Discriminating Between Supraventricular and Ventricular Tachycardias from EGM Onset Analysis

Support Vector Preprocessing and Incremental Learning Approaches for Increased Sensitivity and Specificity of ICDs

José L. Rojo-Álvarez[1], Ángel Arenal-Maíz[2], Antonio Artés-Rodríguez[1]
[1]Escuela Politécnica Superior, Universidad Carlos III, Madrid
[2]Department of Cardiology, Hospital GU Gregorio Marañón, Madrid

Automatic implantable cardioverter defibrillators (ICDs) have supposed a great advance in arrhythmia treatment in the last two decades [1]. The automatic tasks carried out by these devices include 1) the continuous monitoring of the heart rate; 2) the detection and the classification of the cardiac arrhythmia; and 3) the delivering of an appropriate, severity-increasing therapy (cardiac pacing, cardioversion, and defibrillation).

The usual cardiac activation pattern is known as sinus rhythm (SR), and cardiac arrhythmias stem from alterations of it. According to their anatomical origin, fast arrhythmias or tachyarrhythmias are classified into two groups: a) supraventricular tachycardias (SVTs), originated in the atria, and b) ventricular tachyarrhythmias, originated in the ventricles. The former can be either ventricular tachycardia (VT), whose electric pattern has a well-defined shape, or ventricular fibrillation (VF), which represents a more random and ineffective activation. The most dangerous tachyarrhythmias are VF and fast VT; in these cases there is no effective muscular pumping, thus causing sudden cardiac death, and immediate delivering of therapy is required. Nevertheless, SVTs rarely imply acute hemodynamic damage, so they do not require shock delivery. Apart from battery lifetime shortening and deterioration of the patient's quality of life, inappropriate therapy can even start a new VT or VF episode.

Due to the limited lifetime of the ICD batteries, the discrimination algorithms should demand quite low computational burden. The most commonly implemented algorithms are the *Heart Rate Criterion* [2], the *QRS Width Criterion* [3], and the *Correlation Waveform Analysis* [4]. While VF detection cycle rank is commonly accepted as appropriate, there is an overlapping between SVT and VT cycle ranks, and estimations of the number of inappropriate shocks (i.e., delivered to SVT) spread between 10 and 30% [5, 6].

In this article we hypothesize that the analysis of the ventricular electrogram onset (EGM onset) can discriminate between SVT and VT to obtain a simultaneous increase in sensitivity and specificity. We will discuss our analysis of EGMs obtained during SVT and VT together with their preceding SRs in 38 SVT and 68 VT far field records from 16 patients. The algorithmic implementation and the preprocessing tasks were performed through the support vector method (SVM), avoiding the overfitting by means of the statistical bootstrap resampling. To improve the safety for an individual patient, two new methods of incremental learning, based on the SVM, will be proposed and tested on an independent set of spontaneous arrhythmia episodes.

Analyzing Changes in Ventricular EGM Onset

The analysis of the initial changes in the ventricular EGMs has been recently proposed as an alternative arrhythmia discrimination criterion [7], given that it does not suffer from the drawbacks of the Heart

 0739-5175/02/$17.00©2002IEEE January/February 2002

Rate, the QRS Width, or the Correlation Waveform Analysis algorithms. The clinical hypothesis is as follows:

During any supraventricular originated rhythm both ventricles are depolarized through the conduction-specific His-Purkinje system, whose conduction speed is high (4 m/s); however, the electric impulse for a ventricular originated depolarization travels initially through the myocardial cells, whose conduction speed is slower (1 m/s). Then, we hypothesize that changes in the ventricular EGM onset can differentiate between SVT and VT.

Changes in the waveform are found by calculating the EGM first derivative. Figure 1 shows the anatomical elements involved in the hypothesis, and Fig. 2 depicts an SR, an SVT, and a VT example record, with the first-derivative superposed beats on the right, next to each corresponding record; the noisy activity preceding the beat onset has been previously removed. Note how EGM is a sudden activation in both SR and SVT beats but an initially less energetic activation in VT beats.

Once the criterion has been stated, two main questions arise:

- How can the criterion be implemented into an efficient algorithm?
- How must it be adjusted for a single patient?

To date, there is no statistical model for the cardiac impulse propagation that could be detailed enough to allow simulation research. Moreover, the assembling of ICD-stored EGMs is a difficult task because of the need for correct labeling, thus only small databases are available. A robust approach coping with both problems is the SVM, which is a sample-based learning procedure assuming no a priori statistical distribution from the data. Despite the stated robustness of the SVM when dealing with extremely small training sets, it is not yet an overfitting-free learning procedure, especially when additional free parameters must be previously determined in nonlinear machines. Under these circumstances, nonparametric bootstrap resampling will be proposed here to increase the generalization capabilities of the SVM classifiers. On the other hand, the adaptation of the learning scheme to a single patient is carried out through incremental learning (IL) schemes, which consist of finding a trade-off between the relevance of the population and the individual available episodes. The SVM and the bootstrap resampling allow the design of useful IL algorithms.

Statistical Learning Analysis

The Support Vector Method

The SVM was first proposed to obtain maximum margin separating hyperplanes in classification problems, but in a short time it has grown to a more general learning theory, and it has been applied to real data problems such as handwritten character identification or three-dimensional (3-D) object recognition [8, 9]. A comprehensive description of this method can be found in [10].

A set of observed and labeled data is

$$\mathbf{V} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_l, y_l)\} \quad (1)$$

where $\mathbf{x}_i \in \mathbf{R}^{\mathbf{n}}$ and $y_i \in \{+1, -1\}$. A nonlinear transformation $\phi(\mathbf{x}_i)$ to a generally unknown, higher-dimensional space $\mathbf{R}^{\mathbf{m}}$ where a separating hyperplane is given by:

$$(\phi(\mathbf{x}_i) \cdot \mathbf{w}) + b = 0. \quad (2)$$

We want to find the minimum of:

$$\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{l}\xi_i \quad (3)$$

with respect to $\mathbf{w}, b$, and ξ_i, and constrained to:

$$y_i\{(\phi(\mathbf{x}_i) \cdot \mathbf{w}) + b\} - 1 + \xi_i \geq 0, \; i = 1, \ldots, l \quad (4)$$

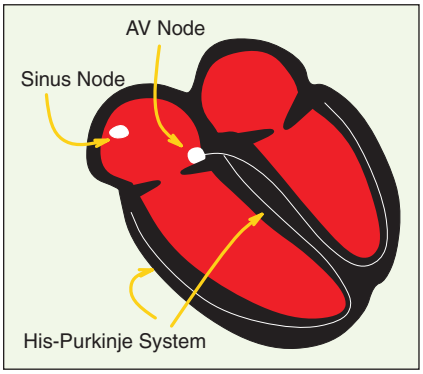
$$\xi_i \geq 0, \quad i = 1, \ldots, l \quad (5)$$

where ξ_i represent the losses; $||\mathbf{w}||^2$ can be shown to be the inverse of the class distance (and it will be called *margin*); C represents a trade-off between margin and losses; and $(\cdot)$ expresses the dot product in $\mathbf{R}^{\mathbf{m}}$. By the Lagrange Theorem, Eq. (3) can be rewritten into:

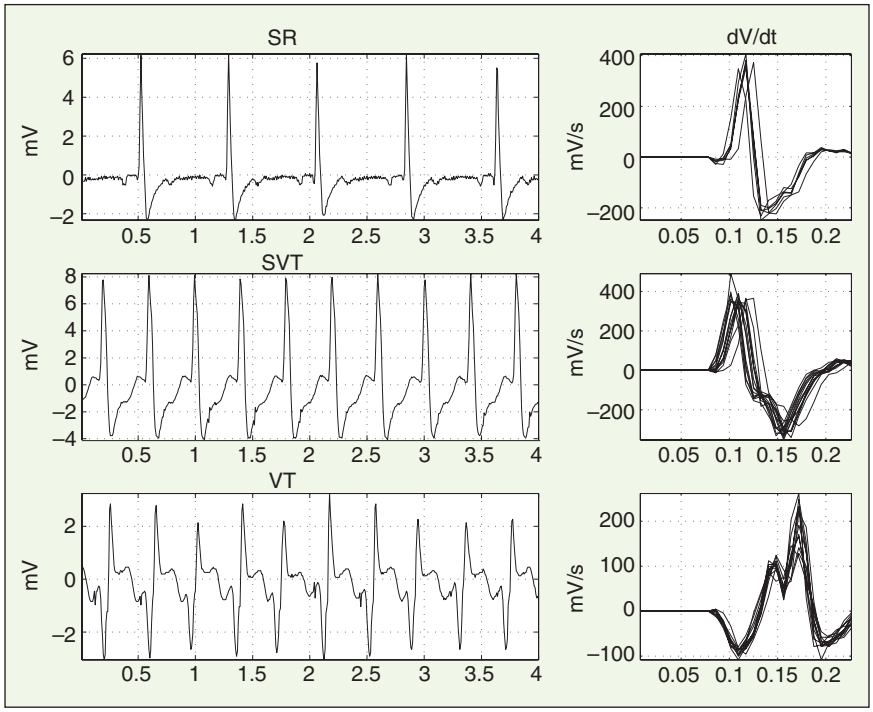
$$L_P = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{l}\mu_i\xi_i - \sum_{i=1}^{l}\alpha_i(y_i\{(\phi(\mathbf{x}_i) \cdot \mathbf{w}) + b\} - 1 + \xi_i) \quad (6)$$

which is minimized with respect to $\mathbf{w}, b, \xi_i$ and maximized with respect to α_i, μ_i, constrained to:

$$\alpha_i, \mu_i \geq 0, \quad i = 1, \ldots, l. \quad (7)$$



1. Anatomical elements involved in the ventricular depolarization.



2. Examples of SR, SVT, and VT, together with their superposed EGM first derivative.

The solution is a linear combination of the training data, where the samples with $\alpha_i \neq 0$ are called the *support vectors*. The classification function is built as a function of the support vectors, as depicted in Fig. 3. Nonlinear classifiers are obtained by taking the dot product in kernel-generated spaces. This product uses kernels satisfying the Mercer conditions; that is, semi-defined positive kernels [10]. The problem consists then in maximizing:

$$L_d = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (8)$$

constrained to:

$$C \geq \alpha_i \geq 0; \sum_{i=1}^{l} \alpha_i y_i = 0 \quad (9)$$

and the general classifier function is:

$$y = f(\mathbf{x}) = \text{sign}\left(\sum_i \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (10)$$

Some common kernels are linear, with $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})$; polynomial, with $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d$; and radial basis functions (RBFs), with

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

Note that one or more free parameters must be previously settled in the nonlinear kernels (polynomial degree d, width σ), together with the trade-off parameter C. This is usually done by cross validation, but, when working with small data sets, a dramatic reduction in the size of the training set arises, leading to poor generalization capabilities.

Bootstrap Resampling

We propose below an algorithm to adjust the free parameters in nonlinear SVM, based on the bootstrap resampling techniques [11], which preserves the size of the training data set. A dependence-estimation process between the pairs of data in a classification problem, where the data are drawn from a joint distribution $p(\mathbf{x}, y)$, is:

$$p(\mathbf{x}, y) \rightarrow \mathbf{V} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_l, y_l)\}. \quad (11)$$

The estimated SVM coefficients over the data set are:

$$\hat{\alpha} = s(\mathbf{V}, \theta) \quad (12)$$

where $s(,)$ symbolizes the SVM estimation process and θ represents a previously settled value of an SVM free parameter. The empirical risk is defined as the training error fraction for the current coefficients:

$$\hat{R}_{\text{emp}} = t(\hat{\alpha}, \mathbf{V}) \quad (13)$$

where $t(,)$ represents the empirical risk estimation process. A *bootstrap resample* is a data subset drawn from the training set following their empirical distribution; i.e., it consists of sampling with replacement of the observed pairs of data:

$$\hat{p}(\mathbf{x}, y) \rightarrow \mathbf{V}^* = \{(\mathbf{x}_1^*, y_1^*), (\mathbf{x}_2^*, y_2^*), \ldots, (\mathbf{x}_l^*, y_l^*)\}. \quad (14)$$

Consequently, $\mathbf{V}^*$ will contain elements of $\mathbf{V}$ appearing none, one, or several times. The resampling is repeated for $b = 1, \ldots, B$ times. A partition of $\mathbf{V}$ in terms of the resample $\mathbf{V}^*(b)$ is

$$\mathbf{V} = (\mathbf{V}_{\text{in}}^*(b), \mathbf{V}_{\text{out}}^*(b)) \quad (15)$$

with $\mathbf{V}_{\text{in}}^*(b)$ the subset of pairs included in the resample b and $\mathbf{V}_{\text{out}}^*(b)$ the subset of nonincluded pairs. The SVM coefficients for each resample will be given by:

$$\hat{\alpha}^*(b) = s(\mathbf{V}_{in}^*(b), \theta). \quad (16)$$

The empirical risk estimation over this population is known as its bootstrap replication:

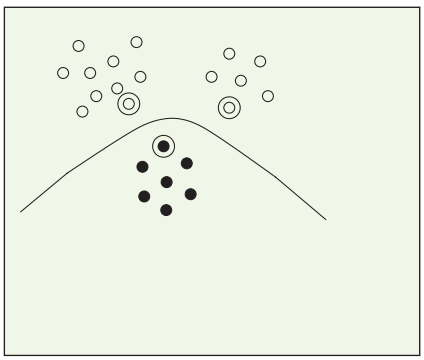
$$\hat{R}_{\text{emp}}^*(b) = t(\hat{\alpha}(b)^*, \mathbf{V}_{\text{in}}^*(b)) \quad (17)$$

and, through the B resamples histogram, it approximates the empirical risk density function. However, further advantage can be taken from:

$$\hat{R}^*(b) = t(\hat{\alpha}(b)^*, \mathbf{V}_{\text{out}}^*(b)), \quad (18)$$

which is a better approximation to the actual (i.e., whole and not only empirical) risk. A nonbiased estimate of the actual risk will be obtained by simply taking the replication average. This average estimate can be achieved for a set of values of the SVM free parameter, allowing determination of the most suitable value to train the SVM with the whole training set. A good range for B is typically 200 to 500 resamples.

Above we have assumed that θ represents either the trade-off C or the kernel parameter. This result only holds assuming they are mutually independent. Nonetheless, this will not be true in general. A good heuristic approach is to start with an intermediate value of C to give an initial guess of the kernel parameter, then estimate C again, and continue from one to another until a stable pair of parameters is obtained.



3. A simple 2-D classification example with nonlinear SVM; the support vectors appear as rounded.

Patients

Two different databases were assembled for the analysis, one of them (which we will refer to as *Base C*) for control (training) and the other (*Base D*) for validation.

Base C: Control Episodes

Twenty-six patients, with a third-generation ICD (Micro-Jewel 7221 and 7223, Medtronic), were included in this study. In these patients, monomorphic VT EGMs were obtained during an electrophysiologic study performed three days after the implant in the postabsorptive state. The EGM source between the subpectoral can and the defibrillation coil in the left ventricle was programmed, as it was previously shown to be the most appropriate electrode configuration for the criterion. The ICD pacing capabilities were used to induce monomorphic VT. The EGMs were stored in the ICD during induced sustained monomorphic VTs and during its preceding SRs. In order to obtain a group of SVTs, a treadmill test (modified Bruce protocol) was performed in the postabsorptive state, at least four days after the implantation procedure, if no contraindication was present. At least two therapy zones were programmed. The VT zone detection was activated at the peak of exercise if no angina, hypotension, or intense dyspnea were present. The detection interval was programmed 20 ms longer than the cycle length present at this particular moment. The EGMs recorded during VT, sinus tachycardia, and SR were downloaded in a computer system (A/D conversion: 128

Hz, 8 bits per sample, range ±7.5 mV). In this group, spontaneous tachycardias stored in the device during the follow-up were included if the EGM morphology of the recurrence was identical to either the induced VT morphology or the exercise-induced SVT morphology. A total of 38 SVT episodes (493 ± 54 ms) and 68 VT episodes (314 ± 49 ms) were analyzed. The EGM duration during SR was 90 ± 24 ms, with a range of (60, 144) ms.

Base D: Spontaneous Episodes

An independent group of spontaneous tachycardias from 54 patients with a double-chamber ICD (Micro-Jewel 7271, Medtronic) was assembled. We only admitted data from this type of device in order to reduce the diagnostic error during arrhythmia classification. Let $\mathbf{V}(A)$ denote the time for the ventricular (atrial) EGM; a VT was diagnosed when there was: a) V-A dissociation; b) irregular atrial rhythm, even if this was faster than a regular ventricular rhythm; or c) V-A association with a V-A < A-V. This allowed us to label 299 SVTs (498 ± 61 ms) and 1088 VTs (390 ± 81 ms) episodes.

The number of available episodes is high enough to be considered as significant in a clinical study. However, learning machines are usually trained with a far greater number of data in order to avoid the overfitting, hence the necessity of a regularization procedure in order to extend the SVM capabilities to clinical studies.

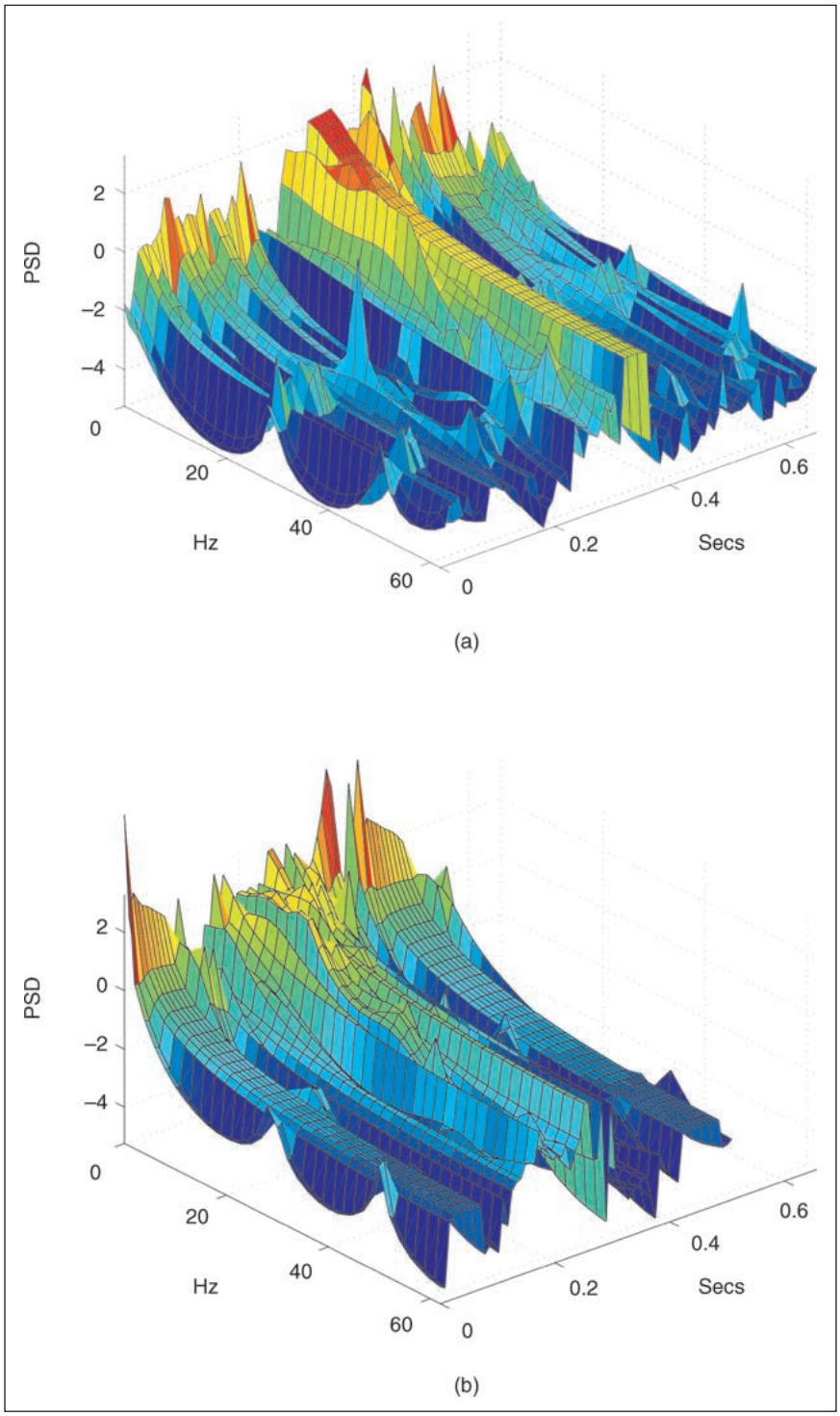
Algorithmic Implementation

Three issues arise when designing the algorithmic implementation of the ventricular EGM onset criterion, namely: where to measure (depending on time interval and bandwidth), which features to measure, and which preprocessing to use in order to enhance the differences in the onset.

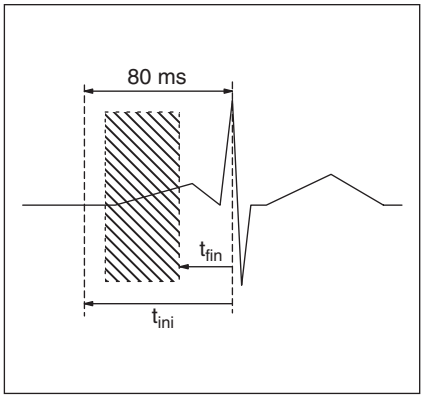
EGM Time-Frequency Analysis

The work hypothesis in the EGM onset criterion points to the analysis of the frequency components during the onset time. However, the EGM onset extends during a dramatically short time interval, about 40-80 ms, which for a 128-Hz sampled signal means just about ten time samples. This will lead to frequency indetermination, according to the uncertainty principle of the time-frequency analysis [12].

Qualitatively, Fig. 4 shows the comparison between two spectrograms, one from a SR beat and another from a VT beat. Both have been obtained in 110-ms segments, 90% overlapping, with an AR modeling on each segment; the model order was determined according to the Akaike Criterion [13]. In both cases, the R-wave appears nearby 300 ms. Note that for the SR beat the activation is a sudden, marked wall, occupying all the bandwidth, with a short duration of the previous activity. However, the VT beat shows a less energetic wall, with minor high-frequency content and lasting a shorter period of time; furthermore, the wideband activation is preceded by a longer low-frequency transient.



4. Time-frequency analysis for (a) an SR beat and (b) a VT beat.



5. Determination of the time interval for the analysis of the EGM onset.

Working in the time-frequency domain leads to a much higher input feature space. As all the information will be contained in the onset samples, it is preferable to separately analyze both domains' influence; i.e., to try to determine the time interval where the greater differences between SVT and VT arise, and to look for the most suitable frequency band to process.

Methods

The optimum time interval was determined by examining all the possible time subwindows inside the 86 ms previous to the R-wave of the EGM (Fig. 5), as this value was previously shown to contain the onset in every episode. The minimum time subwindow allowed was 20 ms (three samples). For each subwindow length and its generated feature space, a nonlinear (RBF) SVM classifier was built; the trade-off parameter was fixed to a moderate value in each case, $C = 100$, while the value of σ was individually adjusted with bootstrap resampling. The adequacy of each subwindow was quantified by the bootstrap estimates of actual sensitivity, specificity, and machine complexity (or number of support vectors). For notation, the R-wave was considered to take place at 0 ms.

The optimum band was assessed in a similar way. The feature space was given by the samples in the 86 ms previous to the R-wave; the EGMs were previously filtered by a bandpass FIR filter, order 32, and each possible sub-band (in the range from 0 to 45 Hz, step 5 Hz, minimum bandwidth 10 Hz) was analyzed by means of the SVM classifier performance as described above.

Results

Tables 1 and 2 show the results for the time interval and for the band analysis, respectively. In Table 1, sensitivity increases for initial times before −31 ms and for final times reaching 0 ms; specificity increases for initial times before −39 ms and remarkably for final times between −16 and −47 ms; and the number of support vectors decreases for initial times before −55 ms and for final times after −8 ms. In Table 2, sensitivity increases when the (15,30) Hz band is included; specificity increases when the (0,35) Hz band is included; and the number of support vectors decreases when the (0,20) Hz band is included.

A higher specificity in (−86,−16) ms is reasonably due to the presence of ventricular potentials during wide QRS, VT episodes. A higher sensitivity for ranges beginning soon enough (before −39 ms) and including the R-wave might be due in part to the fact that the SVM extracts the features in the time interval where the differences are less evident. On the other hand, a higher sensitivity in the (15,30) Hz band points to an evident concentration of the VT episodes in this band. However, a higher specificity can be achieved including practically the whole band, as SVT EGMs contain information in low frequency.

Table 1. Observation Subwindows for the Ventricular EGM Onset: (a) Bootstrap Averaged Sensitivity; (b) Bootstrap Averaged Specificity; (c) Bootstrap Averaged Percentage of Support Vectors (complexity). Start and End Window Time Are Expressed in ms.

– 70	85									
– 62	88	83								
– 55	93	91	76							
– 47	91	89	83	80						
– 39	90	89	86	87	88					
– 31	88	87	83	82	84	90				
– 23	90	88	84	86	83	84	86			
– 16	88	90	87	86	86	88	80	89		
– 8	91	89	89	88	89	85	85	85	83	
0	91	92	92	90	92	90	91	90	76	82
End/Start	–86	–78	–70	–62	–55	–47	–39	–31	–23	–16

(a)

– 70	92									
– 62	95	93								
– 55	96	93	93							
– 47	97	96	87	84						
– 39	99	95	91	92	93					
– 31	97	97	98	99	97	94				
– 23	98	98	97	96	94	93	87			
– 16	97	93	94	95	94	94	89	81		
– 8	96	95	96	95	94	94	91	83	75	
0	94	95	92	93	92	90	87	79	75	75
End/Start	–86	–78	–70	–62	–55	–47	–39	–31	–23	–16

(b)

– 70	44									
– 62	41	50								
– 55	41	45	49							
– 47	46	48	52	51						
– 39	52	54	52	50	47					
– 31	45	47	48	48	51	48				
– 23	51	41	41	44	44	50	53			
– 16	39	40	40	41	44	50	58	62		
– 8	41	39	41	40	43	48	52	61	66	
0	40	41	41	39	41	44	48	57	65	67
End/Start	–86	–78	–70	–62	–55	–47	–39	–31	–23	–16

(c)

Conclusions

A (– 80,0) ms time interval and all the band should be considered for the analysis of the ventricular EGM onset.

Featuring the EGM Onset

The first approaches to the ventricular EGM onset analysis were developed using a group of electrophysiologic parameters, chosen according to the clinical knowledge, and related to the beginning of the ventricular activation time (T_o) over the EGM first derivative. Figure 6 represents these parameters: first peak amplitude (A_1); time interval from T_o to the first peak (T_1); maximum amplitude (A_{max}); time interval from T_o to the maximum amplitude (T_{max}); averaged energies during the initial 20, 30, 40, and 50 ms (E20, E30, E40, and E50). This approach has two main problems: first, the automatic determination of the fiducial point T_o is not an easy task, especially in fast VT where the ending repolarization interferes with the following initial depolarization; second, any heuristic featuring will probably lead to a loss of information. Both issues were studied as described below.

Methods

The T_o for all the EGMs was automatically settled (see Appendix A); then, three different featuring methods of the EGM first derivative were established:

- automatic T_o detection and measurement of the electrophysiological parameters during tachycardia and its preceding SR;
- automatic T_o detection and featuring by the time samples in the following 80 ms; and
- R-wave detection and featuring by the time samples in the preceding 80 ms.

For each input space, a nonlinear (RBF) SVM classifier was designed (C = 100 for all of them, σ individually adjusted with bootstrap resampling).

Results

Table 3 shows the sensitivity, specificity, and complexity for each SVM detector. Classifier 1 needs a high number of support vectors. Classifier 2 shows the best performance in all the scores. Classifier 3 presents a similar specificity but a reduced sensitivity.

Conclusions

Heuristic featuring leads to information loss, and sample-based featuring is more appropriate. Previous T_o detection enhances the classification capabilities, but at the expense of a dramatically high computational burden (see Appendix A for details). Nevertheless, the performance of the R-wave-based scheme is not much lower than the optimal, and it is computationally simpler. A carefully designed preprocessing could allow us to adopt this more attractive scheme. This topic is the subject of the next section.

Enhancing by Preprocessing

Preprocessing can be a determining issue. We will focus here on the following steps.

- Our work hypothesis suggests the observation of changes through the EGM first derivative; still this corresponds to a rough high-pass filtering and it has to be shown that this does not degrade the algorithm performance.
- A previous discriminant analysis upon the electrophysiological features revealed the onset energies as significant, so that rectification could benefit the classification.
- Although R-wave synchronization has been used, synchronization with the maximum of the first derivative is also possible. This maximum will be denoted as the M_d wave.
- Inter-patient variability could be reduced by normalization with the SR maximum.

Methods

The averaged samples in the 80 ms previous to the synchronization wave were used as the feature space of an RBF SVM. Starting from a basic preprocessing scheme, where the EGM first derivative was obtained and R-wave synchronization was used, one preprocessing block was changed each time; rectification incorporation, first derivative removal, M_d synchronization, and SR normalization lead in each case to a different feature space and to a different SVM classifier (C

Table 2. Band Analysis: (a) Bootstrap Averaged Sensitivity; (b) Bootstrap Averaged Specificity; (c) Bootstrap Averaged Percentage of Support Vectors (Complexity). Start and End Window Time Are Expressed in ms.

10	83							
15	89	84						
20	89	91	96					
25	89	91	90	90				
30	91	90	92	89	85			
35	92	94	88	88	90	86		
40	89	91	90	94	92	89	83	
45	94	90	93	90	84	93	84	92
End/Start	0	5	10	15	20	25	30	35

(a)

10	88							
15	92	74						
20	93	93	87					
25	93	94	87	94				
30	96	96	89	95	89			
35	99	92	87	96	81	73		
40	94	93	86	94	86	81	77	
45	98	93	85	93	87	73	83	67
End/Start	0	5	10	15	20	25	30	35

(b)

10	45							
15	48	62						
20	42	54	46					
25	41	49	50	38				
30	39	48	54	35	38			
35	41	50	50	34	41	47		
40	39	48	54	36	41	49	65	
45	41	48	55	35	40	47	65	55
End/Start	0	5	10	15	20	25	30	35

(c)

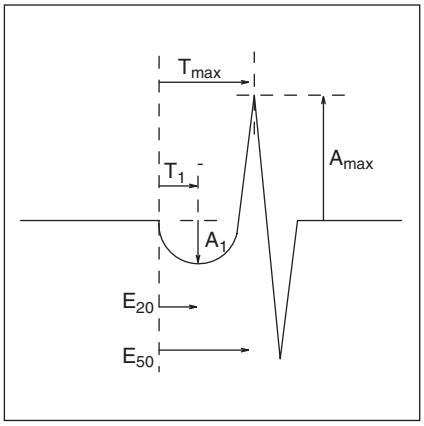
= 200) for all of them, individual σ for bootstrap tuning).

Results

Table 4 shows the sensitivity, the specificity, and the complexity for each classifier: neither rectifying nor the first derivative yield any advantage; M_d synchronization worsens all the classification rates, probably due to the higher instability of this fiducial point; finally, the SR normalization increases the complexity of the related SVM.

Conclusions

Nonlinear filters, such as the SVM, are robust when facing changes in the feature space that do not imply information cancellation. On the other hand, information distortion (like unstable synchronization) can deteriorate the classifier performance.



6. Clinical proposed parameters to feature the EGM onset.

Final Scheme

The algorithmic implementation in a set of patients is closed with the proposal of an automatic discrimination system, which is depicted in Fig. 7. The EGM goes through the next stages:

1) Noise filtering: cascade of a low-pass (45 Hz) and a notch (50 Hz) FIR 32.

2) Segmentation: includes a conventional beat detector, and it extracts the 80 ms previous to the R-wave to be used as the feature vector.

3) SR record: periodically stores the SR feature vector.

4) Commuter: allows to switch off the system (when the rate criterion is not assessed) and to switch on the periodic SR storing or the transmission of the arrhythmic beat.

5) Trained SVM classifier.

The optimum pair of values that minimize the error rate was iteratively adjusted by bootstrap resampling ($C = 10$, $\sigma = 5$). For this pair, empirical and bootstrap sensitivity, specificity and complexity were obtained for both the Base C (training) and the Base D (independent) data sets. The result is shown in Table 5. For the training set, there is both a high sensitivity and a high specificity, and all the SVT are correctly classified. For the independent data set, the output of the previously trained classifier agreed with the results in the training set in terms of specificity, but not in sensitivity. This was later observed to be due, in general, to new VT episodes showing a very different morphology from the Base C observations. Therefore, the learning procedure has correctly extracted the features in Base C, but it cannot generalize wisely when facing not previously observed VTs. Problems solved to the date with SVM have limited feature spaces, such as the handwritten alphabet or the handwritten digits, whereas the VTs with different anatomical origins cannot be considered as a limited set.

One possible solution to this drawback comes from the IL proposal discussed in the next section.

Table 3. Bootstrap Average ± Standard Deviation for the Proposed EGM Onset Featuring Schemes.

	Sensitivity	Specificity	% Support Vectors
Detector #1	79 ± 13	94 ± 7	89 ± 4
Detector #2	87 ± 12	97 ± 5	23 ± 4
Detector #3	79 ± 14	96 ± 6	36 ± 6

Table 4. Bootstrap Average ± Standard Deviation for the Proposed Preprocessing Schemes.

	Sensitivity	Specificity	% Support Vectors
Basic	90 ± 7	91 ± 10	56 ± 4
Rectifying	91 ± 7	90 ± 9	52 ± 47
M_d synchro	84 ± 7	76 ± 11	78 ± 4
N_o derivative	91 ± 6	93 ± 8	41 ± 4
SR normalize	89 ± 7	93 ± 8	99 ± 7

Incremental Learning

In the process of analyzing some diagnostic by way of a given feature, the first step, or analysis step, is to show this feature being predictive enough in a population set. The next one, or synthesis step, is to apply the criterion to a particular patient diagnostic, which can take further advantage of the patient's available information. This is a common operation in today's ICDs, where the cardiologist must program an algorithm by determining some patient-dependent thresholds. An *incremental learning* procedure is the process of establishing a trade-off between the population-available information and the patient-available information. Two main aspects to cope with are storing the population information in an effective, nonredundant way, and stating a population-individual trade-off criterion.

The first issue can be resolved within a SVM framework; the support vectors of the population classification represent the information in terms of the most critical, uncertain episodes. For the trade-off criterion, two models are proposed in this work: margin trade-off and cost trade-off. The first, still general, approach to the SVM cost function (in a Bayesian sense) can be found in [14].

Margin Learning

Let there be two feature vector sets in a classification problem. One of them represents the population boundary and contains the N_p support vectors of an SVM trained on a bigger set:

$$\{(\mathbf{x}_1^P, y_1^P), \ldots, (\mathbf{x}_{Np}^P, y_{Np}^P)\}. \quad (19)$$

The other includes the *Ni* feature vectors observed to the date in the patient:

$$\{(\mathbf{x}_1^I, y_1^I), \ldots, (\mathbf{x}_{Ni}^I, y_{Ni}^I)\}. \quad (20)$$

The relative importance of each subset in an SVM classifier can be controlled

through the subset margin; a higher margin for a subset will lead to a higher number of support vectors in the classifier. Assuming that the margin M_o is decreased in one subset and increased in the same scale in the other subset, the problem corresponds to maximize:

$$\frac{1}{2}||\mathbf{w}||^2+C\sum_{k=1}^{Np+Ni}\xi_k \tag{21}$$

constrained to:

$$y_i^P((\mathbf{w}\cdot\mathbf{x}_i^P)+b)1-\xi_i^P-M_o,\quad i=1,\ldots,Np \tag{22}$$

$$y_j^I((\mathbf{w}\cdot\mathbf{x}_j^I)+b)1-\xi_j^I+M_o,\quad j=1,\ldots,Ni \tag{23}$$

$$\xi_k\geq 0,\quad k=1,\cdots,Ni+Np. \tag{24}$$

An equivalent problem is to maximize:

$$L_d(\alpha_k,M_o)=-\frac{1}{2}\sum_{k,l=11}^{Ni+Np}\alpha_k\alpha_l y_k y_l K(\mathbf{x}_k,\mathbf{x}_l)-\sum_{i=1}^{Np}\alpha_i^P(1-M_o)-\sum_{j=1}^{Ni}\alpha_j^I(1+M_o) \tag{25}$$

constrained to:

$$\sum_{k=1}^{Np+Ni}\alpha_k y_k=0 \tag{26}$$

$$C\geq\alpha_k\geq 0. \tag{27}$$

The margin parameter M_o cannot be analytically calculated. Instead, it must be settled a priori. Nevertheless, it can be adjusted as an SVM free parameter by just minimizing the actual risk bootstrap estimation for the patient rate error:

$$\hat{R}_I^*(b)=\varphi(\alpha^*(b),\mathbf{V}_{exI}^*(b)) \tag{28}$$

where $\mathbf{V}_{exI}^*(b)$ represents the patient feature vectors, which are from resample b.

Cost Learning

Another possible approach to the SVM incremental learning consists of assigning a different cost to the population and to the individual feature vector losses. Let C_o be the trade-off parameter for the population-trained SVM. The problem now is to minimize:

$$\frac{1}{2}||\mathbf{w}||^2+C_oC_P\sum_{i=1}^{Np}\xi_i^P+C_oC_I\sum_{j=1}^{Ni}\xi_i^I \tag{29}$$

constrained to:

$$y_i^P((\mathbf{w}\cdot\mathbf{x}_i^P)+b)\geq 1-\xi_i^P,\quad i=1,\ldots,Np \tag{30}$$

$$y_j^I((\mathbf{w}\cdot\mathbf{x}_j^I)+b)\geq 1-\xi_j^I,\quad j=1,\ldots,Ni \tag{31}$$

$$\xi_k\geq 0,\quad k=1,\cdots,Ni+Np. \tag{32}$$

where C_p is the population loss cost and C_i the individual loss cost. Supposing the normalization condition $C_p\,C_i=1$, the ratio:

$$\frac{C_p}{C_i} \tag{33}$$

specifies the behavior of the algorithm. The problem is equivalent to maximize:

$$L_d(\alpha_k)=-\frac{1}{2}\sum_{k,l=1}^{Ni+Np}\alpha_k\alpha_l y_k y_l K(\mathbf{x}_k,\mathbf{x}_l)-\sum_{k=1}^{Ni+Np}\alpha_k \tag{34}$$

constrained to Eq. (26) and:

$$C_pC_o\geq\alpha_i\geq 0 \tag{35}$$

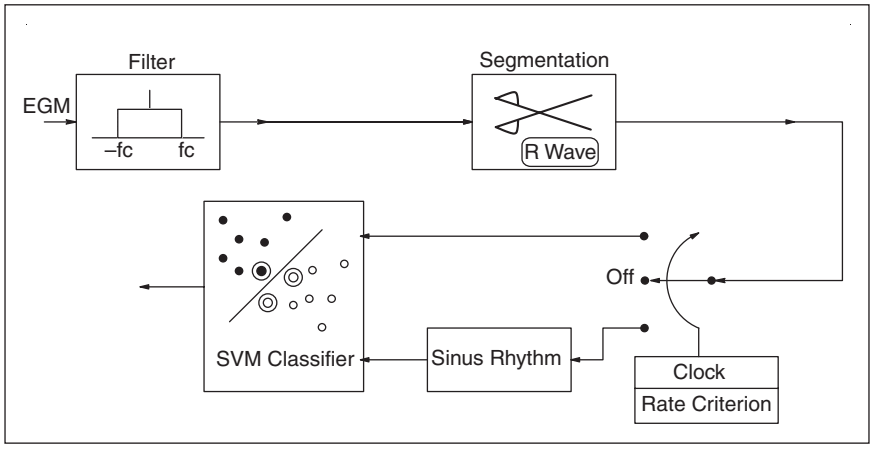
$$C_iC_o\geq\alpha_i\geq 0. \tag{36}$$

As in the preceding case, the cost ratio can be determined with bootstrap resampling on a set of SVM classifiers with different ratios.

A Single-Patient Case Study

Methods

An RBF SVM population classifier was built as described in the preceding section, on Base C, giving 35 support vectors (from 106 vectors); from these, 22 were saturated coefficients (8 SVTs and 14 VTs) and 13 were nonsaturated coefficients (5 SVTs and 8 VTs). A patient from Base D was selected, with a total of 257 episodes (77 SVTs, 180 VTs). About a 10% of the episodes were randomly selected (8 SVTs, 18 VTs), leaving apart the rest as a test set. The two SVM IL proposed methods were evaluated: a) margin learning, with M_o in (−1.5,1.5), discretized with resolution $\Delta M_o=0.1$; and b) cost learning, with a *Cp/Ci* ratio of (10^{-3},10^3), logarithmic scaling. Four different SVM soft outputs were represented for all of the patient episodes: a) conventional SVM and training set consisting of just the populational support vectors; b) conventional learning and training set consisting of the populational support vector plus observed (10%) individual episodes; and c) and d) margin and cost IL, as just described.



7. Proposed implementation for the ventricular EGM onset criterion.

Table 5. Final Scheme Classification Rates (Sensitivity, Specificity, and Complexity-Percentage of Support Vectors) in: (a) Control Set (Base C); (b) Validation Set (Base D).

Base C	Sen	Spe	% Nsv
Empirical	94.1	100	33
Bootstrap	91 ± 6	94 ± 7	41 ± 4

(a)

Base D	Sen	Spe
Empirical	76	94

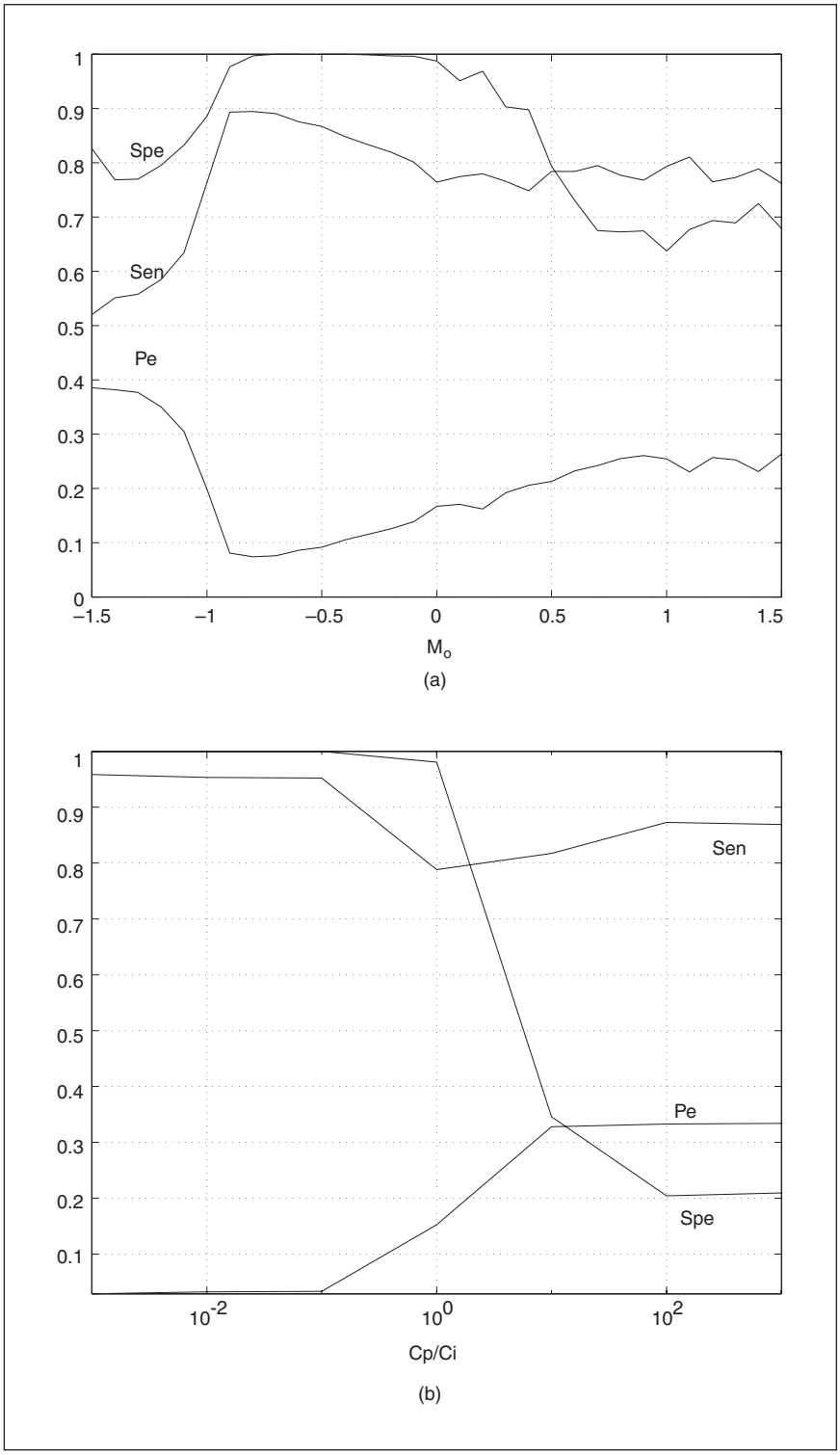
(b)

Results

Figure 8 depicts the bootstrap estimated sensitivity, specificity, and complexity for both schemes. With margin learning, the optimum value takes place at about $M_o = -0.75$ (sensitivity = 100%, specificity = 90%), whereas the range for cost learning is wider, $10^{-3} < Cp/Ci < 10^{-1}$ (sensitivity = 100%, specificity = 96%). Figure 9 shows the soft outputs for the mentioned SVM classifiers.

Discussion

Both margins point to the individual episodes being more significant than populational information in this case ($M_o < 0$ and $Cp/Ci << 1$). This can be due to a high enough number of available patient episodes. Nevertheless, cost learning offers a better performance, as it could be seen in the soft outputs.



8. Incremental learning results: (a) margin learning; (b) cost learning.

Conclusions

The analysis of changes in the ventricular EGM onset has been shown to accurately discriminate between SVT and VT in patients with ICD with high sensitivity and specificity performance. Simple preprocessing can enhance the performance of a nonlinear SVM classifier. The resulting algorithmic implementation has a low computational burden, meeting the requirements of present ICDs. Bootstrap resampling has been employed in order to avoid the overfitting of sample-based learning with a nonlinear SVM, making it an efficient analysis tool in a clinical data study. Two new methods of IL have been proposed for the patient-implementation of the algorithm, allowing an increase in the ventricular onset algorithm performance.

The proposed algorithm should be analyzed in more extensive data base sets. A special emphasis should be put on SVT episodes with presence of bundle branch block, as they usually cause morphological algorithms (such as the correlation waveform analysis and the width criterion) to fail. With respect to the algorithmic implementation, the fact that the criterion is enclosed into a black-box statistical model makes it unattractive for a medical professional, as long as the nature of the problem remains in the SVM formula. This subject will be studied in the companion article also in this issue [16].

Acknowledgments

Thanks to M. Ortiz for her kind help with the database management and thanks to F.J. Escribano for the useful comments and review on the manuscript.

José Luis Rojo-Álvarez received the bachelor's degree in telecommunication engineering from the Universidad de Vigo, Vigo, Spain, in 1996, and the Ph.D. in telecommunication from the Universidad Politécnica de Madrid, Madrid, Spain, in 2000. He is an assistant professor at the Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Spain. His research interests include statistical learning theory, digital signal processing, and complex system modelling, focusing on EKG and

intracardiac EGM signal processing, arrhythmia-genesis mechanisms, robust analysis of HRV, echocardiographic imaging, and hemodynamic function evaluation.

Ángel Arenal-Maíz received the Doctor en Medicina degree from the Universidad Complutense de Madrid in 1980, and he spent his residence term in the Hospital General Universitario Gregorio Marañón (HGU GM), Madrid, from 1981 to 1986. He received a cardiac electrophysiology fellowship at the Montreal Cardiology Institute in 1986-87 and then joined the Arrhythmia Unit of HGU GM. During 1991 and 1992 he was working at the Basic Electrophysiology Laboratory with Dr. Stanley Natel. Dr. Arenal is now in charge of the Arrhythmia Unit of the HGU GM, Madrid, Spain. His research interests are focused on the automatic discrimination of cardiac arrhythmia in implantable devices, on the study of arrhythmia mechanisms, and on clinical cardiac electrophysiology.

Antonio Artés-Rodríguez was born in Alhama de Almería, Spain, in 1963. He received the Ingeniero de Telecomunicación and Doctor Ingeniero de Telecomunicación degrees, both from the Universidad Politécnica de Madrid, Spain, in 1988 and 1992, respectively. He is now an associate professor at the Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Spain. His research interests include detection, estimation, and statistical learning methods and their application to signal processing, communication, and biomedicine.

Address for Correspondence: José Luis Rojo Álvarez, Depto Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, 28911-Leganés, Madrid. Spain. Phone: +34 1 624 87 69. Fax: +34 1 624 87 49. E-mail: jlrojo@tsc.uc3m.es.

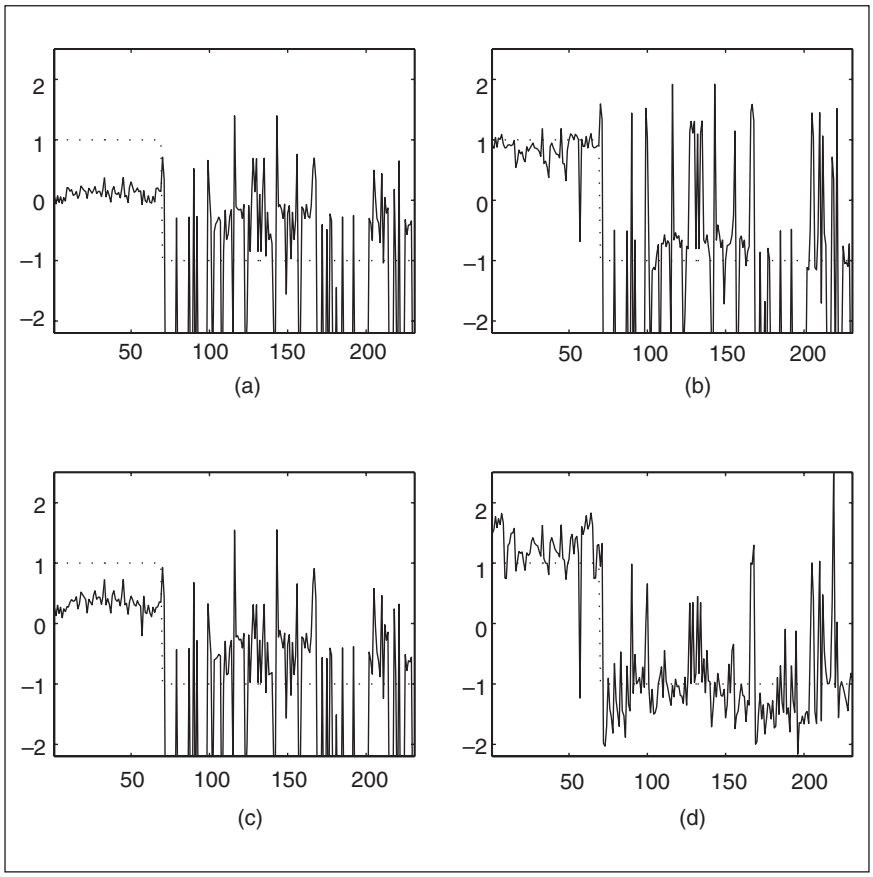
Appendix
Automatic EGM Onset Detector

A wide variety of methods have been suggested to automatically determine the QRS onset and offset timing. The ventricular EGM onset criterion initially seemed to need accurate knowledge of the QRS-onset time T_o, but neither threshold-based proceedings or slope considerations were accurate enough for SR, SVT, and VT beats simultaneously, due to the uncertainty in the determination of T_o in the fast rhythms where the ending ventricular repolarization interferes with the following initial depolarization. Therefore, a learning machine accomplishing this task was designed.

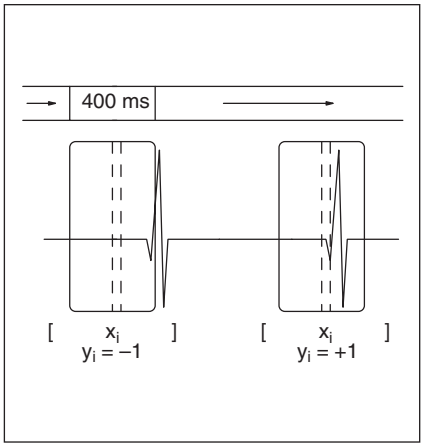
Methods

The proposed T_o detector was based on a sliding window applied to the EGM record; it was required to discriminate between T_o crossing the middle point of the window (high output-level target) and T_o either in any other position or its absence in the window (low output-level target), as depicted in Fig. 10. A 400-ms (51-sample) window was employed, since it was known to be long enough to determine T_o for an expert observer. The SR, SVT, and VT records were normalized in power and low-pass filtered (45 Hz). Each record window was considered as a training vector, yielding a total of 78.964 feature vectors (for all of the records in Base C). Onset smoothing was achieved by also considering as EGM onset the segments preceding and following the T_o segments. A nonlinear (RBF) SVM classifier was built with the large training set optimization procedure introduced in [15]. The value of the RBF width σ was settled heuristically to $\sigma = 7$. The resulting T_o detector was tested on a new set of spontaneous, ICD-recorded episodes from 12 patients (12 SR, 18 SVTs, and 55 VTs). However, the resulting computational burden was extremely high. An output example for an SVT episode is depicted in Fig. 11, and Table 6 shows the results of all the validation episodes. Note the averaged precision in the true positives (12.7 ms) being near the time resolution (7.8 ms).

Then, it is possible to build a high-precision QRS-onset detector with decision capabilities comparable to an expert human observer. It can cope with different kinds of rhythms (SR, SVTs, VTs). However, the computational burden of the detector is very high, due to 3.896 (9.8%) segments remaining as support vectors, and it is not suitable for ICD implementation. We need as much as 51 sums plus



9. Incremental learning. Soft output (continuous) and desired output (SVT = +1, VT = – 1) for the selected patient in Base D. The x-axes represent the number of patient episodes.



10. Automatic onset detector: sliding window for the feature input space.

one product per kernel evaluation, 3.896 kernel evaluation per segment, and 128 segments per second; near 26×106 operations per second. Clearly, it is desirable to have a reduction in the number of support vectors.

Although it can still be used as an automatic onset detector for this qualitative study, further work should be carried on to reduce the complexity. Nevertheless, this is still an interesting proceeding to detect fiducial points in ECG analysis systems.

References

[1] B.N. Singh, "Controlling cardiac arrhythmias: An overview with a historical perspective," *Am. J. Cardiol.,* vol. 80, no. 8A, pp. 4G-15G, 1997.

[2] I. Singer, *Implantable Cardioverter Defibrillator*. Mt. Kisco, NY: Futura, 1994.

[3] T. Klingenheben, C. Sticherling, M. Skupin, S.H. Hohnloser, "Intracardiac QRS electrogram width. An arrhythmia detection feature for implantable cardioverter defibrillator. Exercise induced variation as a base for device programming," *Pacing Clin. Electrophysiology*, vol. 8, no. 2, pp. 1609-1617, 1998.

[4] J.M. Jenkins and S.A. Caswell, "Detection algorithms in implantable cardioverter defibrillators," *Proc. IEEE*, vol. 84, pp. 428-445, 1996.

[5] J. Neuzner, H.F. Pitschner, and M. Schlepper, "Programmable VT detection enhancements in implantable cardioverter defibrillator therapy," *Pacing Clin. Electrophysiology*, vol. 18, no. 3, Part 2, pp. 539-547, 1995.

[6] A. Schaumann, F. von zur Muhlen, B.D. Gonska, and H. Kreuzer, "Enhanced detection criteria in implantable cardioverter defibrillators to avoid inappropriate therapy," *Am. J. Cardiol.*, vol. 78, no. 12, pp. 42-50, 1996.

[7] J.L. Rojo-Álvarez, A. Arenal, A. Artés, J. Villacastín, and J. Almendral, "Discrimination between ventricular and supraventricular tachycardia based on implantable defibrillator stored electrogram analysis," in *Proc. 47th Amer. College of Cardiology*, vol. 47, Mar. 1998., pp. 294A.

[8] B. Schölkopf and K. Sung, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Trans. Signal Processing*, vol. 45, pp. 2758-2765, 1997.

[9] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 637-646, 1998.

[10] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowl. Discovery*, vol. 2, no. 2, pp. 1-32, 1998.

[11] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*. London, U.K.: Chapman & Hall, 1998.

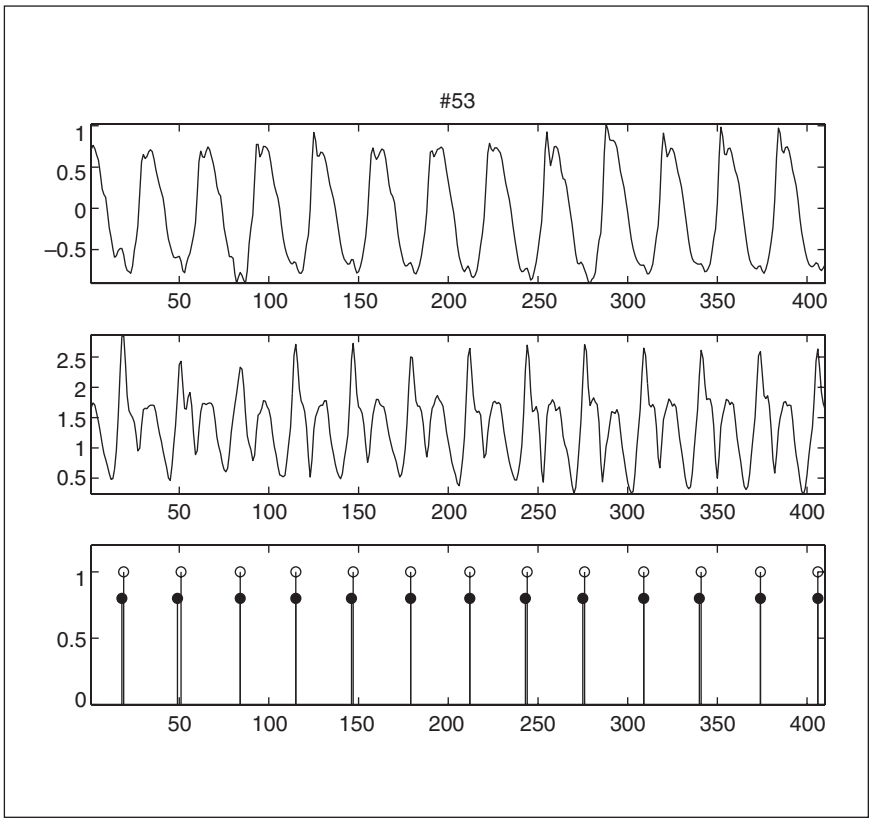
[12] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1995.

[13] S.L. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, N.J.: Prentice-Hall, 1987.

[14] G. Whaba, Y. Lin, and H. Zhang, "Support Vector Machines, reproducing kernel Hilbert spaces and the randomized GACV," in *Advances in Kernel Methods: Support Vector Learning,* A.J. Smola, B. Schölkopf, and D. Schurmans, Eds. Cambridge, MA: MIT Press, 1999.

[15] E. Osuna and F. Girosi, "Reducing the run-time complexity in Support Vector Machines," in in *Advances in Kernel Methods: Support Vector Learning,* A.J. Smola, B. Schölkopf, and D. Schurmans, Eds. Cambridge, MA: MIT Press, 1999.

[16] J.L. Rojo-Álvarez, A. Arenal-Maíz, and A. Artés-Rodríguez, "Support vector black-box interpretation in ventricular arrhythmia discrimination," *IEEE Eng. Med. Biol. Mag.*, vol. 21, pp. 27-35, Jan./Feb. 2002.



11. Automatic onset detector: example episode. Top: VT record. Middle: detector soft output. Bottom: T_o sequences as determined by a human expert assessment (white circles) and by the automatic detector (black circles). The x-axes represent the sample number.

Table 6. Percentage of True Positives (Tp), False Positives (Fp) and False Negatives (Fn) for the Automatic EGM Onset Detector (Mean ± std on the Record Detection Averages).

	Tp	Fp	Fn
SR	100 ± 0	0 ± 0	0 ± 0
SVT	98.0 ± 0.0	0.02 ± 0.04	0 ± 0
VT	91.5 ± 15.4	8.1 ± 15.7	3.6 ± 1.8