CHOOSING REAL-TIME PREDICTORS FOR VENTRICULAR ARRHYTHMIA DETECTION

BERNARDETE RIBEIRO†, AMÂNDIO MARQUES and JORGE HENRIQUES
Department of Computer Engineering
CISUC — Center for Informatics and Systems, University of Coimbra
Polo II, P-3030-290 Coimbra, Portugal
bribeiro@dei.uc.pt

MANUEL ANTUNES
Department of Cardiothoracic Surgery
University Hospital, Coimbra, Portugal
antunes.cct.huc@sapo.pt

The risk of developing life-threatening ventricular arrhythmias in patients with structural heart disease is higher with increased occurrence of premature ventricular complex (PVC). Therefore, reliable detection of these arrhythmias is a challenge for a cardiovascular diagnosis system. While early diagnosis is critical, the task of its automatic detection and classification becomes crucial. Therefore, the underlying models should be efficient, albeit ensuring robustness. Although neural networks (NN) have proven successful in this setting, we show that kernel-based learning algorithms achieve superior performance. In particular, recently developed sparse Bayesian methods, such as, Relevance Vector Machines (RVM), present a parsimonious solution when compared with Support Vector Machines (SVM), yet revealing competitive accuracy. This can lead to significant reduction in the computational complexity of the decision function, thereby making RVM more suitable for real-time applications.

Keywords: Pattern recognition; Relevance Vector Machines (RVM); Support Vector Machines (SVM); ventricular arrhythmias.

1. Introduction

Arrhythmias represent a serious threat, especially ventricular arrhythmias like ventricular tachycardia (VT) and ventricular fibrillation (VF). Other arrhythmias, including premature ventricular contraction (PVC), are not as lethal as VF, but are important for diagnosing heart disorders. PVC is due to an ectopic cardiac pacemaker located in the ventricle. Detecting ectopic beats is of particular interest.

†Author for correspondence
to support the detection of ventricular tachycardia and to evaluate the regularity of ventricles depolarization. For example, the risk of sudden death in patients with structural heart disease is higher with increased occurrence of premature ventricular complex. Therefore, understanding this disorder would be fundamental for predicting the development of ventricular arrhythmias. On the other hand, the reliable detection of these arrhythmias poses a challenge for a cardiovascular diagnostic system. While early diagnosis of ventricular arrhythmias trends is critical, automatic detection and classification becomes crucial. Thus it is very important that the algorithms perform the following functions: (i) Quality signal measurement, i.e. to distinguish between ECG signals and noise signals; (ii) QRS complex detection; (iii) PVC detection. Consequently, significant amount of research has focused on the development of algorithms for accurate ventricular arrhythmias diagnosis. Given the complexity constraints, i.e. many and large-sized samples are required, it is not possible to use the ECG signal directly. Thus, average wave amplitude, duration and area measurements, are usually adopted to perform a quantitative description of the signal and is used for parameter extraction. Once this set of extracted ECG parameters has been evaluated, several techniques for PVC detection\cite{2,4} can then be applied, ranging from regression models\cite{3} and fuzzy systems\cite{9} to neural networks (NN).\cite{1,11} Probabilistic approaches, such as Relevance Vector Machines (RVM)\cite{12} can infer a class prediction rather than a class membership whilst maintaining the accuracy property of state-of-art Support Vector Machines (SVM).\cite{14} Both lead to a sparse model, which is a linear combination of a certain set of basis functions defined by a kernel.

In the proposed approach (shown in Fig. 1), prior to PVC classification, an algorithm for on-line QRS detection inspired by Pan–Thompkins\cite{8} was implemented. Our improved version includes both threshold adjusting and a set of adaptive heuristics defined to allow accurate localization of QRS complexes. Feature extraction is

![Fig. 1. PVC classification system.](image-url)
then accomplished in order to set up the input vector with the most relevant fea-
tures for PVC detection. We then extend our previous results with NN to the use of
kernel-based learning machines SVM and RVM for ventricular arrhythmia predic-
tion. The results attest so far the efficiency and accuracy of the proposed system.

The present paper is organized as follows. Section 2 provides a brief introduction
to Support Vector Machines. In Sec. 3, we will introduce Relevance Vector Machines
reviewing sparse Bayesian learning, prior model specification and relevance vec-
tor classification. Section 4 illustrates the proposed algorithm for QRS detection,
as well as the threshold setups, adaptive heuristics defined and feature extrac-
tion procedure. Section 5 describes the experimental setup. Section 6 details the
results. Section 7 draws the conclusions and provides possible guidelines for future
research.

2. Support Vector Machines

Support Vector Machines (SVM) combine essentially two strong concepts: max-
imum margin classifiers with low capacity and implicit features spaces defined
by kernel functions. In other words, they conjoin the following properties: low
Vapnik–Chervonenkis (VC) dimension solutions through maximization of the mar-
gin and kernel nonlinearity. These properties being lead to better generalization
of the Vapnik learning machine. Given a training data set consisting of input–output
pairs \( \{x_n, t_n\}_{n=1}^N \), SVM use the convolution of the scalar products to build, in input
space, the nonlinear decision functions of the form:

\[
f(x) = \sum_{n=1}^{N} w_n K(x, x_n) + w_0 \tag{1}
\]

\( K \) stands for the kernel function, which must be a positive semi-definite matrix:

\[
K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle = \exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)
\]

and where \( \phi \) is the nonlinear mapping from input space to feature space. In the
expression above, the model weights \( \{w_n\} \) are given by the nonzero Lagrange mul-
tipliers which are called Support Vectors (SV). The convex optimization prob-
lem is solved by a quadratic programming procedure being the cost of complexity
\( \mathcal{O}(N^2) \).

3. Relevance Vector Machines

The RVM was proposed by Tipping as a Bayesian treatment of the sparse learning
problem. The RVM preserves the generalization and sparsity of the SVM, yet it also
yields a probabilistic output, as well as circumvents other limitations of SVM, such
as the need for Mercer kernels and the definition of the error/margin trade-off
parameter \( C \). The output of an RVM model is very similar to the Vapnik proposed
SVM model, and can be represented as:

\[ y(x; w) = \sum_{n=1}^{N} w_n K(x, x_n) + w_0 \]  

such that \( K(x, x_n) \) is a kernel function which defines a basis function \( \phi_n(x) \), for each example in the training data set, and \( w \) is the weight parameter vector. A discriminative approach is thus assumed where the focus is on learning the function \( 2 \) directly from the training data set. Unlike the support vector classifiers, the nonzero weights of RVM are not associated with examples close to the decision boundary, but rather appear to represent prototypical examples of classes. These examples are called relevance vectors and, in our case, they can be perceived as representations of PVCs occurring in the signals of a cardiac patient.

### 3.1. Sparse Bayesian learning

In a Bayesian learning approach a prior probability distribution \( p(w|\alpha) \) over the feature weight parameters can be specified by assigning a zero-mean Gaussian probability distribution to the weights:

\[ p(w|\alpha) = \prod_{n=0}^{N} \mathcal{N}(w_n|0, \alpha_n^{-1}) \]  

where the notation \( \mathcal{N}(w_n|0, \alpha_n^{-1}) \) stands for a Gaussian density over \( w_n \) with zero mean and variance \( \alpha_n^{-1} \). The vector of hyperparameters \( \alpha \) of dimension \( N + 1 \) specifies how far from zero the weights are in Ref. 10. These explicit Gaussian priors are described in the ARD (Automatic Relevance Determination) model. Unlike SVM where overfitting is prevented by the margin term, this technique promotes simplicity and smoothness allowing tractability and computability. The hyperparameters are obtained from a training procedure by maximizing the evidence \( p(t|\alpha) \). The outcome of this optimization process is that many of the values of \( \alpha \) go to infinity, allowing only a few \( w_n \) of vector \( w \) to be nonzero. Therefore, in a way, the irrelevant data points are pruned out from the problem data, thus resulting in a parsimonious solution.

### 3.2. Relevance vector classification

For an input vector \( x_n \) sparse Bayesian classification allows to model the probability distribution of its class label \( t_n \in \{0, 1\} \) by applying the logistic sigmoid link function \( \sigma = 1/(1 + e^{-y}) \) to \( y(x) \).\(^{12}\) Adopting a Bernoulli distribution for \( P(t|w) \) the likelihood is

\[ p(t|w) = \prod_{n=1}^{N} \sigma(y(x_n; w))^{t_n}[1 - \sigma(y(x_n; w))]^{1-t_n}, \]  

\( \text{(4)} \)
such that $t = (t_1 \cdots t_N)^T$, $w = (w_0 \cdots w_N)^T$. It is not possible to find the weight vector $w$ such that $p(t|w)$ is maximized due to the likelihood discontinuity. Therefore, closed forms for the posterior $p(w|t, \alpha)$ and the marginal likelihood $p(t|\alpha)$ cannot be obtained. We must use Laplace approximation scheme developed in Ref. 6 for approximating the posterior. The method approximates the evidence by a Gaussian distribution around the maximum a posteriori value of $w$

$$p(w|t, \alpha) \approx N(w|w_{MP}, \Sigma).$$

In the above expression $\Sigma$ is the covariance matrix of the posterior probability over the weights at $w_{MP}$

$$\Sigma = (\Phi^T \Phi + A)^{-1}. \quad (5)$$

$$w_{MP} = \Sigma \Phi^T B t. \quad (6)$$

Here, $\Phi$ is the $N \times (N + 1)$ design matrix with $\Phi = [\phi(x_1); \phi(x_2); \cdots; \phi(x_N)]^T$, where $\phi(x_n) = [1; K(x_n; x_1); K(x_n; x_2); \cdots; K(x_n; x_N)]^T$, $A = \text{diag}(\alpha)$ is the hyperparameter diagonal matrix, and $B = \text{diag}\{\beta_1, \ldots, \beta_n\}$ with $\beta_n = \sigma(y(x_n; w))(1 - \sigma(y(x_n; w)))$. The iterative approximation finds, for a fixed $\alpha$, the locally most probable weights $w_{MP}$. This process typically involves a gradient descent optimization over the parameters. Using the new $w_{MP}$ the new target $\hat{t}$ is then obtained through\textsuperscript{13}:

$$\hat{t} = \Phi w_{MP} + B^{-1}(t - \sigma(y(x; w))). \quad (7)$$

Using $\Sigma$ and $w_{MP}$, $\alpha_i$ parameters can be updated by:

$$\alpha_i = \frac{\gamma_i}{\Sigma_{ii}}$$

$$\gamma_i = 1 - \alpha_i \Sigma_{ii}. \quad (8)$$

The RVM training algorithm involves a computationally costly technique with complexity $O(N^3)$ due to the matrix inversion, achieved robustly through Cholesky decomposition.

4. QRS Detection, Feature Selection and Feature Extraction

A large number of methods and algorithms have been developed for off-line and real-time QRS detection. For an excellent review, see the paper by Kohler.\textsuperscript{5} The algorithm for QRS detection proposed here is an improved version of the Pan–Tompkins algorithm.\textsuperscript{8} Basically, the algorithm consists of two main steps: in the first one, the energy of the ECG signal is computed through a preprocessing method (band-pass filter, differentiator, squaring operation and moving-window integration); in the second, the evaluated energy signal is compared to a predefined threshold to recognize the occurrences of QRS complexes. The preprocessing block diagram is shown in Fig. 2. Moreover, a proper adjustment is done of parameters and thresholds, through the implementation of adaptive heuristics. Details of the implementation can be seen elsewhere.\textsuperscript{7}
4.1. Band-pass filter

The band-pass filter aims at reducing the influence of muscle noise, baseline Wander, and T-wave interference. The band-pass filter is obtained by cascading a low-pass filter with a high-pass filter. Since the QRS complex contains the highest frequencies in a normal ECG signal, a [5–15]Hz band-pass filter was used. The implementation of the filter consists of a fast, real-time recursive filter design with integer coefficients.

4.1.1. Low-pass filter

The transfer function of the low-pass filter is given by:

\[ H(z) = \frac{(1 - z^{-6})^2}{(1 - z^{-1})^2}. \]  \hspace{1cm} (9)

If \( T \) is the sampling period and \( n \) is the sampling instant, the difference equation of the filter is:

\[ y(nT) = 2y(nT - T) - y(nT - 2T) + x(nT) - 2x(nT - 6T) + x(nT - 12T). \]  \hspace{1cm} (10)

4.1.2. High-pass filter

The transfer function of the high-pass filter is

\[ H(z) = \frac{-1/32 + z^{-16} - z^{-17} + z^{-32}/32}{1 - z^{-1}} \]  \hspace{1cm} (11)

and the difference equation of the filter is:

\[ y(nT) = y(nT - T) - x(nT)/32 + x(nT - 16T) - x(nT - 17T) \]
\[ + x(nT - 32T)/32. \]  \hspace{1cm} (12)

4.2. Differentiator

After filtering, the signal is differentiated to provide the QRS complex slope information. Using a five-point derivative, with the transfer function

\[ H(z) = 0.1(-2z^{-2} - z^{-1} + z^1 + 2z^2) \]  \hspace{1cm} (13)

the difference equation is given by

\[ y(nT) = (2x(nT) + x(nT - T) - x(nT - 3T) - 2x(nT - 4T))/8. \]  \hspace{1cm} (14)
4.3. **Squaring function**

After differentiation, the signal is squared point by point, in order to compute its energy:

\[ y(nT) = [x(nT)]^2. \]  \hspace{1cm} (15)

4.4. **Moving-window integration**

The purpose of the moving window integration is to obtain wave-form feature information in addition to the R wave slope. It is calculated as follows:

\[ y(nT) = \frac{1}{N} \sum_{i=0}^{N-1} x(nT - iT) \]  \hspace{1cm} (16)

where \( N \) is the number of samples within the width of the integration window.

4.5. **Peak detection**

In order to determine the R peaks, the signal provided by the preprocessing phase is considered an input to a moving window. Its duration is found in order to minimize the effect of noise in the threshold adjustment. A peak is considered when the amplitude of the signal at the center of the window is higher than its value in the window extremities. The detected peak is classified as signal if its amplitude does not surpass the \( \text{thr}_1 \) threshold and it will be considered as noise otherwise.

4.6. **Setting thresholds**

The thresholds are automatically adjusted to float over the noise. Low thresholds are justified due to the improvement of the signal-to-noise ratio by the band-pass filter. The highest of the two thresholds, \( \text{thr}_1 \), is used for the first analysis of the signal. The lowest threshold \( \text{thr}_2 \) is used if no QRS is detected in a certain time interval so that a search-back technique is necessary to search back in time for the QRS complex. The set of thresholds initially applied to the integration waveform is calculated based on the following:

\[ \text{spki} = 0.125 \times \text{peaki} + 0.875 \times \text{spki} \]  \hspace{1cm} (17)

\[ \text{npki} = 0.125 \times \text{peaki} + 0.875 \times \text{npki} \]  \hspace{1cm} (18)

\[ \text{thr}_1 = 2 \times \text{npki} + 0.125 \times (\text{spki} - \text{npki}) \]  \hspace{1cm} (19)

\[ \text{thr}_2 = 0.5 \times \text{thr}_1 \]  \hspace{1cm} (20)

where the variables \( \text{spki} \) and \( \text{npki} \) define the running estimate of signal and noise peaks respectively.

Unlike Pan–Tompkins Algorithm, which considers only peaks, when they are detected simultaneously in the integration and bandpass filtered waveform, our algorithm uses only one set of thresholds, more restricted, and applied only to the integration waveform. Thus, it reduces the computational effort and improves the classification results significantly.
4.7. Heart beat rate estimation

The estimated heart beat rate is calculated based on the average of the time intervals between peaks R ($RR_{av}$) in the last QRS complexes detected. In order to calculate this average value, the last eight RR’s higher than $92\% \, RR_{av}$ and lower than $116\% \, RR_{av}$ are considered. If it is not possible to find eight RR’ values in the last 12 RR values that satisfy the above conditions, an indiscriminate average of the last eight RR values is adopted. The moving average of the last eight values will allow a smooth adjustment in the estimation procedure. On the other hand, a search window of only 12 values endows the system with the capability to quickly adapt to abnormal variations.

4.8. Adaptive heuristics

4.8.1. Heuristic 1

After a QRS complex has been detected and within a period of 200ms the occurrence of new energy peaks are not considered to calculate new QRS complexes, since their occurrence is not morphologically feasible. However, if a peak of energy occurs with an energy 150% higher than the previous QRS complex found, and this value is lower than 150% of the thr$_1$ threshold [Eq. (19)], the last QRS complex will be discarded and the actual peak is taken as a candidate.

4.8.2. Heuristic 2

After a period of 200ms, up to 360ms, the occurrence of a peak of energy may correspond to a T-wave. It will be considered as a QRS complex if the amplitude is higher than half of the amplitude of the last QRS complex detected.

4.8.3. Heuristic 3

On the basis of the last estimated heart beat rate, if the algorithm does not find one QRS complex up to $166\% \, RR_{av}$, the second thr$_2$ threshold, Eq. (20) will be used in the attempt to find the last heart beat. The highest energy peak will be considered a QRS complex if it surpasses the second defined threshold.

4.9. Feature selection and extraction

We take into account the morphology of PVCs to select the relevant features for classification. Morphologically, PVCs are characterized by the premature occurrence of strangely shaped QRS complexes (see Fig. 3), usually with a duration greater than 120ms. These complexes are not preceded by a P wave, and the T wave is usually large and opposite in direction to the major QRS deflection.

It is not possible to apply a classification method directly to the ECG samples due to complexity constraints, i.e. given the large amount and the high dimension of required samples. Therefore, measurements of average wave amplitudes,
Choosing Real-Time Predictors for Ventricular Arrhythmia Detection

Fig. 3. Normal and PVC beat.

Table 1. Selected features from ECG signal.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRav</td>
<td>RR mean interval</td>
</tr>
<tr>
<td>RR0</td>
<td>Last RR interval</td>
</tr>
<tr>
<td>SN</td>
<td>Signal/Noise estimation</td>
</tr>
<tr>
<td>Ql</td>
<td>Q-wave length</td>
</tr>
<tr>
<td>(Qcx, Qcy)</td>
<td>Q-wave mass centre (x,y) coordinates</td>
</tr>
<tr>
<td>(Qpx, Qpy)</td>
<td>Q-wave peak (x,y) coordinates</td>
</tr>
<tr>
<td>Rl</td>
<td>R-wave length</td>
</tr>
<tr>
<td>(Rcx, Rcy)</td>
<td>R-wave mass centre (x,y) coordinates</td>
</tr>
<tr>
<td>(Rpx, Rpy)</td>
<td>R-wave peak (x,y) coordinates</td>
</tr>
<tr>
<td>Sl</td>
<td>S-wave length</td>
</tr>
<tr>
<td>(Scx, Scy)</td>
<td>S-wave mass centre (x,y) coordinates</td>
</tr>
<tr>
<td>(Spx, Spy)</td>
<td>S-wave peak (x,y) coordinates</td>
</tr>
</tbody>
</table>

durations and areas, are usually adopted to perform a quantitative description of the signal and a parameter extraction regarding PVC classification. Once this set of extracted ECG parameters is evaluated, several techniques for medical diagnostic classification are then applied, such as probabilistic approaches, heuristic models, knowledge-based systems as well as neural networks as previously referred. The features extracted and computed from the output of the QRS complex detector are listed in Table 1. Figure 4 depicts graphically the QRS complex with indicated features for clarification. These features are able to distinguish normal from abnormal PVC beats, as will be shown in Sec. 6. The advantage of the feature extraction technique so far is that we were able to reduce the size of the descriptive feature vector and, hence, the size of the input space.

5. Experimental Data Sets

5.1. MIT-BIH arrhythmia database

The database used consists of 48 annotated records. Each record is about 30 min long. Four records include paced beats. Several records were chosen specifically
because features of the rhythm, QRS Morphology or signal quality may be expected
to present significant difficulty to arrhythmia detectors. The availability of anno-
tated MIT-BIH\textsuperscript{a} database enables the evaluation of the classifier performance.

5.2. Performance metrics
To estimate the quality of the algorithm, sensitivity (SE) and specificity (SP) were
used. SE refers to the algorithm’s ability to detect QRS complex, while SP refers
to the algorithm’s ability to detect other points where there are no QRS. True
positive (TP) occurs when the algorithm finds a QRS complex which coincides
with the annotation in the database record (see Table 2). A true negative (TN)
is assumed in between two consecutive TP. A false positive (FP) occurs when the
algorithm detects a QRS-like artifact. A false negative (FN) is an annotated QRS
complex that has not been detected by the algorithm.

Also error types I and II are rather important to analyze detection ability of
classifiers. Further, we have extended the use of these measures to PVC detection

\begin{table}[h]
\centering
\caption{Contingency table for binary classification.}
\begin{tabular}{lll}
\hline
 & Class Positive & Class Negative \\
\hline
Assigned Positive & TP & FP \\
Assigned Negative & FN & TN \\
\hline
\end{tabular}
\end{table}

\textsuperscript{a}http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm
Choosing Real-Time Predictors for Ventricular Arrhythmia Detection

Based on a similar interpretation as above. Next section will discuss the results obtained based on these performance metrics.

\[
SE = \frac{TP}{TP + FN} \quad E_{\text{Type I}} = \frac{FN}{FN + TP}
\]

\[
SP = \frac{TN}{TN + FP} \quad E_{\text{Type II}} = \frac{FP}{FP + TN}
\]

The correct classification (CC) or accuracy has been calculated and is defined as

\[
CC = \frac{TP + TN}{TP + FN + FP + FN}
\]

6. Results and Discussion

Results with the adaptive online algorithm in MIT-BIH database are shown in Table 3. As observed, the algorithm performs excellently, thus improving the results published in the same databases in Ref. 8. Possibly this might be due to the strength of heuristics that enables robustness. This is essential for a correct evaluation of patient clinical diagnosis. The algorithm is 100% correct in records (“100”, “102”, “118”, “201”, “202”, “209”, “212”, “213” and “219”). In the majority of databases

<table>
<thead>
<tr>
<th>Record</th>
<th>Beats</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>SE</th>
<th>SP</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>“100”</td>
<td>2273</td>
<td>2269</td>
<td>0</td>
<td>0</td>
<td>2268</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>“101”</td>
<td>1865</td>
<td>1858</td>
<td>7</td>
<td>3</td>
<td>1852</td>
<td>99.84</td>
<td>99.62</td>
<td>99.46</td>
</tr>
<tr>
<td>“102”</td>
<td>2187</td>
<td>2183</td>
<td>0</td>
<td>0</td>
<td>2182</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>“103”</td>
<td>2084</td>
<td>2080</td>
<td>0</td>
<td>1</td>
<td>2078</td>
<td>99.95</td>
<td>100.0</td>
<td>99.95</td>
</tr>
<tr>
<td>“104”</td>
<td>2229</td>
<td>2215</td>
<td>23</td>
<td>10</td>
<td>2196</td>
<td>99.55</td>
<td>98.96</td>
<td>98.52</td>
</tr>
<tr>
<td>“105”</td>
<td>2572</td>
<td>2559</td>
<td>41</td>
<td>9</td>
<td>2521</td>
<td>99.65</td>
<td>98.40</td>
<td>98.05</td>
</tr>
<tr>
<td>“106”</td>
<td>2027</td>
<td>2025</td>
<td>1</td>
<td>0</td>
<td>2024</td>
<td>100.0</td>
<td>99.95</td>
<td>99.95</td>
</tr>
<tr>
<td>“107”</td>
<td>2137</td>
<td>2131</td>
<td>0</td>
<td>2</td>
<td>2128</td>
<td>99.91</td>
<td>100.0</td>
<td>99.91</td>
</tr>
<tr>
<td>“118”</td>
<td>2278</td>
<td>2274</td>
<td>0</td>
<td>0</td>
<td>2273</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>“119”</td>
<td>1987</td>
<td>1984</td>
<td>1</td>
<td>0</td>
<td>1983</td>
<td>99.95</td>
<td>99.95</td>
<td>99.95</td>
</tr>
<tr>
<td>“200”</td>
<td>2601</td>
<td>2593</td>
<td>5</td>
<td>4</td>
<td>2584</td>
<td>99.85</td>
<td>99.81</td>
<td>99.65</td>
</tr>
<tr>
<td>“201”</td>
<td>1963</td>
<td>1915</td>
<td>0</td>
<td>44</td>
<td>1874</td>
<td>97.75</td>
<td>100.0</td>
<td>97.75</td>
</tr>
<tr>
<td>“202”</td>
<td>2136</td>
<td>2129</td>
<td>0</td>
<td>5</td>
<td>2123</td>
<td>99.77</td>
<td>100.0</td>
<td>99.77</td>
</tr>
<tr>
<td>“203”</td>
<td>2980</td>
<td>2933</td>
<td>26</td>
<td>42</td>
<td>2877</td>
<td>98.59</td>
<td>99.10</td>
<td>97.71</td>
</tr>
<tr>
<td>“205”</td>
<td>2656</td>
<td>2651</td>
<td>0</td>
<td>1</td>
<td>2649</td>
<td>99.96</td>
<td>100.0</td>
<td>99.96</td>
</tr>
<tr>
<td>“206”</td>
<td>2955</td>
<td>2923</td>
<td>8</td>
<td>27</td>
<td>2901</td>
<td>99.08</td>
<td>99.72</td>
<td>98.81</td>
</tr>
<tr>
<td>“209”</td>
<td>3005</td>
<td>3001</td>
<td>0</td>
<td>0</td>
<td>3000</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>“210”</td>
<td>2650</td>
<td>2596</td>
<td>3</td>
<td>49</td>
<td>2543</td>
<td>98.15</td>
<td>98.88</td>
<td>98.03</td>
</tr>
<tr>
<td>“212”</td>
<td>2748</td>
<td>2744</td>
<td>0</td>
<td>0</td>
<td>2743</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>“213”</td>
<td>3251</td>
<td>3245</td>
<td>0</td>
<td>0</td>
<td>3244</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>“214”</td>
<td>2262</td>
<td>2255</td>
<td>0</td>
<td>3</td>
<td>2252</td>
<td>99.87</td>
<td>100.0</td>
<td>99.87</td>
</tr>
<tr>
<td>“215”</td>
<td>3363</td>
<td>3353</td>
<td>0</td>
<td>5</td>
<td>3347</td>
<td>99.85</td>
<td>100.0</td>
<td>99.85</td>
</tr>
<tr>
<td>“217”</td>
<td>2208</td>
<td>2202</td>
<td>2</td>
<td>3</td>
<td>2197</td>
<td>99.86</td>
<td>99.91</td>
<td>99.77</td>
</tr>
<tr>
<td>“219”</td>
<td>2154</td>
<td>2151</td>
<td>0</td>
<td>0</td>
<td>2150</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Global Result 99.65 99.80 99.46
tested improvement is above 99%. In the case of record “203” it presents the lowest performance value: 97.71%.

As for PVC detection, we extend our previous results with NNs to the use of statistical learning algorithms, such as SVM and RVM. The evaluation results regarding correct classification, SE and SP are summarized in Tables 4 and 5. Regarding the MLP approach, the best neural network was found with 12 neurons in the hidden layer. The study of its selection, based on the minimization of the mean square error of desired and network output, has been published elsewhere.

The training, test and validation sets consist each of 19,391 samples of 18 extracted features (see Table 1). The output of the classifier is a decision about which class the input data is drawn from, i.e. normal QRS complex or abnormal QRS complex (PVC). To estimate detection algorithm quality, sensitivity (SE) and specificity (SP) were used. As said above, SE is algorithm’s ability to detect a PVC successfully, whereas SP expresses its ability to detect successfully a normal QRS complex (or other annotated signal rather than a PVC). Also two types of errors were defined, which account for misclassification rate. A true positive (TP) occurs when the algorithm finds a PVC which coincides with its annotation in the database. A false positive (FP) occurs when the algorithm wrongly detects a PVC. A false negative (FN) is an annotated PVC which has not been detected by the algorithm.

ROC (Receiver Operating Characteristic) curves were used to compare the tested classifiers. The results shown in Fig. 5 for all tested classifiers are very good, since the Areas Under the Curve (AUC) are clearly over 99% while the diagonal represents a worthless classifier in which the AUC is 50%. While best results (see Tables 4 and 5) are obtained with SVM and also MLP, RVM yield higher sparsity in the final model solution (see the number of basis functions or expansion
vectors (EVs) used in Table 5), allowing them to be more suitable for real-time applications.

7. Conclusions

In this paper we proposed a system for PVC detection, a critical ventricular arrhythmia which may evolve into a serious life-threatening risk. First, an improved algorithm for on-line QRS detection was implemented, based on a set of adaptive heuristics and proper parameter adjustment. Second, extraction of critical features resulted in a set of relevant attributes for successful PVC detection. Finally, the classification results extend our previously used neural network approach to kernel-based methods, which are the state of art in machine learning. The effectiveness of the PVC classification system was validated on the annotated MIT-BIH database, where several occurrence of PVC episodes can be found. Classification results demonstrate that the proposed approach can be used to discriminate between different arrhythmias. Moreover, they are comparable to the recently published results on cardiac arrhythmias classification. We show that although good results were obtained with MLP for PVC detection, SVM are amongst the methods providing the best performance results. However, recently developed RVM provide a much sparser model, making it a choice predictor for real-time applications. Sparsity is interesting, both with respect to fast training and predictions, and ease of solution interpretation. Besides, SVM present the drawback of definitional constraints and absence of posterior class probability estimates, whereas RVM make a probabilistic prediction of a PVC without the need for regularization parameters. Probabilistic predictions are desirable because inference is most naturally formulated in terms of the probability theory, i.e. we can manipulate probabilities through Bayes’ theorem, reject uncertain predictions, etc.
Future work will focus on improving both sensitivity and specificity values, as well as on ensuring robustness of changes in measurement conditions (including noise). We will also investigate further ways of predicting critical ventricular arrhythmia trends. In particular, the validation phase will definitely be a challenge: there are no available databases that can be used to perform this kind of validation, thus specific data must to be collected.

Acknowledgments
We thank the anonymous reviewers for the useful comments which helped to improve the paper.

References
Choosing Real-Time Predictors for Ventricular Arrhythmia Detection

Bernardete Ribeiro received the degree in chemical engineering, the M.Sc. and Ph.D. degrees in informatics engineering from the University of Coimbra, Portugal. Since 2005, she is associate professor at the Department of Informatics of the University of Coimbra. She is a research member of CISUC-Center for Informatics and Systems at the University of Coimbra.

Her current research interests include machine learning, pattern recognition and signal processing techniques for clinical, industrial and information management applications.

Amândio Marques received the degree in systems and informatics engineering in 2001, from the Superior School of Technology of the Polytechnical Superior Institute of Viseu, Portugal. He is Professor of Secondary Education at Viriato Secondary School in Viseu. He is a research member of CISUC-Center for Informatics and Systems of University of Coimbra.

His current research interests include soft computing methodologies and pattern recognition techniques for industrial and clinical applications.

Jorge Henriques received the degree in informatics engineering, the M.Sc. and Ph.D. degrees in 1990, 1995 and 2001, all in informatics engineering, from the University of Coimbra, Portugal. Since 2001, he is assistant professor at the Department of Informatics of University of Coimbra, and a research member of CISUC-Center for Informatics and Systems of University of Coimbra.

His current research interests include soft computing methodologies, e-health solutions, and signal processing and pattern recognition techniques for industrial and clinical applications.

Manuel Antunes graduated in medicine and surgery in 1971 and received the degrees of Master of Medicine in 1988, Ph.D. in 1987 and D.Sc. in 2004 at the University of the Witwatersrand, Johannesburg, South Africa, where he was Professor and Head of the Department of Cardiothoracic Surgery until 1988. Since then, he occupies a similar position at the University of Coimbra, Portugal.

One of his main interests is surgery of valvular heart diseases, with special emphasis on mitral valve repair, about which he has written a book, several book chapters and more than 100 papers published in international peer-review journals.