Heartbeat Time Series Classification With Support Vector Machines

Argyro Kampouraki, George Manis, and Christophoros Nikou, *Member, IEEE*

***Abstract*—In this study, heartbeat time series are classified using support vector machines (SVMs). Statistical methods and signal analysis techniques are used to extract features from the signals. The SVM classifier is favorably compared to other neural network-based classification approaches by performing leave-one-out cross validation. The performance of the SVM with respect to other state-of-the-art classifiers is also confirmed by the classification of signals presenting very low signal-to-noise ratio. Finally, the influence of the number of features to the classification rate was also investigated for two real datasets. The first dataset consists of long-term ECG recordings of young and elderly healthy subjects. The second dataset consists of long-term ECG recordings of normal subjects and subjects suffering from coronary artery disease.**

***Index Terms*—Feature extraction, heartbeat time series, heart rate variability (HRV), support vector machine (SVM).**

I. Introduction

HEART RATE Variability analysis is based on measuring the variability of heart rate signals, and more specifically, variations per unit of time of the number of heartbeats (also referred to as the RR interval, since it is the time interval between consecutive R points of the QRS complex of the electrocardiogram). A large value of this index reveals a complicated system that can response better to a wide variety of conditions. Thus, a healthy person usually presents large values of HRV, while a decreased value may indicate pathological cases. HRV analysis has gained significant clinical attention as can be seen from the large number of research efforts of the past two decades. Several categorizations for heart rate variability measures were proposed. Guidelines for standards are summarized in [1], a summary of measures and models is presented in [2], and a review examining the physiological origins and mechanisms of heart rate may be found in [3].

Several techniques have been proposed for the investigation of HRV time series. Among them, spectral methods [4] based on fast Fourier transform (FFT) or standard autoregressive modeling, nonlinear approaches, including Markov modeling [5], entropy-based metrics [6], the mutual information measure [7], and probabilistic modeling [8] are widely used. The application of the Karhunen–Loëve transformation [9] and modulation analysis [10] have also been considered.

Artificial intelligence and machine learning methods constitute a powerful tool in HRV analysis. Radial basis function networks were applied for learning and predicting the HRV dynamics [11]. In [12], neural networks were used as a prediction and approximation tool for HRV analysis and the mean prediction error was used as a HRV index. In [13], coronary disease risk was predicted based on short-term RR interval measurements.

In this study, we investigate the potential benefit of using support vector machine (SVM) learning [14], [15] to classify heart rate signals. In order to enforce independence of the results of our study to the recordings and their physiological conditions, we have experimented with two different datasets. The first dataset is available on the Web [16] and consists of recordings acquired from young and elderly subjects. The second dataset consists of normal subjects and subjects suffering from coronary artery disease [17].

Support vector classifiers are based on recent advances in statistical learning theory [18]. They use a hypothesis space of linear functions in a high-dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. In the last decade, SVM learning has found a wide range of applications [19], including image segmentation [20] and classification [21], object recognition [22], image fusion [23], and stereo correspondence [24]. More recently, SVMs have been employed in several applications in biomedicine: gait degeneration due to age [25], EEG signal classification [26], brain computer interfacing (BCI) [27], [28], analysis and prediction of scoliosis [29], [30], electrogastrogram analysis [31], and color Doppler echocardiography [32]. Relevant studies involving SVM and heart rate time series are the Hermite characterization of QRS complex [33], where a heartbeat is characterized as normal or abnormal and the detection of *risky* situations for fetal assessment [34]. Previous research of our group on HRV was presented in [35] and [36]. This study investigates the problem further, examines a larger number of HRV computation methods, extracts more features, and uses different datasets. It also compares SVM with other neural network-based classifiers and examines their robustness to noisy data.

Heart rate variability analysis is applied to the estimation of the autonomic nervous balance, to the estimation of the stress or relaxation condition, and to the evaluation of mental or physiological workload. All of these issues may be handled using classification methods. In machine learning theory, SVMs are considered as being the state-of-the-art classifier. Therefore, they could be employed in such situations as the problem addressed here. Since they have not been used for the aforesaid cases, their performance is still an open issue.

Manuscript received November 11, 2007; revised April 14, 2008. First published August 4, 2008; current version published July 6, 2009.

The authors are with the Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece (e-mail: akampour@cs.uoi.gr; manis@cs.uoi.gr; cnikou@cs.uoi.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org

Digital Object Identifier 10.1109/TITB.2008.2003323

The rest of the paper is structured as follows: Section II presents a background on SVMs and Section III discusses the most common HRV features that are the input to the SVM classifier. Description of the datasets employed in our experimentation is provided in Section IV, numerical experiments are presented in Section V, and conclusions are drawn in Section VI.

II. Background on SVMs

Support vector learning strategy is a principled and very powerful method that has outperformed most other systems in a wide variety of applications [19]. The learning machine is given a training set of examples (or inputs), belonging to two classes, with associated labels (or output values). The examples are in form of attribute vectors and the SVM finds the hyperplane separating the input data and being furthest from both convex hulls. If the data are not linearly separable, a set of slack variables is introduced representing the amount by which the linear constraint is violated by each data point.

In this study, we are concerned with a two-class pattern classification problem. Let vector $\mathbf{x} \in \Re^n$ denote a pattern to be classified and let scalar y denote its class ($y \in \{\pm 1\}$). Also let $\{(\mathbf{x}_i, y_i),\ \ i = 1, \ldots, l\}$ denote a set of l training examples. The problem is how to construct a decision function $f(\mathbf{x})$ that correctly classifies an input pattern that is not necessarily in the training set.

A. Linear SVM Classifiers

If the training patterns are linearly separable, there exists a linear function of the form

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{1}$$

such that $y_i f(\mathbf{x}_i) \geq 0$, or $f(\mathbf{x}_i) \geq 0$ for $y_i = +1$ and $f(\mathbf{x}_i) < 0$ for $y_i = -1$. Vector $\mathbf{w}$ and scalar b represent the hyperplane $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$ separating the two classes.

While there may exist many hyperplanes separating the two classes, the SVM classifier finds the hyperplane that maximizes the separating margins between the two classes [14], [15]. This hyperplane can be found by minimizing the cost function

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}\|\mathbf{w}\|^2 \tag{2}$$

subject to the separability constraints

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \quad i = 1, \ldots, l. \tag{3}$$

If the training data is not completely separable by a hyperplane, a set of slack variables $\xi_i \geq 0$, $i = 1, \ldots, l$ is introduced that represents the amount by which the linearity constraint is violated

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, l. \tag{4}$$

In that case, the cost function is modified to take into account the extent of the constraint violations. Hence, the function to be minimized becomes

$$J(\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \tag{5}$$

subject to the constraints in (4). Here, C gives the significance of the constraint violations with respect to the distance between the points and the hyperplane and ξ is a vector containing the slack variables. The cost function in (5) is called *structural risk* and is a tradeoff between the empirical risk (the training errors reflected by the second term) with model complexity (the first term) [37]. The purpose of using model complexity to constrain the optimization of empirical risk is to avoid overfitting, a situation in which the decision boundary corresponds to the training data, and thereby, fails to perform well on data outside the training set.

The problem in (5) with the constraints in (4) can be solved by introducing Lagrange multipliers. With some manipulation, it can be shown that the vector $\mathbf{w}$ is formed by linear combination of the training vectors

$$\mathbf{w} = \sum_{i=1}^{l}\alpha_i y_i \mathbf{x}_i \tag{6}$$

where $\alpha_i \geq 0, \quad i = 1, \ldots, l$ are the Lagrange multipliers associated with the constraints in (4). The Lagrange multipliers are solved for the dual problem of (5), which is expressed as

$$\max_{\alpha_i}\left\{\sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i(y_i y_j \mathbf{x}_i\mathbf{x}_j)\alpha_j\right\} \tag{7}$$

subject to the constraints

$$\alpha_i \geq 0, \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \tag{8}$$

The cost function to be maximized in (7) is convex and quadratic with respect to the unknown parameters α_i, and in practice, it is solved numerically through quadratic programming.

Note that only a few parameters α_i will have values satisfying the constraints in (8), i.e., will be nonzero. The corresponding training vector $\mathbf{x}_i$ is called a *support vector*. Vector $\mathbf{w}$ is computed from (6) while scalar b is computed from $y_i(\mathbf{w}\mathbf{x}_i + b) = 1$ for any support vector. The classification of a vector $\mathbf{x}$ outside the training set is performed by

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{l}(\alpha_i y_i \mathbf{x}\mathbf{x}_i + b)\right). \tag{9}$$

B. Kernel-based SVM Classifiers

For many datasets, it is unlikely that a hyperplane will yield a good classifier. Instead, we want a decision boundary with more complex geometry. One way to achieve this is to map the attribute vector into some new space of higher dimensionality and look for a hyperplane in that new space, leading to kernel-based SVMs [38], [39]. The interesting point about kernel functions is that although classification is accomplished in a space of higher dimension, any dot product between vectors involved in the optimization process can be implicitly computed in the low-dimensional space [15].

Let $\Phi(\cdot)$ be a nonlinear operator mapping the input vector $\mathbf{x}$ to a higher dimensional space. The optimization problem for the

new points $\Phi(\mathbf{x})$ becomes

$$\min \ J(\mathbf{w},\xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \tag{10}$$

subject to the constraints

$$y_i(\mathbf{w}^T\Phi(\mathbf{x}_i)+b) \geq 1-\xi_i, \quad \xi_i \geq 0, \quad i=1,\ldots,l. \tag{11}$$

Following the same principles as in the linear case, we note that the only form in which the mapping appears is in terms of $K(\mathbf{x}_i,\mathbf{x}_j,) = \Phi^T(\mathbf{x}_i)\Phi(\mathbf{x}_j)$. That is, the mapping appears only implicitly through the kernel function $K(\cdot,\cdot)$. There are a variety of possible kernels. However, when choosing a kernel, it is necessary to check that it is associated with the inner product of some nonlinear mapping [37]. Some typical choices for kernels are polynomials and radial basis functions.

Finally, the dual problem to be solved is

$$\max_{\alpha_i}\left\{\sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i(y_iy_jK(\mathbf{x}_i,\mathbf{x}_j))\alpha_j\right\} \tag{12}$$

subject to the constraints $\alpha_i \geq 0, \quad \sum_{i=1}^{l}\alpha_i y_i = 0$, and the classifier becomes

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{l}(\alpha_i y_i K(\mathbf{x},\mathbf{x}_i)+b)\right) \tag{13}$$

III. Heart Rate Variability Features

At first, we describe how the RR features are obtained by a QRS detection algorithm [40]. Then, the most commonly used HRV analysis methods were selected as features for the Gaussian kernel-based SVM classifier. We employed the results of the widely used HRV analysis methods [1], [2].

A. RR Detection Algorithm

In the first step, the algorithm [40] passes the signal through a low-pass and a high-pass filter in order to reduce the influence of the muscle noise, the power line interference, the baseline wander, and the T-wave interference. After filtering, the signal is differentiated to provide the QRS slope information and it is squared making all data points positive and emphasizing the higher frequencies.

After squaring, the algorithm performs sliding window integration in order to obtain the waveform feature. A temporal location of the QRS is marked from the rising edge of the integrated waveform. In the last step, two thresholds are adjusted. The higher of the two thresholds identifies peaks of the signal. The lower threshold is used when no peak has been detected by the higher threshold in a certain time interval. In this case, the algorithm has to search back in time for a lost peak. When a new peak is identified (as a local maximum–change of direction within a predefined time interval), then this peak is classified as a signal peak if it exceeds the high threshold (or the low threshold if we search back in time for a lost peak) or as a noise peak otherwise. In order to detect a QRS complex, the integration waveform and the filtered signals are investigated and different values for the aforesaid thresholds are used. To be identified as a QRS complex, a peak must be recognized as a QRS in both integration and filtered waveform.

B. Statistical HRV Features

Let x_i, $i = 1, 2, \ldots, N$ be the series of the RR intervals (the time interval between consecutive R points of the QRS) of a heartbeat signal. Let, also, $\bar{x}$ be the RR signal mean value. The statistical methods considered in this study are:

1) the standard deviation

$$\text{sdnn} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

2) the standard deviation of mean values of intervals (sdann) is defined by the standard deviation of mean values of successive equal-sized window intervals. A typical value of the window size is 5 min of recording;
3) the root-mean-square of successive differences

$$\text{rmssd} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(x_{i+1}-x_i)^2}$$

4) the mean standard deviation of intervals (sdnni) is described by the mean standard deviation of successive equal-sized window intervals, in a way similar to sdann. Again, we define a window size, but this time we first calculate the standard deviation for every successive window and then compute the mean value of the standard deviations;
5) the percentage of differences greater than x ($pNNx$) calculates how much percent of the differences between successive samples are greater than a given value x;
6) the standard deviation of differences

$$\text{sdsd} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(dx_i-\bar{dx})^2}$$

where $dx_i = x_{i+1} - x_i$, x_i is a sample point, $\bar{dx}$ the mean value of all dx_i and N the total number dx_i intervals;
7) the autocorrelation

$$\text{corr}(\tau) = \frac{\sum_{i=1}^{N-\tau}(x_i-\bar{x})(x_{i+\tau}-\bar{x})}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

where τ is a time lag;
8) the Shannon entropy

$$\text{entr} = -\sum_{i=1}^{B} f_i \log f_i$$

where f_i is the relative frequency of the ith bin of the RR intervals histogram. The RR intervals are quantized into B bins that span the range $[0, \max\{x_i\}]$.

C. Prediction-Based HRV Feature

1) The local linear prediction (llp) [17], [41] is a simple autoregressive prediction method in which future samples of a time series $x_1, x_2, \ldots, x_i, \ldots, x_N$ are predicted by using a linear combination of previous k samples:

$$\hat{x_i} = \frac{1}{k} \sum_{j=i-k}^{i-1} x_j.$$

The index is calculated by the mean values of absolute differences between predicted and actual values

$$\text{llp} = \frac{1}{N-k} \sum_{i=k}^{N} |\hat{x_i} - x_i|.$$

D. Wavelet HRV Features

1) The signal is decomposed by the discrete wavelet transform using the Haar wavelet whose mother function is simply a step function. The standard deviation of the detail coefficients, representing the high-frequency content of the signal, of each scale of analysis is computed. A detailed description of the method is described in [42].

IV. Materials and Methods

We applied the Gaussian kernel-based SVM classification to two different datasets. The first dataset consists of long-term ECG recordings, in which twenty young (21–34 years old) and twenty elderly (68–85 years old) rigorously screened healthy subjects underwent 120 min of continuous supine resting while continuous electrocardiographic signals were collected. Each subgroup of subjects includes equal numbers of men and women. All subjects remained in a resting state in sinus rhythm while watching the movie Fantasia (Disney, 1940) to help maintain wakefulness. The continuous ECG signals were digitized at 250 Hz. Each heartbeat was annotated using an automated arrhythmia detection algorithm, and each beat annotation was verified by visual inspection. The data are available and further described in [16] and [43]. We refer to these recordings as *Data Set I*.

The second data set consists of long-recording ECGs (approximately 2 h long) of six normal subjects and six subjects suffering from coronary artery disease [17]. The normal subjects were young males aged 25–29 years, with unremarkable medical histories and normal physical examinations. All of the subjects were nonsmokers, received no drugs, and abstained from caffeine for 24 h prior to acquisition. The recordings were performed in a controlled environment that is similar for all patients and a cardiologist is always present to ensure that preparation and procedure details during cardiogram acquisition are followed properly. The patient subjects were hospitalized, had one- or two-vessel coronary disease, which was angiographically confirmed, and normal left ventricular function (defined as an ejection fraction greater than or equal to 50%). Subjects with a history of myocardial infarction, coronary angioplasty or bypass grafting, cardiac rhythm disturbances, left ventricular dysfunction (defined as an ejection fraction less than 50%), severe

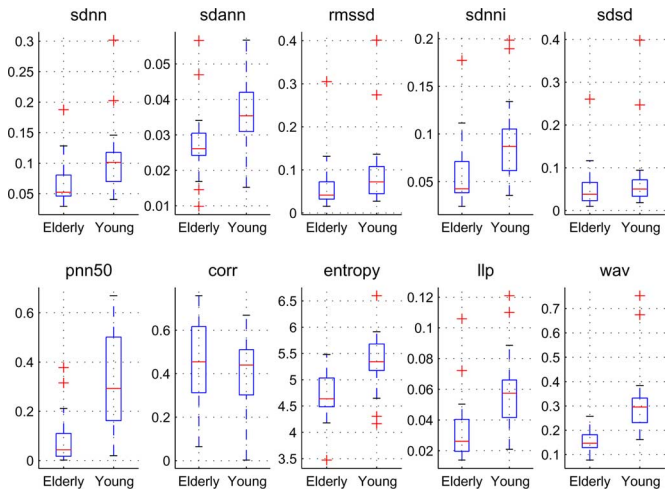

Fig. 1. Features computed for the heartbeat signals of *Data Set I*.

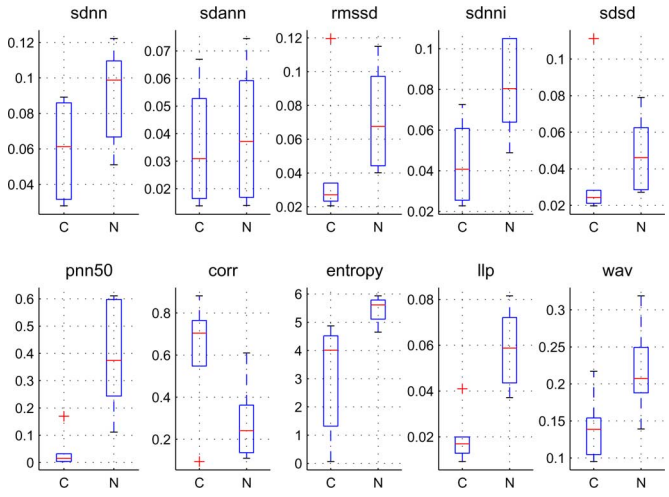

Fig. 2. Features computed for the heartbeat signals of *Data Set II*. *N* indicates normal subjects and C indicates subjects suffering from coronary artery disease.

arterial hypertension, and medical conditions affecting heart rate variability (e.g., diabetes mellitus, hormonal disturbances, treatment with psychotropic drugs, and respiration diseases) were excluded. All patient subjects were undergoing treatment with nitrates, angiotensin converting enzyme inhibitors, salicylics, and calcium antagonists. No one had a history of stroke, peripheral vascular disease, or clinically significant valvular abnormalities. Their ages were 42–52 years old. To guarantee that valid and precise data were acquired, a cardiologist was present to ensure that all preparation and procedure details during electrocardiogram acquisition were followed properly. All recordings were performed in a quiet room, between the hours of 15.00 and 17.00, in the supine position under continuous monitoring by the cardiologist who confirmed the absence of any cardiac rhythm disturbances throughout the recording. Subjects were told to breathe normally and an attempt was made to maintain the respiratory rate at around 12/min. The continuous ECG signals were digitized at 300 Hz. We refer to these recordings as *Data Set II*.

In HRV analysis methods, a set of parameters should be specified for the feature extraction processes. This is nontrivial and usually a data-dependent task. We have selected these parameters after a trial and error procedure. In the performed experiments, the window size for both sdann and sdnni is set to 5 min of recording, as suggested in [1]. Llp has been applied with window size equal to 8, as suggested in [17]. The τ parameter of the autocorrelation function has been set to 5 beats. The Haar wavelet was applied to extract multiresolution features. Since the seventh scale of analysis produced the clearest classification compared to all other scales of analysis, we selected this scale as the output of the method.

V. Numerical Experiments

To motivate the use of vector-valued features and nonlinear classification using SVM, we present in Figs. 1 and 2 the features computed for the signals of the two datasets. Also, in Table I, the *p*-values of the features of the datasets used in the experiments are presented. As can be easily observed, although the features are statistically significant (small *p*-values), the signals may not be classified correctly by simple thresholding.

In our evaluation framework, we also compared the SVM approach with the learning vector quantization (LVQ) neural network [44] and a backpropagation neural network using the Levenberg–Marquardt (LM) minimization algorithm [45].

LVQ is an autoassociative nearest-neighbor classifier that classifies arbitrarily patterns into classes using an error correction encoding procedure related to competitive learning. The main idea is to cover the input space of samples with codebook vectors, each representing a region labeled with a class. A codebook vector can be seen as a prototype of a class member, localized in the center of a class or decision region (Voronoi cell) in the input space. A class can be represented by an arbitrary number of codebook vectors, but one codebook vector represents one class only.

Neural network minimization problems are often very ill-conditioned due to the Hessian matrix. For such problems, the LM algorithm is often a good choice. It approximates the Hessian by a product of the Jacobian, which is a matrix that contains the first derivatives of errors with respect to the weights. The Jacobian is much less complex to be computed and the approximation speeds up the minimization algorithm.

Let us also notice that SVM has the computational cost of a quadratic programming optimization algorithm that is not prohibitive if the vector dimensions are not very large, as is the case in our experiments.

A first experiment consists in applying the SVM classifier to feature vectors generated by the signals of *Data Set I* and *Data Set II*. In that framework, leave-one-out cross validation was implemented for all of the compared methods (SVM, LVQ, and LM). The classifier was trained with all but one signal and the remaining signal was used for testing the classifier. The procedure is repeated in a cyclic way until all of the signals have been used for testing. The percentage of correctly classified signals for the SVM, the LVQ, and the LM neural networks are summarized in Table II. As can be seen, SVM achieves to correctly differentiate young and elderly subjects at 100% for *Data Set I*. The same stands for *Data Set II*, where healthy and pathological cases are correctly classified with the SVM while the other methods present weaker performances. This is an important conclusion that reveals the efficiency of the SVM methodology for heart rate variability signal classification.

A second experiment consists in investigating the robustness to noise of the SVM classifier. The whole set of the original signals (the original heartbeat time series) was corrupted by zero mean white Gaussian noise. The standard deviation of the noise was selected appropriately in order to obtain the SNRs between 5 and 0 dB. Attribute vectors were created from the degraded signals and the classifiers were trained by a total number of 500 signals for each SNR level. Two hundred new test signals were then created by the same procedure for each SNR level and the classifiers were evaluated on them. Table III summarizes the performances of the compared algorithms where the SVM classifier overwhelms the other methods. Let us notice that when a significant amount of noise is added to the signal, the statistical attributes totally missclassify the signals by simple thresholding. Table III summarizes the results.

We have also investigated the relation between the number of features and the classification performance which is a first step toward feature selection. To this end, we have randomly selected 20 signals of *Data Set I* (10 young and 10 elderly) and used them as a training set for the SVM classifier. The remaining signals (10 young and 10 elderly) were used as a test set. At first, we removed one feature from the feature vectors and performed both training and test phases with the remaining 9 features. This was done for all possible combinations of 9 features chosen out of 10 features (totally 10 configurations). Then, the same experiment was repeated by omitting 2 features out of 10 resulting in 45 configurations. This procedure was repeated until keeping 4 features out of 10 (210 configurations). In every configuration, the values of *Recall* and *Precision* were computed. Let us remind that *Recall* and *Precision* are defined as

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}, \quad \text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{14}$$

where tp, fp, and fn are the true positive, false positive, and false negative classification results for the examined signals. More specifically, *Recall* shows the percentage of the ground truth that was retrieved and *Precision* represents the percentage of the retrieved signals that were relevant (i.e., correctly classified).

A statistical representation of the results is presented in Fig. 3. As can be observed, the fact of leaving out one feature at a time gradually deteriorates the *Recall* measure. However, the *Precision* is less affected. The same experiment was repeated for the case of a noisy signal with SNR of 2 dB and the respective results are presented in Fig. 4. In that case, the *Precision* measure provides less coherent classification rates as is intuitively expected.

In order to study the relationship between the number of training signals and the classification performance, one has to perform a huge amount of experiments. In our case, we have performed leave-one-out cross validation and a second experiment,

TABLE I
p-VALUES OF THE FEATURES OF THE DATASETS USED IN THE EXPERIMENTS

Feature	$sdnn$	$sdann$	$rmssd$	$sdnni$	$pNN50$	$sdsd$	$corr$	$entr$	llp	wav
Data Set I	0.010	0.014	0.219	0.010	0.000	0.432	0.268	0.000	0.002	0.000
Data Set II	0.070	0.753	0.175	0.017	0.003	0.573	0.003	0.037	0.018	0.031

TABLE II
PERCENTAGE OF CORRECTLY CLASSIFIED SIGNALS USING LVQ, LM, AND SVM CLASSIFIERS WITH LEAVE-ONE-OUT CROSS VALIDATION. FOR EACH SIGNAL, A FEATURE VECTOR WAS CREATED WITH COMPONENTS OF THE HRV CHARACTERISTICS, DESCRIBED IN SECTION III

Algorithm	LVQ	LM	SVM
Data Set I	80%	85%	100%
Data Set II	92%	92 %	100%

TABLE III
PERCENTAGE OF CORRECTLY CLASSIFIED SIGNALS USING LVQ, LM, AND SVM CLASSIFIERS TO DATA CORRUPTED BY ZERO-MEAN WHITE GAUSSIAN NOISE. IN EACH CASE, 500 NEW DEGRADED SIGNALS WERE USED FOR TRAINING AND 200 SIGNALS FOR TESTING THE ALGORITHMS. FOR EACH CORRUPTED SIGNAL, A FEATURE VECTOR WAS CREATED WITH COMPONENTS OF THE HRV CHARACTERISTICS, DESCRIBED IN SECTION III

	Data Set I			*Data Set II*		
SNR	LVQ	LM	SVM	LVQ	LM	SVM
5 dB	67%	81%	100%	74%	78%	100%
4 dB	66%	80%	98%	72%	77%	100%
3 dB	65%	78%	97%	71%	76%	100%
2 dB	64%	76%	93%	69%	74%	97%
1 dB	62%	84%	87%	68%	71%	88%
0 dB	60%	68%	79%	66%	69%	78%

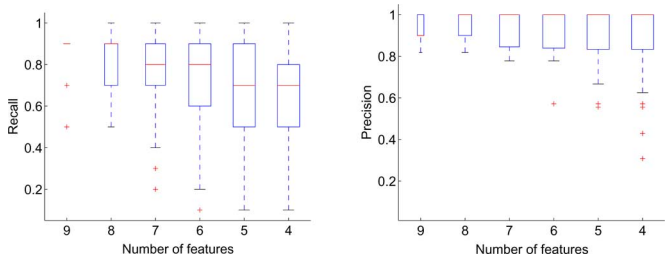

Fig. 3. *Recall* and *Precision* values as a function of the number of features. For every value of features in the ordinate, all the possible combinations of the initial ten features were examined. The boxes summarize the statistics of the respective experiments.

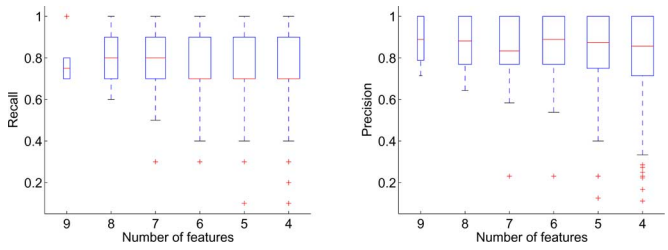

Fig. 4. *Recall* and *Precision* values as a function of the number of features for the signals degraded by white Gaussian noise at 2 dB. For every value of features in the ordinate, all the possible combinations of the initial ten features were examined. The boxes summarize the statistics of the respective experiments.

where we have randomly selected half of the signals as training signals. The obtained results showed that there is some difference in the classification rates though not significant. For instance, in Fig. 3, in the *Recall* measure, we can observe that leaving one feature out (i.e., with nine features), the median value is 0.9 and that there are two cases with lower rates. This is due to the selection of the training signals. On the other side, the majority of the values are around 0.9 and the *Precision* rates are much higher and not significantly affected.

VI. DISCUSSION AND CONCLUSION

The categorization of the ECGs into two dinstinct groups according to their heart rate variability can be very accurate using an SVM in cases where standard methods fail to present a satisfactory categorization. Experiments comparing SVM classification of heart rate signals with the classifications obtained by other nonlinear classifiers have also confirmed the effectiveness of the former methodology even in the presence of important amount of noise.

In the relevant literature, the features employed for HRV characterization are either applied as scalars [1], [2] or used in totally different contexts by being adapted to a specific application [4]–[10]. For instance, in [6], the authors present a study for the detection of regularities in the time series, while in [9], the authors detect abrupt changes in the signal. The main characteristic of the already proposed methods is that they focus on a specific feature and try to improve its performance. Generally, there is no effort for a combination of different features. However, we agree that the majority of the presented studies are application dependent. Therefore, the authors try to improve the performance of a specific feature. From this point of view, our approach is more general and flexible.

In our experiments, for the specific datasets examined, we achieved a very accurate classification of subjects, contrary to the most common HRV analysis methods that failed to categorize the same signals accurately. An important open issue and a perspective of this study is the problem of feature selection. We would preferably provide the classifier with a feature space that would make some signal characteristic obvious. Although a first approach was proposed in this study, these aspects of signal classification are still the subject of ongoing research.

REFERENCES

[1] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, pp. 354–381, 1996.

[2] M. Teich, S. Lowen, K. Vibe-Rheymer, and C. Heneghan, "Heart rate variability: measures and models," in *Nonlinear Biomedical Signal Processing, vol. II, Dynamic Analysis and Modelling*. New York: IEEE Press, 2001, pp. 159–213.

[3] G. Berntson, J. Bigger, D. Eckberg, P. Grossman, P. Kaufmann, M. Malik, H. Nagaraja, S. Porges, J. Saul, P. Stone, and M. van der Molen, "Heart rate variability: Origins, methods, and interpretive caveats," *Psychophysiology*, vol. 34, no. 6, pp. 623–648, Nov. 1997.

[4] M. Kamath and E. Fallen, "Power spectral analysis of HRV: A noninvasive signature of cardiac autonomic functions," *Crit. Rev. Biomed. Eng.*, vol. 21, pp. 245–311, 1993.

[5] R. Silipo, G. Deco, R. Vergassola, and C. Gremigni, "A characterization of HRV's nonlinear hidden dynamics by means of Markov models," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 8, pp. 978–986, Aug. 1999.

[6] M. Ferrario, M. Signorini, G. Magenes, and S. Cerutti, "Comparison of entropy-based regularity estimators: Application to the fetal heart rate signal for the identification of fetal distress," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 119–125, Jan. 2006.

[7] D. Hoyer, B. Pompe, K. Chon, H. Hardhalt, C. Wicher, and U. Zwiener, "Mutual information function assesses autonomic information flow of heart rate dynamics at different time scales," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 584–592, Apr. 2005.

[8] R. Barbieri and E. Brown, "Analysis of heartbeat dynamics by point process adaptive filtering," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 4–12, Jan. 2006.

[9] B. Aysin, L. Chaparro, I. Gravé, and V. Shusterman, "Orthonormal basis partitioning and time frequency representation of cardiac rhythm dynamics," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 878–889, May 2005.

[10] J. Mateo and P. Laguna, "Improved heart rate variability signal analysis from the beat occurence times according to the IPFM model," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 8, pp. 997–1009, Aug. 2000.

[11] A. Bezerianos, S. Papadimitriou, and D. Alexopoulos, "Radial basis function neural networks for the characterization of heart rate variability dynamics," *Artif. Intell. Med.*, vol. 15, no. 3, pp. 215–234, 1999.

[12] A. Alexandridi, C. D. Stylios, and G. Manis, "Neural networks and fuzzy logic approximation and prediction for HRV analysis," presented at the Eur. Symp. Intell. Technol., Hybrid Syst. Implement. Smart Adapt. Syst., Oulu, Finland, Jul. 2003.

[13] F. Azuaje, W. Dubitzky, P. Lopes, N. Black, K. Adamson, X. Wu, and J. A. White, "Predicting coronary disease risk based on short-term RR interval measurements: A neural network approach," *Artif. Intell. Med.*, vol. 15, no. 3, pp. 275–297, 1999.

[14] C. Cortes and V. N. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, pp. 1–25, 1995.

[15] N. Christianini and J. Shawe-Taylor, *Support Vector Machines and Other Kernel-based Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[16] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000, circulation Electronic Pages [Online]. Available: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[17] G. Manis, S. Nikolopoulos, A. Alexandridi, and C. Davos, "Assessment of the classification capability of prediction and approximation methods for HRV analysis," *Comput. Biol. Med.*, vol. 37, no. 5, pp. 642–654, 2007.

[18] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[19] H. Byun and S. W. Lee, "A survey of pattern recognition applications of support vector machines," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 17, no. 3, pp. 459–486, 2003.

[20] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, and R. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1552–1563, Dec. 2002.

[21] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1542–1550, Nov. 2002.

[22] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 637–646, Jun. 1998.

[23] S. Li, T. Y. Kwok, I. Wai-Hang, and Y. Wang, "Fusing images with different focuses using support vector machines," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1555–1561, Nov. 2004.

[24] G. Pajares and J. M. de la Cruz, "On combining support vector machines and simulated annealing in stereovision matching," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1646–1657, Aug. 2004.

[25] R. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 828–838, May 2005.

[26] I. Güler and E. D. Übeyli, "Multiclass support vector machines for EEG signals classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 117–126, Mar. 2007.

[27] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaunner, and B. Scholkopf, "Support vector channel selection for BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, Jun. 2004.

[28] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI competition 2003-Data set IIb: Support vector machines for the P300 speller paradigm," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1073–1076, Jun. 2004.

[29] L. Ramirez, N. Durdle, D. Hill, and J. Raso, "A support vector classifier approach to predicting the risk of progression of adolescent idiopathic scoliosis," *IEEE Trans. Inf. Technol. Biomed.*, vol. 9, no. 2, pp. 276–282, Jun. 2005.

[30] L. Ramirez, N. Durdle, J. Raso, and D. Hill, "A support vector machines classifier to assess the severity of idiopathic scoliosis from surface topography," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 84–91, Jan. 2006.

[31] H. Liang and Z. Lin, "Detection of delayed gastric emptying from electrogastrograms with support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 5, pp. 601–604, May 2001.

[32] J. L. Rojo-Alvarez, J. Bermejo, V. M. Juarez-Caballero, R. Yotti, C. Cortina, M. A. Garcia-Fernandez, and J. C. Antoranz, "Support vector analysis of color-doppler images: A new approach for estimating indices of left ventricular function," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 1037–1043, Aug. 2006.

[33] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 584–589, Apr. 2004.

[34] G. Georgoulas, C. Stylios, and P. Groumpos, "Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 5, pp. 875–884, May 2006.

[35] A. Kampouraki, C. Nikou, and G. Manis, "Classification of heart rate signals using support vector machines," in *Proc. BioSignal Conf.*, Brno, Czech Republic, Jun. 2006, pp. 9–11.

[36] A. Kampouraki, C. Nikou, and G. Manis, "Robustness of support vector machine-based classification of heart rate signals," in *Proc. IEEE Conf. Eng. Med. Biol. Soc. (EMBS 2006)*, New York, Aug.–Sep., pp. 2159–2162.

[37] B. Schölkopf, C. Burges, and A. Smola, *Advances in Kernel Methods: Support Vector Learning*. New York: MIT Press, 1999.

[38] R. Kondor and T. Jebara, "A kernel between sets of vectors," in *Proc. 20th Int. Conf. Mach. Learn. (ICML)*, Washington, DC, 2003, pp. 361–368.

[39] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–202, Mar. 2001.

[40] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, Mar. 1985.

[41] G. Manis, S. Nikolopoulos, and A. Alexandridi, "Prediction techniques and HRV analysis," presented at the Medicon Health Telematics, Naples, Italy, 2004.

[42] M. C. Teich, "Multiresolution wavelet analysis of heart-rate variability for heart-failure and heart-transplant patients," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, vol. 3, Hong Kong, 1998, pp. 1136–1141.

[43] N. Iyengar, C.-K. Peng, R. Morin, A. Goldberger, and L. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *Amer. J. Physiol.*, vol. 271, pp. 1078–1084, 1996.

[44] T. K. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.

[45] M. T. Hagan and M. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.

Authors' photographs and biographies not available at the time of publication.