σ_1, we implemented two sets of simulations in order to study collinearity effects. In the first set, we took $\sigma_5 = 3$ to obtain a correlation between variables $\mathbf{x}_1$ and $\mathbf{x}_5$ above 90%. In the second, we decreased this correlation by taking $\sigma_5 = 10$.

Tables 2 and 3 show the selected features obtained from both FS algorithms (SVM-BR, SVM-RFE) and the proposed relevance criteria (S-PCI, H-PCI) operating over the two linear classification problems under study ($\sigma_5 = 10$) and ($\sigma_5 = 3$), respectively. In order to compare the performance of the obtained model, we present the test error (mean and confidence intervals) over 500 trials for both the original complete model ($P_{e,c}$), and the reduced set that was finally selected ($P_{e,r}$). In addition, we include the theoretical error probability associated with the classification problem $P_{e,t}$ and the correlation coefficient R between variables $\mathbf{x}_1$ and $\mathbf{x}_5$. As shown, performances of both SVM-BR and SVM-RFE were identical for low correlation values ($\sigma_5 = 10$, Table 2). Using the S-PCI criterion, the selection procedure is optimal for all error probability working scenarios, where as H-PCI selected the collinear variable. This, however, did not significantly affect the performance of the selected model $P_{e,r}$, showing slight differences compared to the optimal values. Results for a high correlation scenario ($\sigma_5 = 3$, Table 3) were also very similar between SVM-BR and SVM-RFE, except for the most favourable case in terms of error probability ($\sigma_1 = 0.5$), where SVM-RFE selected the redundant variable $\mathbf{x}_5$ for both criteria, thus abruptly reducing performance of the algorithm. In conclusion, the S-PCI criterion presents optimal results, and our SVM-BR algorithm shows a more robust behavior than SVM-RFE. It is worth noting that the value of the SVM free parameter C was calculated once for the complete model. We checked that the optimal value of C did not vary significantly during the FS procedure, which is consistent with the fact that C does not depend on the dimension but on the signal variance (Cherkassky & Ma, 2004).

Table 2
Performance of SVM-BR and SVM-RFE algorithms in a linear classification problem and for moderate correlation values between variables $\mathbf{x}_1$ and $\mathbf{x}_5$, ($\sigma_5 = 10, N = 1000, B = 500$).

Method	Criterion	$\sigma_1 = 0.5$	$\sigma_1 = 1.0$	$\sigma_1 = 2.5$	$\sigma_1 = 5$
SVM-BR	S-PCI	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$
	H-PCI	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_5$
SVM-RFE	S-PCI	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$
	H-PCI	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_5(7)$
	$P_{e,c}$	$3.9(0.0, 100.0) \cdot 10^{-5}$	$2.4(1.5, 3.4) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.35(0.32, 0.38)$
SVM-BR	$P_{e,r}$ S-PCI	$3.4(0.0, 100.0) \cdot 10^{-5}$	$2.3(1.4, 3.2) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.34(0.32, 0.37)$
	$P_{e,r}$ W-PCI	$3.4(0.0, 100.0) \cdot 10^{-5}$	$2.3(1.4, 3.2) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.37(0.34, 0.40)$
SVM-RFE	$P_{e,r}$ S-PCI	$3.4(0.0, 100.0) \cdot 10^{-5}$	$2.3(1.4, 3.2) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.34(0.32, 0.37)$
	$P_{e,r}$ W-PCI	$3.4(0.0, 100.0) \cdot 10^{-5}$	$2.3(1.4, 3.2) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.37(0.34, 0.40)$
	$P_{e,t}$	$3.2 \cdot 10^{-5}$	$2.3 \cdot 10^{-2}$	0.21	0.34
	R	0.53	0.55	0.71	0.86

Table 3
Performance of SVM-BR and SVM-RFE algorithms in a linear classification problem and for high correlation values between variables $\mathbf{x}_1$ and $\mathbf{x}_5$, ($\sigma_5 = 3$, $N = 1000$, $B = 500$).

Method	Criterion	$\sigma_1 = 0.5$	$\sigma_1 = 1.0$	$\sigma_1 = 2.5$	$\sigma_1 = 5$
SVM-BR	S-PCI	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1$
	H-PCI	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_5(9)$	$\mathbf{x}_5(8)$
SVM-RFE	S-PCI	$\mathbf{x}_5$	$\mathbf{x}_1$	$\mathbf{x}_1$	$\mathbf{x}_1(8)$
	H-PCI	$\mathbf{x}_5$	$\mathbf{x}_1$	$\mathbf{x}_5(6)$	$\mathbf{x}_5(9)$
	$P_{e,c}$	$10.6(0.0, 100.0) \cdot 10^{-5}$	$2.5(1.5, 3.5) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.35(0.32, 0.37)$
SVM-BR	$P_{e,r}$ S-PCI	$4.2(0.0, 100.0) \cdot 10^{-5}$	$2.3(1.4, 3.3) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.34(0.31, 0.37)$
	$P_{e,r}$ W-PCI	$4.2(0.0, 100.0) \cdot 10^{-5}$	$2.3(1.4, 3.3) \cdot 10^{-2}$	$0.23(0.20, 0.26)$	$0.35(0.32, 0.38)$
SVM-RFE	$P_{e,r}$ S-CI	$3.7(2.6, 4.8) \cdot 10^{-2}$	$2.3(1.4, 3.3) \cdot 10^{-2}$	$0.21(0.19, 0.24)$	$0.34(0.31, 0.37)$
	$P_{e,r}$ W-PCI	$3.7(2.6, 4.8) \cdot 10^{-2}$	$2.3(1.4, 3.3) \cdot 10^{-2}$	$0.23(0.20, 0.26)$	$0.35(0.32, 0.38)$
	$P_{e,t}$	$3.2 \cdot 10^{-5}$	$2.3 \cdot 10^{-2}$	0.21	0.34
	R	0.90	0.92	0.95	0.98

Table 4
Performance of SVM-BR and SVM-RFE algorithms in a XOR nonlinear classification ($N = 1000, B = 500$).

Method	Criterion	$\sigma_{12} = 0.5$	$\sigma_{12} = 1.0$	$\sigma_{12} = 1.5$	$\sigma_{12} = 2.0$
SVM-BR	S-PCI	$(\mathbf{x}_1,\mathbf{x}_2)$	$(\mathbf{x}_1,\mathbf{x}_2)$	$(\mathbf{x}_1,\mathbf{x}_2)(7)$	$(\mathbf{x}_1,\mathbf{x}_2)(7)$
	H-PCI	$(\mathbf{x}_1,\mathbf{x}_2)$	$(\mathbf{x}_1,\mathbf{x}_6)(4)$	$\mathbf{x}_7(4)$	$\mathbf{x}_7(5)$
SVM-RFE	S-PCI	$\mathbf{x}_6(7)$	$\mathbf{x}_5(5)$	$\mathbf{x}_5(4)$	$\mathbf{x}_5(4)$
	H-PCI	$\mathbf{x}_6(7)$	$\mathbf{x}_3(7)$	$\mathbf{x}_3(5)$	$(\mathbf{x}_4,\mathbf{x}_5)(4)$
	$P_{e,c}$	$3.3(0.0, 14.0) \cdot 10^{-3}$	$6.3(4.6, 8.3) \cdot 10^{-2}$	$0.19(0.16, 0.23)$	$0.29(0.26, 0.34)$
SVM-BR	$P_{e,r}$ S-PCI	$8.4(0.0, 100.0) \cdot 10^{-5}$	$4.6(3.4, 6.1) \cdot 10^{-2}$	$0.17(0.15, 0.20)$	$0.30(0.28, 0.33)$
	$P_{e,r}$ W-PCI	$8.4(0.0, 100.0) \cdot 10^{-5}$	$8.2(5.9, 11.2) \cdot 10^{-2}$	$0.23(0.20, 0.27)$	$0.32(0.28, 0.36)$
SVM-RFE	$P_{e,r}$ S-PCI	$2.8(1.9, 4.0) \cdot 10^{-2}$	$5.0(4.7, 5.3) \cdot 10^{-1}$	$0.50(0.47, 0.53)$	$0.50(0.47, 0.53)$
	$P_{e,r}$ W-PCI	$2.8(1.9, 4.0) \cdot 10^{-2}$	$4.9(4.7, 5.2) \cdot 10^{-1}$	$0.50(0.47, 0.53)$	$0.50(0.47, 0.53)$
	$P_{e,t}$	$6.3 \cdot 10^{-5}$	$4.5 \cdot 10^{-2}$	0.18	0.32
	R	0.69	0.69	0.7	0.7

5.3. Nonlinear classification problem

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_7\}$ be a set of random variables, where $\mathbf{x}_1$ and $\mathbf{x}_2$ define an XOR classification problem: $x_{1,i} = z + \mathcal{N}(0, \sigma_{12})$ and $x_{2,i} = z + \mathcal{N}(0, \sigma_{12})$, being z a random variable such as $z \in \{-2, +2\}$ and the probability of $z = 2$ or $z = -2$ is equal, for $i = 1, 2, \ldots, N$. From $\mathbf{x}_3$ to $\mathbf{x}_5$ different noisy variables are introduced: $x_{3,i} = \mathcal{N}(0, 3.5)$, $x_{4,i} = \mathcal{U}(-0.5, 0.5)$ and $x_{5,i} = \mathcal{R}(1) - 1$, respectively. Collinearity is introduced with $\mathbf{x}_6$ and $\mathbf{x}_7$, defined as $x_{6,i} = 3(x_{1,i} + x_{2,i}) + \mathcal{N}(0, 2)$ and $x_{7,i} = 2(x_{1,i} + x_{2,i})^2 + \mathcal{N}(0, 2)$, respectively. Together with $\mathbf{x}_1$ and $\mathbf{x}_2$, note that both $\mathbf{x}_6$ and $\mathbf{x}_7$ are also relevant features (in weak sense (Kohavi & John, 1997)) since they contain discriminatory information and therefore they can contribute to the classification performance. The theoretical error probability for this XOR problem is given by

$$P_{e,t} = \mathrm{erfc}\left(\frac{\sqrt{2}}{\sigma_{12}}\right) \quad (16)$$

We simulated different error probability scenarios through the parameter $\sigma_{12} = \{0.5, 1, 1.5, 2\}$. Table 4 presents the selected variables for both methods and criteria. We calculated also the test error (mean and confidence intervals) over 500 trials for both the original complete model $P_{e,c}$, and the reduced set that was finally selected $P_{e,r}$. In addition, we include the theoretical error probability associated with the classification problem $P_{e,t}$ and the correlation coefficient R between variables $(\mathbf{x}_1, \mathbf{x}_2)$, and $\mathbf{x}_6$. As shown in Table 4, the SVM-BR algorithm using the S-PCI selected the optimal subset of variables for all error probability scenarios, therefore reducing the error probability compared to the complete model. Conversely, the SVM-RFE method did not behave correctly, selecting noisy variables. This behavior could be attributed to the fact that, in a nonlinear scenario, input variables are transformed to a high dimensional space (RKHS), where the SVM weight vector is defined. Therefore

Table 5
SVM performance for FV detection in terms of sensitivity (Ss) and specificity (Sp).

	FV-Flutter (Ss) (%)	Normal (Sp) (%)	Others (Sp) (%)	TV (Sp) (%)	Global (Sp) (%)
5-*fold*	74.7	99.7	99.6	65.0	95.1
Test	69.0	99.7	99.2	59.0	93.7

this weight vector cannot be directly associated to the input space variables to evaluate their relevance. Consequently, as stated in Statnikov, Hardin, and Aliferis (2006), SVM-RFE algorithm might assign higher weights to irrelevant variables than to the relevant ones.

As in the linear case, SVM free parameters C and σ just needed to be calculated once for the complete model. We also checked that optimal values of C and σ did not vary significantly during the FS procedure.

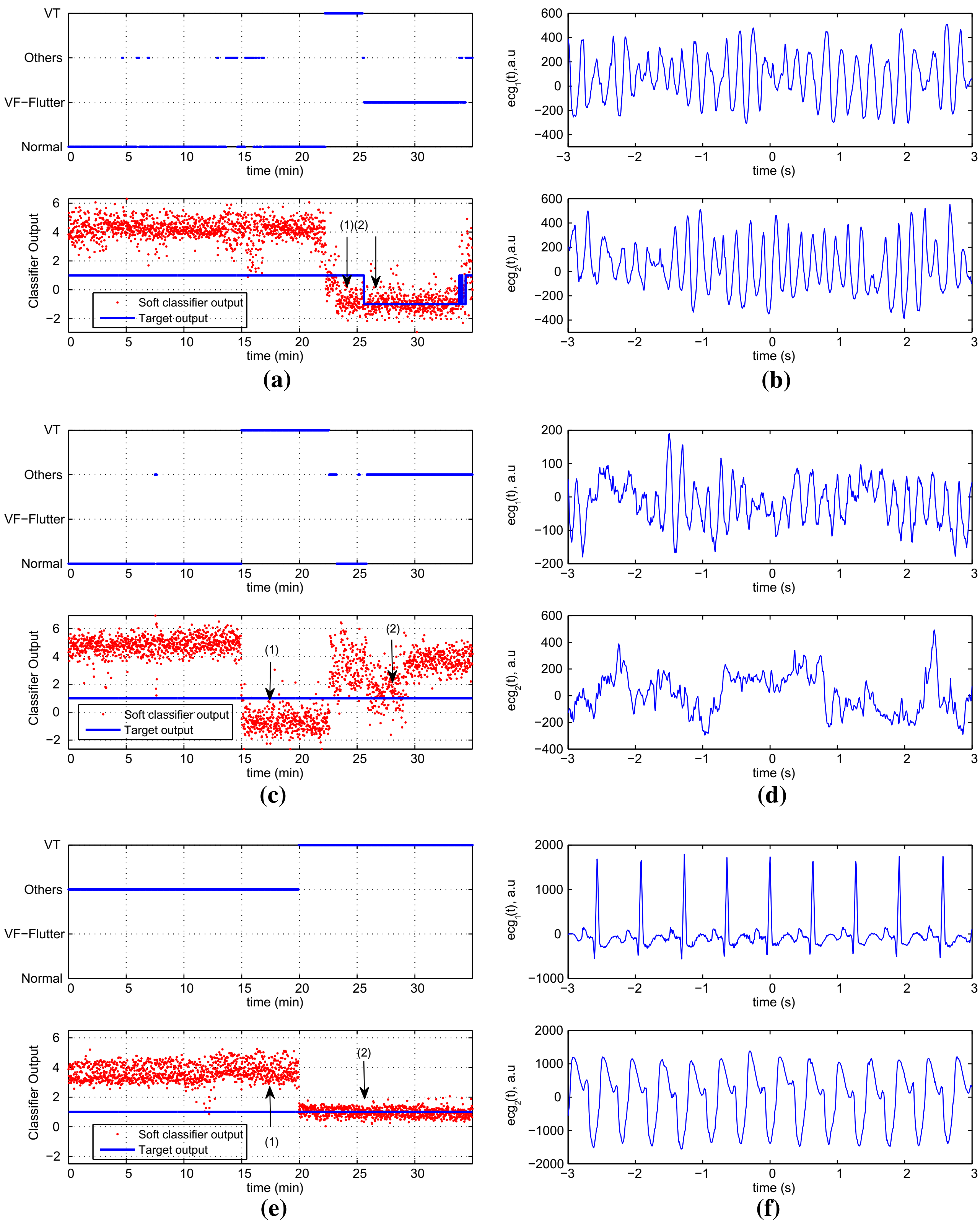

Fig. 1. Detection example of VF episodes with SVM. Panels (a), (c) and (e) show labels and the classifier output for each ECG segment; Panels (b), (d) and (f) represent six window segments ECG registered in locations marked as (1) ($ecg_1(t)$) and (2) ($ecg_2(t)$) in panels (a), (c) and (e) respectively, in arbitrary units (a.u.).

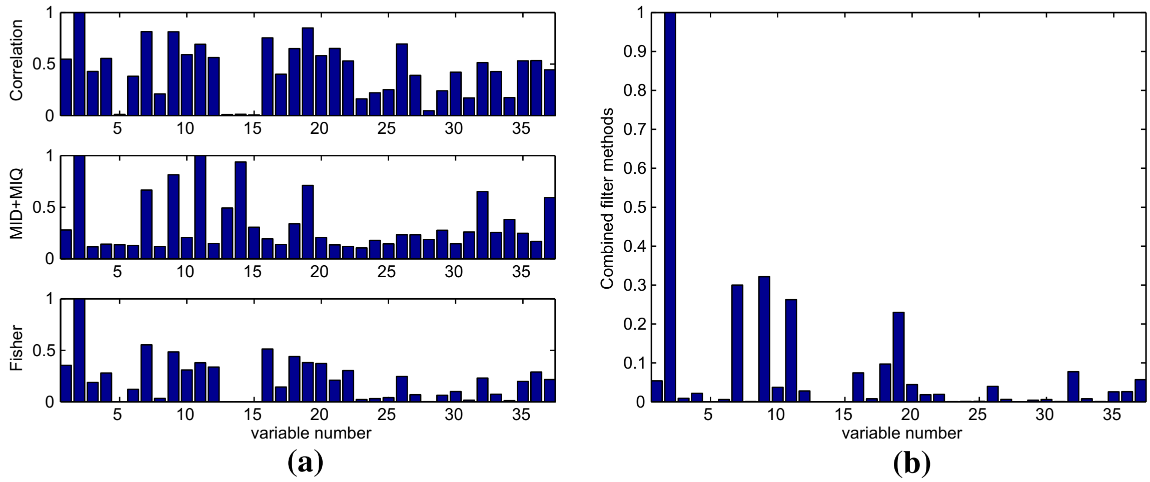

Fig. 2. Normalized variable ranking weights of different filter methods under consideration. (a) Correlation, difference and quotient mutual information (MID + MIQ) and Fisher criteria. (b) Combination of filters methods.

Based on the above presented results, we propose the SVM-BR method using the S-PCI criterion as the FS algorithm to analyze the relevance of extracted ECG parameters for VF detection.

6. Results on VF databases

In this section we analyze the proposed SVM-BR algorithm in the problem of VF detection. We first characterize the complete set of temporal, spectral, and time–frequency ECG parameters by examining the performance of SVM classifiers for detecting VF. Then, we study the combination of filter methods to reduce the high-dimensional input space set. Finally, our SVM-BR algorithm is applied to the resulting set of ECG parameters after filtering.

6.1. SVM performance

Given that our purpose was VF detection, a binary output target was considered for discriminating VF episodes from other rhythms (labeled as {−1} and {+1}, respectively). Conventional cross-validation strategy (*n-fold* with $n = 5$) was followed for setting the free parameters of the SVM. Due to the large amount of available ECG 1-s segments, the training set was defined as a random subset (20%) of the original data, and the remaining samples were used as test set, suitable for measuring the generalization capabilities of the classifier. Unbalance between the examples of each class was corrected by pre-weighting *C* free parameter for the two different classes according to their priors. Additionally, we decided to use the complete databases, and not selected segments, as far as these are conventionally used standard databases.

As shown in Table 5, acceptable VF detection capabilities were obtained, nevertheless, most significant errors were present in a number of VT segments. Fig. 1 shows application examples of SVM for VF detection. The upper parts of Fig. 1(a), (c) and (e) show the label of each ECG segment, whereas the lower parts represent the classifier output. Fig. 1, panels (b), (d) and (f) represent two six-window segment ECGs registered at locations (1) and (2) marked with arrows in Fig. 1(a), (c) and (e), respectively. In the first example, Fig. 1(a) shows the evolution of the soft classifier output towards a VF episode, where the transition from normal sinus rhythm to VF is progressive. This transition interval corresponds to a VT episode that precedes the VF onset. The upper part of Fig. 1(b) shows an ECG record labeled as VT according to the annotation file, where as the lower part depicts an ECG recording annotated as VF. Both records, however, show a similar morphology and, in the absence of a gold standard to discriminate FV, their annotation might be different depending on the specialist. This discrepancy reflects the difficulties when discriminating between VT and VF. Fig. 1(c) represents an example of erroneous discrimination between VT and VF, where VT samples are labeled as VF. Representative ECGs registered at locations (1) y (2) are presented in Fig. 1(d). A correct discrimination between VT and VF is shown in Fig. 1(e). However, the corresponding ECG (location (2)) presents a quite regular morphology, indicating a monomorphic VT for which specialist would clearly differentiate from VF. On the other hand, note the differences in those ECG recordings labeled as OTHERS (panels (d) and (f)), indicating the broad spectrum of pathologies considered within this group.

6.2. Filter methods performance

Following a similar approach as in Cho, Baek, Youn, Jeong, and Taylor (2009), we applied filter methods to reduce the high-dimensional input space data set. Specifically, we considered a combined strategy of filter methods, accounting for second order methods (correlation criterion), mutual information methods (difference and quotient schemes), and the maximum separability Fisher criterion. Fig. 2(a) shows the normalized variable ranking weights obtained from the three filter methods under consideration for the complete set of ECG features. We multiplied these variable ranking

Table 6
SVM classifier performance for FV detection in terms of Ss and Sp after using a combination of filter methods.

	FV-FLUTTER (Ss) (%)	NORMAL (Sp) (%)	OTHERS (Sp) (%)	TV (Sp) (%)	Global (Sp) (%)
5-*fold*	74.1	99.8	99.5	62.0	94.7
Test	69.7	99.7	99.1	57.0	93.5

Table 7
SVM classifier performance for FV detection using the selected variables obtained from our SVM-BR method.

	FV-Fsc lutter (Ss) (%)	NORMAL (Sp) (%)	OTHERS (Sp) (%)	TV (Sp) (%)	Global (Sp) (%)
5-*fold*	72.1	99.7	99.3	57.0	93.9
Test	71.9	99.7	99.2	56.6	93.8

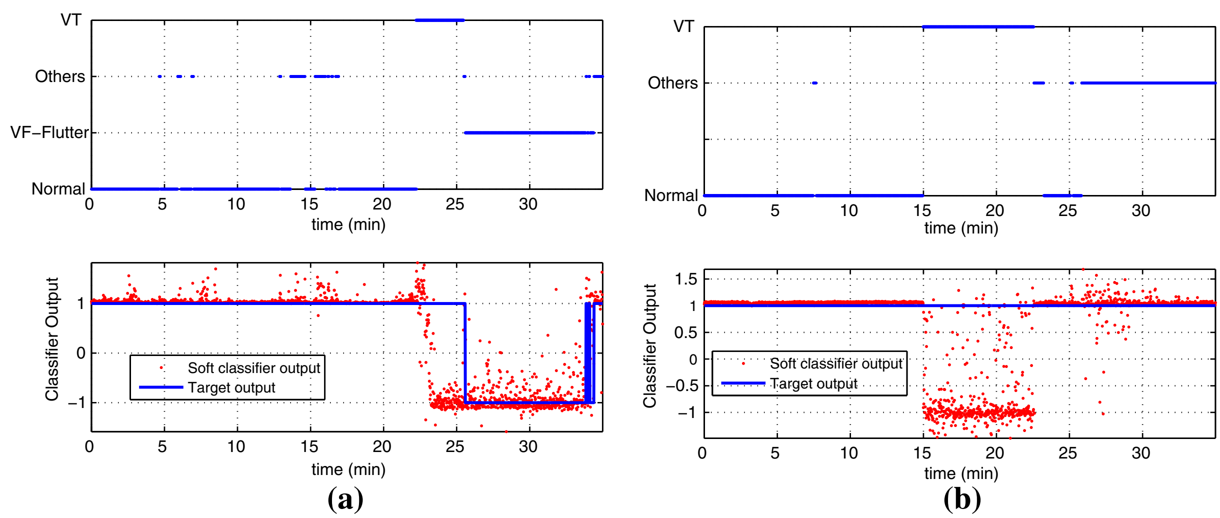

Fig. 3. Detection example of VF episodes with SVM using a reduced set of selected ECG parameters.

by each other and normalized the resultant weights, as presented in Fig. 2(b). Then, variables under a threshold level set at $1 \cdot 10^{-3}$ were removed. Referring to Table 1, discarded variable are numbered as {5,8,13,14,15,23,28,31,34}.

The reduction of the input space dimension using a combination of filter methods did not reduce the performance of the VF detection, as shown in Table 6. These results ensure that discriminatory information has not been eliminated after removing variables. However, it highlights the great amount of redundant information that it is conveyed by the complete set of variables.

6.3. SVM-BR method performance

We applied our SVM-BR method to the resultant input set of features after filtering. Due to the large amount of observations ($N = 57{,}908$), we constructed bootstrap resamples of reduced size ($N_B = 5000$) and $B = 100$ resamples iterations. Referring to Table 1, the finally selected variables were: `RatioVar`, `QTL`, and `Curve`. The performance of SVM for VF detection using this reduced set of variables is presented in Table 7.

Note that, after applying our SVM-BR algorithm, the original input space of variables has been drastically reduced while improving the performance of the VF detector compared to previous examples (see Test results). As stated before, this result evidences that the original set of data consists principally of redundant variables. On the other hand, it proves that the application of our FS algorithm is useful to select a reduced set of variables which might be used to develop new VF detectors. Detection examples using the selected set of variables are presented in Fig. 3(a) and (b), which correspond to the examples depicted in Fig. 1(a) and (c), respectively. It can be seen, that both classes can be distinguished more clearly, reducing the number of possible misclassified outliers.

7. Discussion and conclusions

A FS procedure has been proposed for its application to VF automatic detection, which compares the performance of a classifier for a complete set of data and a reduced subset. Comparison is achieved by using a hypothesis test based on nonparametric BR, and the confidence interval width is contrasted to discard variables whenever the decision statistic lacks of discriminant capabilities, a common situation in highly redundant variables scenarios.

7.1. SVM-BR algorithm

The analysis of our FS algorithm on synthetic data has shown its good behavior when working with noisy and collinear variables. Previous studies on the usefulness of SVM for developing FS algorithms (Guyon et al., 2002; Ishak & Ghattas, 2005; Rakotomamonjy, 2003; Weston, Elisseeff, Schölkopf, & Tipping, 2003) follow a similar methodology, the selection process relying on evaluating the differences on a performance measurement when a subset of input variables is removed. Usual performance measurements are either the norm of the classification hyperplane, $\|\mathbf{w}\|^2$, or some upper bound of the structural risk. Nevertheless, these performance measurements can be affected by the data variability, hence making necessary some relevance criterion exploiting the statistical nature of the objective function. In this setting, Ishak and Ghattas (2005) proposed the use of BR over the target functions defined in Guyon et al. (2002) and Rakotomamonjy (2003), aiming to improve the relevance criterion estimation. Resampling, however, is not used therein as a tool for defining a hypothesis test evaluating the relevance of a feature set. Hence, our FS proposal is new with respect to methods to date.

The SVM-BR algorithm has demonstrated to be very efficient when working with high-dimensional complex scenarios, having a great amount of redundant variables. The performance of our FS method over the AHA and MIT-BIH databases using the selected set of variables has been improved in comparison to the original set, highlighting the potential of our algorithm to extract relevant features. In the case of the detection of VF episodes, our SVM-BR can be extended to analyze ECG parameters defined in the literature and to provide a reduced set of discriminatory measurements, thus decreasing the computational requirements to develop real-time VF detectors.

7.2. Limitations of the study

The main limitation of our FS method, generally shared by methods based on SVM, is their dependence on the free parameters. The search of an adequate working point for SVM classification is crucial ir order to ensure the FS working properly. However, after the free parameters are fixed, we do not need to re-train the machine during the selection procedure. The effect of re-training after feature removing has been evaluated before, concluding that it is not generally necessary (Guyon et al., 2002; Ishak & Ghattas, 2005; Rakotomamonjy, 2003). With respect to the computational burden of our algorithm, training process is made just once (for each working scenario), yet this is a costly procedure, specially for nonlinearly separable problems. The burden due to BR is high, hence our FS algorithm can be considered as computationally intensive.

We analyzed continuos ECG signals by means of 1-s window segments to mimic real-time acquisition procedures in EADs and monitoring systems, such as Holter devices. As suggested by others (Amann et al., 2005), a larger window length for processing might improve the performance of detection algorithms. Nevertheless, this second-by-second detection is capable of describing the pathology evolution at the higher episode level, thus demonstrating that SVM constitute an adequate tool for developing VF detection algorithms.

7.3. VF vs VT discrimination

With respect to VF detection, the SVM algorithm can correctly discriminate it from different pathologies, but it misclassifies VF-Flutter as VT. Given that VT is often an early stage of VF, it is well known that VT-VF discrimination is a complex problem. In fact, flutter episodes, which are here included in VF, are often considered as a kind of VT. Results for VT and VF in the literature should be taken with caution. Some of them use previously selected segments of VT and VF for evaluating the performance of their algorithms (Thakor et al., 1990), and others present the comparison between VT–VF and sinus rhythm (Jekova, 2000). However, when complete and non pre-selected ECG recordings are used, sensitivity and specificity in VF detection are around 80% (Amann et al., 2005). Accordingly, our VF detection method can be considered as acceptable, given that we did not pre-select the episodes, and more, sometimes discrepancies can be raised between the databases labels and other specialists opinion on the episodes. Hence, the success rate can be further improved by means of two alternatives. First, aiming to improve VT vs VF discrimination, the labels of VT and VF could be revised by a committee of specialists. This has not been addressed in this work because we wanted to obtain the performance of our method in the actual standard of databases for discrimination algorithms. Second, more sophisticated detection logic could be built, by combining previously proposed techniques for normal rhythm discrimination (Rosado et al., 2001; Rosado-Muñoz et al., 2002) or by developing SVM algorithms specialists on VT–VF discrimination. Another possible future development consists of the use of combination of kernels devoted to temporal, spectral, and time-spectral parameters.

7.4. Feature extraction and VF discrimination system

It is widespread accepted that systems for VF detection must be focused at yielding 100% sensitivity for VF, and then trying to increase the specificity for improving patient's life quality, and in fact, implantable devices follow this guideline in their design. We have proposed here a pattern recognition scheme with improved feature selection as the basis for a VF detection system, and hence, we have devoted our effort to the optimization at the feature extraction stage. The computational burden of the process in its current state is still high as for being introduced in an detection device or system, but our purpose in this research line is to be able to merely optimize the feature selection stage. The 100% sensitivity must be required at a higher level stage, using the 1-s optimized features but using additional episode logic detection, in order to consider the features in a larger time window (typically 6–8 s.), and taking into account information such as the consecutive presence of VF in a certain number of 1-s windows, or other episode-level considerations. Such (more complex) scheme is out of the scope of the paper. Previous work for VF detection in the literature often uses (sometimes implicitly) this same approach. There are previous works that focus on increasing the sensitivity and specificity of their detection simultaneously, and reporting sensitivities lower than 100% required for system implementation. This is acceptable as far as we keep in mind that the final system must provide an episode detection logic yielding 100% sensitivity, and as high as possible specificity (Amann et al., 2005).

7.5. Conclusions

A novel FS algorithm has been defined based on SVM classifiers and BR techniques. Results have shown good performance both in toy examples and in the analysis of AHA and MIT-BIH databases for detecting VF. Further extensions of this work account for improving FV-VT discrimination and analyzing potential discriminatory ECG parameters to develop real-time VF detectors.

Acknowledgments

This work has been partially supported by Research Projects URJC-CM-2010-CET-4882 from Comunidad de Madrid, TEC2010-19263/TCM from the Spanish Ministry of Science and Innovation and TSI-020100-2009-332 from the Spanish Ministry of Industry, Tourism and Commerce.

References

Afonso, V. X., & Tompkins, W. J. (1995). Detecting ventricular fibrillation. *IEEE Engineering in Medicine and Biology, 14*, 152–159.

American Heart Association. Available from http://www.americanheart.org (Accessed: 17.04.10).

Amann, A., Tratnig, R., & Unterkofler, K. (2005). Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators. *Biomedical Engineering Online, 4*.

Atienza, F., Almendral, J., Moreno, J., Vaidyanathan, R., Talkachou, A., Kalifa, J., et al. (2006). Activation of inward rectifier potassium channels accelerates atrial fibrillation in humans: Evidence for a reentrant mechanism. *Circulation, 114*, 2434–2442.

Aubert, A. E., Denys, B. C., Ector, H., & Geest, H. D. (1982). Fibrillation recognition using autocorrelation analysis. In *IEEE computers in cardiology*, (pp. 477–489).

Barro, S., Ruiz, R., Cabello, D., & Mira, J. (1989). Algorithmic sequential decision making in the frequency domain for life threatening ventricular arrhythmias and imitative artifacts: A diagnostic system. *Journal of Biomedical Engineering, 11*, 320–328.

Baykal, A., Ranjan, R., & Thakor, N. V. (1997). Estimation of the ventricular fibrillation duration by autoregressive modeling. *IEEE Transactions on Biomedical Engineering, 44*, 349–356.

Beck, C. S., Pritchard, W. H., Giles, W., & Mensah, G. (1947). Ventricular fibrillation of long duration abolished by electric shock. *Journal of the American Medical Association, 135*, 985–986.

Bi, J., Bennett, K. P., Embrechts, M., Breneman, C. M., Song, M., Guyon, I., et al. (2003). Dimensionality reduction via sparse support vector machines. *Journal of Machine Learning Research, 3*, 1229–1243.

Blum, A., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence, 97*, 245–271.

Camps-Valls, G., Rojo-Álvarez, J. L., & Martínez-Ramón, M. (2007). *Kernel methods in bioengineering, communications and image processing*. Hershey, PA, USA: Idea Group Inc.

Chen, S. W., Clarkson, P. M., & Fan, Q. (1996). A robust sequential detection algorithm for cardiac arrhythmia classification. *IEEE Transactions on Biomedical Engineering, 43*, 1120–1125.

Chen, S., Thakor, N. V., & Mower, M. M. (1987). Ventricular fibrillation detection by a regression test on the autocorrelation function. *Medical and Biological Engineering and Computing, 25*, 241–249.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks, 17*, 113–126.

Cho, H. W., Baek, S., Youn, E., Jeong, M., & Taylor, A. (2009). A two-stage classification procedure for near-infrared spectra based on multi-scale vertical energy wavelet thresholding and SVM-based gradient-recursive feature elimination. *Journal of the Operational Research Society, 60*, 1107–1115.

Claasen, T. A. C. M., & Mecklenbrauker, W. F. G. (1980). The Wigner distribution – A tool for time-frequency signal analysis; part III: relations with other time-frequency signals transformations. *Philips Journal of Research, 35*, 372–389.

Clayton, R. H., & Murray, A. (1998). Comparison of techniques for time-frequency analysis of the ECG during human ventricular fibrillation. In *IEE proceedings science, measurement and technology* (Vol. 145, pp. 301–306).

Clayton, R. H., Murray, A., & Campbell, R. W. (1993). Comparison of four techniques for recognition of ventricular fibrillation from the surface ECG. *Medical and Biological Engineering and Computing, 31*, 111–117.

Clayton, R. H., Murray, A., & Campbell, R. W. (1994). Recognition of ventricular fibrillation using neural networks. *Medical and Biological Engineering and Computing, 32*, 217–220.

Clayton, R. H., Murray, A., & Campbell, R. W. (1995). Evidence for electrical organization during ventricular fibrillation in the human heart. *Journal of Cardiovascular Electrophysiology, 6*, 616–624.

Davidenko, J. M., Pertsov, A. V., Salomonsz, R., Baxter, W., & Jalife, J. (1992). Stationary and drifting spiral waves of excitation in isolated cardiac muscle. *Nature, 355*, 349–351.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. New York, NY, USA: Chapman and Hall.

Everett, T. H., Kok, L. C., Vaughn, R. H., Moorman, J. R., & Haines, D. E. (2001). Frequency domain algorithm for quantifying atrial fibrillation organization to increase defibrillation efficacy. *IEEE Transactions on Biomedical Engineering, 48*, 969–978.

Everett, T. H., Moorman, J. R., Kok, L. C., Akar, J. G., & Haines, D. E. (2001). Assessment of global atrial fibrillation organization to optimize timing of atrial defibrillation. *Circulation, 103*, 2857–2861.

Faddy, S. C. (2006). Reconfirmation algorithms should be standard of care in automated external defibrillators. *Resuscitation, 68*, 409–415.

Forster, F. K., & Weaver, W. D. (1982). Recognition of ventricular fibrillation, other rhythms and noise in patients developing sudden cardiac death. *IEEE computers in cardiology*, (pp. 245–248).

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*, 389–422.

Herschleb, J. N., Heethaar, R. M., de Tweel, I. V., Zimmerman, A. N. E., & Meijler, F. L. (1979). Signal analysis of ventricular fibrillation. *IEEE computers in cardiology*, (pp. 49–54).

Ishak, A.B., & Ghattas, B. (2005). An efficient method for variable selection using svm-based criteria. *Institut de Mathé matiques de Luminy*, preprint.

Jack, C. M., Hunter, E. K., Pringle, T. H., Wilson, J. T., Anderson, J., & Adgey, A. A. (1986). An external automatic device to detect ventricular fibrillation. *European Heart Journal, 7*, 404–411.

Jalife, J., Gray, R. A., Morley, G. E., & Davidenko, J. M. (1998). Evidence for electrical organization during ventricular fibrillation in the human heart. *Chaos, 8*, 79–93.

Jekova, I. (2000). Comparison of five algorithms for the detection of ventricular fibrillation from the surface ECG. *Physiological Measurement, 21*, 429–439.

Jekova, I., & Mitev, P. (2002). Detection of ventricular fibrillation and tachycardia from the surface ECG by a set of parameters acquired from four methods. *Physiological Measurement, 23*, 629–634.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273–324.

Kuo, S., & Dillman, R. (1978). Computer detection of ventricular fibrillation. *IEEE computers in cardiology*, (pp. 2747–2750).

Macfarlane, P. W., & Veitch, T. D. (Eds.). (1989). *Comprehensive Electrocardiology Theory and practice in health and disease*. UK: Pergamon Press.

Mirowski, M., Mower, M. M., & Reid, P. R. (1980). The automatic implantable defibrillator. *American Heart Journal, 100*, 1089–1092.

Massachusetts Institute of Technology, MIT-BIH malignant ventricular arrhythmia database, Accessed 17.04.2010.

Moe, G. K., Abildskov, J. A., & Han, J. (1964). Factors responsible for the initiation and maintenance of ventricular fibrillation. In B. Surawicz, & E. Pellegrino (Eds.), Sudden Cardiac Death. New York: Grune and Stratton.

Murray, A., Campbell, R. W. F., & Julian, D. G. (1985). Characteristics of the ventricular fibrillation waveform. *IEEE computers in cardiology*, (pp. 275–278).

Neumann, J., Schnörr, C., & Steidl, G. (2005). Combined SVM-based feature selection and classification. *Machine Learning, 61*, 129–150.

Neurauter, A., Eftestol, T., Kramer-Johansen, J., Abella, B., Sunde, K., Wenzel, V., et al. (2007). Prediction of countershock success using single features from multiple ventricular fibrillation frequency bands and feature combinations using neural networks. *Resuscitation, 73*, 253–263.

Nolle, F. M., Bowser, R. W., Badura, F. K., Catlett, J. M., Gudapati, R. R., Hee, T. T., et al. (1989). Evaluation of frequency-domain algorithm to detect ventricular fibrillation in the surface electrocardiogram. *IEEE computers in cardiology*, (pp. 337–340).

Nygards, M. E., & Hulting, J. (1978). Recognition of ventricular fibrillation utilizing the power spectrum of the ECG. *IEEE computers in cardiology*, (pp. 393–397).

Osowski, S., Hoai, L., & Markiewicz, T. (2004). Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Transactions on Biomedical Engineering, 51*, 582–589.

Pardey, J. (2007). Detection of ventricular fibrillation by sequential hypothesis testing of binary sequences. *IEEE computers in cardiology*, (pp. 573–576).

Proakis, J. G. (2001). *Digital communications* (4th ed.). McGraw-Hill [International editions].

Rakotomamonjy, A. (2003). Variable selection using SVM based criteria. *Journal of Machine Learning Research, 3*, 1357–1370.

Ribeiro, B., Marques, A., Henriques, J., & Antunes, M. (2007). Premature ventricular beat detection by using spectral clustering methods. *IEEE computers in cardiology*, (pp. 149–152).

Rosado, A., Serrano, A., Martínez, M., Soria, E., Calpe, J., & Bataller, M. (1999). Detailed study of time-frequency parameters for ventricular fibrillation detection. In *Fifth conference of the European Society for Engineering and Medicine (ESEM)* (pp. 379–380).

Rosado, A., Bataller, M., Vicente, J., Guerrero, J., Chorro, J., & Francés, J. (2000). VF detection method based on a fast real-time algorithm. In *World congress on medical physics and biomedical engineering* (pp. 50–54).

Rosado, A., Guerrero, J., Bataller, M., & Chorro, J. (2001). Fast non-invasive ventricular fibrillation detection method using pseudo Wigner–Ville distribution. *IEEE computers in cardiology*, (Vol. 28, pp. 237–240).

Rosado-Muñoz, A., Camps-Valls, G., Guerrero-Martínez, J., Francés-Villoria, J. V., Muñoz-Marí, J., & Serrano-López, A. J. (2002). Enhancing feature extraction for VF detection using data mining techniques. *IEEE computers in cardiology* (pp. 237–240).

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*, 2507–2517.

Salcedo-Sanz, S., Camps-Valls, G., Pérez-Cruz, F., Sepulveda-Sanchís, J., & Bousoño-Calzón, C. (2004). Enhancing genetic feature selection through restricted search and Walsh analysis. *IEEE Transactions on System, Man and Cybernetics Part C, 24*, 398–406.

Sanders, P., Berenfeld, O., Hocini, M., Jaïs, P., Vaidyanathan, R., Hsu, L. F., et al. (2005). Spectral analysis identifies sites of high-frequency activity maintaining atrial fibrillation in humans. *Circulation, 112*, 789–797.

Statnikov, A., Hardin, D., & Aliferis, C. (2006). Using SVM weight-based methods to identify causally relevant and non-causally relevant variables. In *Neural information processing systems (NIPS), workshop on causality and feature selection* (pp. 129–150).

Thakor, N. V. (1984). From Holter monitors to automatic defibrillators: developments in ambulatory arrhythmia monitoring. *IEEE Transactions on Biomedical Engineering, 31*, 770–778.

Thakor, N. V., Zhu, Y. S., & Pan, K. Y. (1990). Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. *IEEE Transactions on Biomedical Engineering, 37*, 837–843.

Ubeyli, E. D. (2008). Usage of eigenvector methods in implementation of automated diagnostic systems for ECG beats. *Digital Signal Processing, 18*, 33–48.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag.

Weston, J., Elisseeff, A., Schölkopf, B., & Tipping, M. (2003). Use of the zero norm with linear models and kernel methods. *Journal of Machine Learning Research, 3*, 1439–1461.

White, R., Asplin, B., Bugliosi, T., & Hankins, D. (1996). High discharge survival rate after out-of-hospital ventricular fibrillation with rapid defibrillation by police and paramedics. *Annals of Emergency Medicine, 28*, 480–485.

Yakaitis, R. W., Ewy, G. A., & Otto, C. W. (1980). Influence of time and therapy on ventricular fibrillation in dogs. *Critical Care Medicine, 8*, 157–163.

Zhang, Z., Lee, S., & Lim, J. (2008). Discrimination of ventricular arrhythmias using NEWFM. In *AIRS* (pp. 176–183).

Zhang, X. S., Zhu, Y. S., Thakor, N. V., & Wang, Z. Z. (1999). Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Transactions on Biomedical Engineering, 46*, 548–555.