Arrhythmia Detection Using Signal-Adapted Wavelet Preprocessing for Support Vector Machines

D Strauss[1,2], G Steidl[2], J Jung[3]

[1]Key Numerics, Germany
[2]Faculty of Mathematics and Computer Science, University of Mannheim, Germany
[3]Internal Medicine III, University Hospital Homburg/Saar, Germany

Abstract

Rate based arrhythmia recognition algorithms in implantable cardioverter–defibrillators are of limited reliability in some clinical situations. Here the inclusion of morphological features of endocardial electrograms can improve the performance. In this study, we present a coupled signal–adapted wavelet–support vector machine (SVM) arrhythmia detection scheme.

Within the scope of an electrophysiological examination, data segments were recorded during normal sinus rhythm (NSR) and ventricular tachycardia (VT). Consecutive beats were selected as morphological activation patterns of NSR and VT. These patters were represented by their multilevel concentrations. For this, a signal–adapted and highly efficient lattice structure based wavelet decomposition technique was employed which maximizes the class separability and takes the final classification of NSR and VT by SVMs with radial compactly supported kernels into account.

In an automated analysis of an independent test–set, our hybrid scheme outperformed other methods and classified all patterns correctly without overlap.

1. Introduction

Sudden cardiac death is a major public health concern worldwide. The implantable cardioverter–defibrillator (ICD) is an automated antitachycardia device and accepted to be the most effective therapy for preventing sudden cardiac death due to tachyarrhythmias [1]. Usually, the information of the endocardial electrogram (EE) utilized by an ICD is the heart rate. However, the rate is of limited reliability in some clinical situations. Although additional detection enhancements are used in third generation ICD–systems, inappropriate ICD therapy occurs in up to 13% of the patients who received such a device [2].

A major challenge for rate–algorithms used in these devices is the discrimination of ventricular tachycardia (VT) with 1:1 retrograde conduction from sinus tachycardia. Here time–domain methods based on template matching [3] or neural networks [4] can be used. A drawback of these methods is that the classification takes place in the original signal space where the dimensionality is often high and features being irrelevant for classification are under consideration. Recently, the superiority of wavelet decompositions before the classification over the direct application of the classifier on the original signal space was shown, e.g., see [5]. Especially, the multilevel concentrations of adapted wavelet decompositions have recently proven to be suitable for arrhythmia detection [6, 7].

In this paper, we present a new hybrid wavelet–support classifier for the detection of VT. Our scheme makes use of lattice structure based signal–adapted wavelet decompositions and support vector machines (SVMs) with radial compactly supported kernel.

2. Methods

2.1. Data segments

Bipolar EEs were obtained from the apex of the right ventricle using the distal pair of a 6-F quadripolar electrode catheter in 10 patients with inducible monomorphic VT. The EEs were amplified (HBV 20, Biotronik, Berlin, Germany), filtered (10–500 Hz), and digitized with 2kHz, 12 bit resolution (DT 2824–PGH, Data Translation, Marlboro, MA, USA). Data segments (DS) of 10 s duration were recorded during normal sinus rhythm (NSR) and VT. Consecutive beats were selected as morphological patterns of NSR (240 beats) and VT (240 beats) within a time–frame of 256ms.

2.2. Feature spaces induced by kernels

The SVM is a novel type of learning machine and very promising for pattern recognition. Basically, it relies on the well known optimal hyperplane classification, i.e., the separation of two classes of points by a hyperplane such

Computers in Cardiology 2001;28:497-500.

that the distance of distinct points from the hyperplane, the so–called *margin*, is maximized. SVMs utilize this linear separation method in very high dimensional feature spaces induced by reproducing kernels to obtain a nonlinear separation of original patterns. We do not discuss the fundamentals of SVMs here and recommend [8, 9] for a tutorial introduction. We only present feature maps and feature spaces in the SVM context that we need for our hybrid strategy.

Let $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ be a positive definite symmetric function in $L^2(\mathcal{X} \times \mathcal{X})$. For a given K, there exists a *reproducing kernel Hilbert space*

$$\mathcal{H}_K = \overline{\mathrm{span}\,\{K(\tilde{\mathbf{x}},\cdot) : \tilde{\mathbf{x}} \in \mathcal{X}\}}$$

of real valued functions on $\mathcal{X}$ with inner product determined by $\langle K(\tilde{\mathbf{x}},\mathbf{x}), K(\bar{\mathbf{x}},\mathbf{x})\rangle_{\mathcal{H}_K} = K(\tilde{\mathbf{x}},\bar{\mathbf{x}})$ which has the reproducing kernel K, i.e., $\langle f(\cdot), K(\tilde{\mathbf{x}},\cdot)\rangle_{\mathcal{H}_K} = f(\tilde{\mathbf{x}}) \quad (f \in \mathcal{H}_K)$. By *Mercer's Theorem*, the reproducing kernel K can be expanded in a uniformly convergent series on $\mathcal{X} \times \mathcal{X}$

$$K(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{\infty} \eta_j \varphi_j(\mathbf{x}) \varphi_j(\mathbf{y}), \tag{1}$$

where $\eta_j \geq 0$ are the eigenvalues of the integral operator $T_K : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ with $T_K f(\mathbf{y}) = \int_{\mathcal{X}} K(\mathbf{x},\mathbf{y}) f(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and where $\{\varphi_j\}_{j\in\mathbb{N}}$ are the corresponding $L^2(\mathcal{X})$–orthonormalized eigenfunctions. We restrict our interest to functions K that arise from a radial basis function (RBF). In other words, we assume that there exists a real valued function k on $\mathbb{R}$ so that

$$K(\mathbf{x},\mathbf{y}) = k(||\mathbf{x}-\mathbf{y}||_2), \tag{2}$$

where $||\cdot||_2$ denotes the Euclidean norm on $\mathbb{R}^d$.

We introduce a so–called *feature map* $\Phi : \mathcal{X} \to \ell^2$ by

$$\Phi(\cdot) = \left(\sqrt{\eta_j}\varphi_j(\cdot)\right)_{j\in\mathbb{N}}.$$

Let ℓ^2 denote the Hilbert space of real valued quadratic summable sequences $a = (a_i)_{i\in\mathbb{N}}$ with inner product $\langle a,b\rangle_{\ell^2} = \sum_{i\in\mathbb{N}} a_i b_i$. By (1), we have that $\Phi(\mathbf{x})$ $(\mathbf{x} \in \mathcal{X})$ is an element in ℓ^2 with

$$||\Phi(\mathbf{x})||_{\ell^2}^2 = \sum_{j=1}^{\infty} \eta_j \varphi_j^2(\mathbf{x}) = K(\mathbf{x},\mathbf{x}) = k(0).$$

We define the *feature space* $\mathcal{F}_K \subset \ell^2$ by the ℓ^2–closure of all finite linear combinations of elements $\Phi(\mathbf{x})$ $(\mathbf{x} \in \mathcal{X})$

$$\mathcal{F}_K = \overline{\mathrm{span}\,\{\Phi(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}}.$$

Then $\mathcal{F}_K$ is a Hilbert space with $||\cdot||_{\mathcal{F}_K} = ||\cdot||_{\ell^2}$. The feature space $\mathcal{F}_K$ and the reproducing kernel Hilbert space $\mathcal{H}_K$ are isometrically isomorph with isometry $\iota : \mathcal{F}_K \to \mathcal{H}_K$ defined by $\iota(\mathbf{w}) = f_{\mathbf{w}}(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x})\rangle_{\ell^2} = \sum_{j=1}^{\infty} w_j \sqrt{\eta_j}\varphi_j(\mathbf{x})$.

2.3. Adaptation in feature spaces

Now we introduce our adaptation strategy for feature spaces that is based on wavelets and filter banks. We do not discuss the fundamentals of filter banks and wavelets here and refer to [10, 11] for an introduction. Let $G_0(z)$ and $G_1(z)$ be the synthesis filters of a normalized paraunitary two–channel FIR filter bank with real filter coefficients and a zero mean highpass. When cascading such a two–channel building block in an octave–band tree, the filters of an equivalent parallel structure are given by $Q_{j,0}(z) = \prod_{m=0}^{j-1} G_0(z^{2^m})$ and $Q_{j,1}(z) = G_1(z^{2^{j-1}}) \prod_{m=0}^{j-2} G_0(z^{2^m})$. Let us denote the translations of the impulse responses $q_{j,k}[\cdot]$ of these filters by $\mathbf{q}_{j,i}^m = \left(q_{j,i}[k-2^j m]\right)_{k\in\mathbb{Z}}$ $(i = 0,1)$ and let J be the maximal decomposition depth. Then the set $\left\{\mathbf{q}_{J,0}^m, \mathbf{q}_{j,1}^m : j = 1,\ldots,J; m \in \mathbb{Z}\right\}$ constitutes an orthonormal basis for ℓ^2 and an arbitrary sequence $\mathbf{x} \in \ell^2$ can be decomposed as

$$\mathbf{x} = \sum_{m\in\mathbb{Z}} d_{J,0}[m]\,\mathbf{q}_{J,0}^m + \sum_{j=1}^{J}\sum_{m\in\mathbb{Z}} d_{j,1}[m]\mathbf{q}_{j,1}^m.$$

We denote the *wavelet coefficients* by $\mathbf{d}_j = (d_{j,1}[m])_{m\in\mathbb{Z}}$.

All paraunitary filter banks can be parameterized by the lattice structure [10]. When using this parameterization for a finite impulse response two–channel building block of order 5 with zero mean highpass, the filters and thus the wavelet coefficients depend on two lattice angles $\boldsymbol{\vartheta}$ in parameter space $\mathcal{P} = \{\boldsymbol{\vartheta} = (\vartheta_0,\vartheta_1) : \vartheta_0,\vartheta_1 \in [0,\pi[\}$ see [12, 6] for details. We use the superscript $\boldsymbol{\vartheta}$ to denote this dependence. Let a set of M endocardial waveforms $\mathbf{x}_i \in \mathcal{X} \subset \mathbb{R}^N$ be given which belong to two distinct classes with corresponding labels $y_i \in \{-1,1\}$ $(i = 1,\ldots,M)$. By M_+ and M_- we denote the sets of indices $i \in \{1,\ldots,M\}$ with $y_i = 1$ and $y_i = -1$, respectively. We label NSR waveforms by 1 and VT waveforms by -1.

For a fixed EE waveform $\mathbf{x}$ we define the function $\boldsymbol{\xi}_{\mathbf{x}} : \mathcal{P} \to \mathbb{R}^J$

$$\boldsymbol{\xi}_{\mathbf{x}}(\boldsymbol{\vartheta}) = (\xi_1(\boldsymbol{\vartheta}),\ldots,\xi_J(\boldsymbol{\vartheta})) = \left(||\mathbf{d}_1^{\boldsymbol{\vartheta}}||_{\ell^p}^p,\ldots,||\mathbf{d}_J^{\boldsymbol{\vartheta}}||_{\ell^p}^p\right)$$

and set $\boldsymbol{\xi}_i(\boldsymbol{\vartheta}) = \boldsymbol{\xi}_{\mathbf{x}_i}(\boldsymbol{\vartheta})$ $(i = 1,\ldots,M)$. In necessary, we may also leave out levels here. This function carries the multilevel concentration of an EE waveform $\mathbf{x}$. Multilevel concentrations belong to a very low dimensional pattern space and are robust against local instabilities in time. Thus, by using such multilevel concentrations for SVMs, we can incorporate a prior knowledge about the EE waveforms. Now we intend to find $\boldsymbol{\vartheta}$ so that

$$\mathcal{A}(\boldsymbol{\vartheta}) = \left\{(\boldsymbol{\xi}_i(\boldsymbol{\vartheta}), y_i) \in \mathcal{X} \subset \mathbb{R}^d \times \{\pm 1\} : i = 1,\ldots,M\right\}$$

is a 'good' training set for a SVM. It is the fundamental concept of SVMs that we expect a good generalization performance if they have a large margin [8, 9].

Our strategy is now to obtain multilevel concentrations that are mapped to far apart points in $\mathcal{F}_K$ for the distinct classes and result in large margin of the SVM. Consequently, we try to find $\hat{\vartheta}$ such that

$$\hat{\vartheta} = \arg\max_{\vartheta \in \mathcal{P}} \left\{ \min_{i \in M_+, j \in M_-} ||\Phi(\xi_i(\vartheta)) - \Phi(\xi_j(\vartheta))||^2_{\mathcal{F}_K} \right\}. \tag{3}$$

By the definition of the inner product in $\mathcal{F}_K$ and (2) it follows that

$$\begin{aligned} &||\Phi(\xi_i(\vartheta)) - \Phi(\xi_j(\vartheta))||^2_{\mathcal{F}_K} \\ &\quad = ||\Phi(\xi_i(\vartheta))||^2_{\mathcal{F}_K} + ||\Phi(\xi_j(\vartheta))||^2_{\mathcal{F}_K} \\ &\qquad -2 \langle \Phi(\xi_i(\vartheta)), \Phi(\xi_j(\vartheta)) \rangle_{\mathcal{F}_K} \\ &\quad = 2k(0) - 2k\left(||\xi_i(\vartheta) - \xi_j(\vartheta)||_2\right). \end{aligned}$$

We suppose that $k(\cdot)$ is monotonely decreasing in $|\cdot|$. Then (3) can be rewritten as

$$\hat{\vartheta} = \arg\max_{\vartheta \in \mathcal{P}} \left\{ \min_{i \in M_+, j \in M_-} ||\xi_i(\vartheta) - \xi_j(\vartheta)||_2 \right\}. \tag{4}$$

For a further simplification of this optimization problem, we reduce the sets M_+ and M_- by averaging the patterns of the respective classes and introduce the notation $\xi_\pm = \frac{1}{|M_\pm|} \sum_{i \in M_\pm} \xi_i$. Instead of (4) we search for $\hat{\vartheta}$ with

$$\hat{\vartheta} = \arg\max_{\vartheta \in \mathcal{P}} \left\{ ||\xi_+(\vartheta) - \xi_-(\vartheta)||_2 \right\}.$$

We applied a genetic algorithm [13] for the solution of this optimization problem with a binary string encoding.

3. Results

In our experiments, we work with a maximal decomposition depth of $J = 8$. We apply the Gaussian RBF $k(t) = e^{-\frac{t^2}{2s^2}}$. We utilize Wendland's compactly support RBFs [14] for our adapted approach. Note that compactly supported RBFs have not been used for the construction of SVMs up to now. In particular, we apply the RBF (up to multiplications with constants) $k_{7,2}(x) = 1/3\,(80\,x^2 + 27\,x + 3)\,(1-x)^9_+ \in C^4(\mathbb{R}^8)$, where $C^4(\mathbb{R}^8)$ denotes the space of 4 times continuously differentiable functions on $\mathbb{R}^8$. We will scale this function by $k_{7,2}(\cdot/s)$ $(s \in \mathbb{R}_+)$. Note that Wendland's RBFs can easily be calculated and evaluated for relatively small space dimensions due to the construction scheme in [14]. For our multilevel concentrations these functions are well suited as we have dimensionality of only 8.

We separate the total data set of 480 beats in a training set of 160 beats (SR: 80, VT: 80) and a test set of 320 beats (SR: 160, VT: 160). Thus, for each patient we have a training set of $|M_-| = |M_-| = 8$ waveforms. The remaining set of 32 beats forms an independent test set for the individual patient.

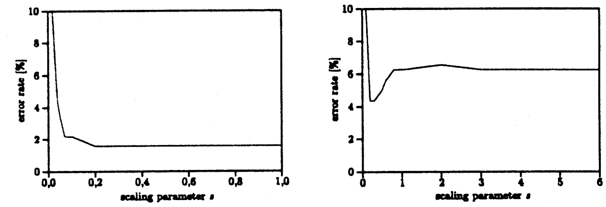

Figure 1. The error rate of a SVM applied to the original waveforms with Gaussian Kernel (left) and to the non–adapted multilevel concentrations with Wendland's function $k_{7,2}$ (right).

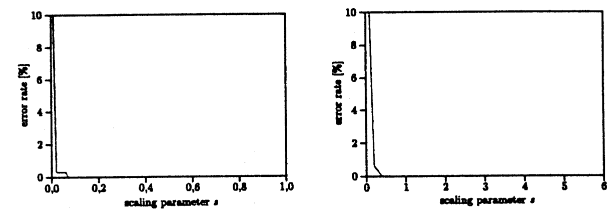

Figure 2. The error rate of a SVM applied to the adapted multilevel concentrations with Gaussian kernel (left) and to the adapted multilevel concentrations with Wendland's function $k_{7,2}$ (right).

In the Figures 1 and 2, we have plotted the error rate in dependence on the scaling factor s for different RBFs. The error rate [%] is determined for all patients in common, that is, the ratio of false classifications on the whole test set to the total number of 320 examples within this set. A comparison of Figure 1 with Figure 2 with respect to the error rate shows that our adapted algorithm with both the Gaussian and Wendland's function is significantly superior to the original SVM and to the SVM on non–adapted multilevel concentrations where we used a Daubechies wavelet with three vanishing moments, a well accepted standard wavelet, see [15].

In Figure 3 (left) we have shown the results for the correlation waveform analysis with best fit alignment [3], a well accepted standard method in morphological arrhythmia recognition. Here a time–frame of 20ms was used for the best fit alignment strategy and the same training set of NSR (as in our experiments above) for constructing the template. In Figure 3 (right) we have shown the results for our hybrid wavelet–support vector classifier with $k_{7,2}$ and $s = 0.8$. For the correlation waveform analysis, the waveforms of NSR and VT do heavily overlap for some patients. For our hybrid wavelet–support vector

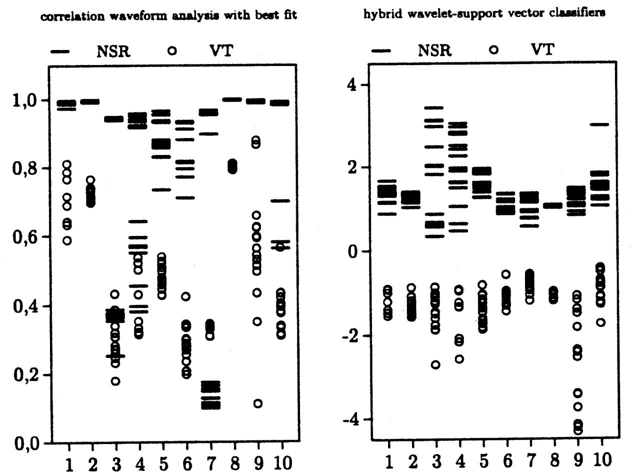

Figure 3. The individual analysis of the 10 patients. Left hand: correlation waveform analysis with best fit alignment. Right hand: our hybrid wavelet–support vector classifier with $k_{7,2}$ and $s = 0.8$.

classifier, the two groups are clearly discriminated without any overlap. This shows that our new method easily outperform the conventional approach.

4. Conclusion

We have developed hybrid wavelet–support vector classifiers for the detection of arrhythmias. Our scheme allows for an inclusion of prior knowledge about EE waveforms, namely local instabilities in time. We have derived an adaptation strategy in feature space induced by radial kernels.

Our scheme classified a large independent test–set of NSR and VT pattern correctly without overlap. This result was neither achievable with SVMs on the raw data nor by the use of a non–adapted standard wavelet. Our new detection scheme significantly outperformed the well accepted correlation waveform analysis with best fit alignment. We also firstly investigated compactly supported RBFs for the construction of SVMs which performed equal to the well accepted Gaussian kernels.

We conclude that hybrid–wavelet support vector classifiers are powerful for arrhythmia detection. The use of the very efficient lattice structure implementation and SVM evaluation of the multilevel concentrations also supports an implementation in implantable devices.

References

[1] The Antiarrhythmics Versus Implantable Defibrillators (AVID) Investigators. A comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from near fatal ventricular arrhythmias. N Engl J Med 1997;337:1576–1583.

[2] Weber M, Böcker D, Bansch D, Brunn J, Castrucci M, Gradaus R, Breithardt G, Block M. Efficacy and safety of the initial use of stability and onset criteria in implantable cardioverter defibrillator. J Cardiovascular Electrophysiology 1999;10:145–153.

[3] Throne RD, Jenkins JM, Winston SA, Finelli CJ, DiCarlo LA. Discrimination of retrograde from antegrade atrial activation using intracardiac electrogram waveform analysis. Pacing and Clinical Electrophysiology 1989;12:1622–1630.

[4] Leong PHW, Jabri MA. MATIC–An intracardiac tachycardia classification system. Pacing and Clinical Electrophysiology 1992;15:1317–1331.

[5] Saito N, Coifman RR. Local discriminant bases and their applications. J Mathematical Imaging and Vision 1995; 4:337–358.

[6] Strauss D, Jung J, Rieder A, Manoli Y. Classification of endocardial electrograms using adapted wavelet packets and neural networks. Annals of Biomedical Engineering 2001; 29:483–492.

[7] Jung J, Strauss D, Siaplaouras S, Buob A, Sinnwell T, Manoli Y, Schieffer H, Heisel A. Wavelet analysis of bipolar endocardial electrograms for morphology based detection of ventricular tachycardias. In Murray A, Swiryn S (eds.), IEEE Computers in Cardiology (26). 2000; 347–351.

[8] Vapnik V. The Nature of Statistical Learning Theory. NY: Springer, 1995.

[9] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines. Cambridge University Press, 2000.

[10] Vaidyanathan PP. Multirate Systems and Filter Banks. Englewood Cliffs, NJ: Prentice Hall, 1993.

[11] Vetterli M, Kovačević J. Wavelets and Subband Coding. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[12] Strauss D, Sinnwell T, Rieder A, Manoli Y, Jung J. A promising approach to morphological endocardial signal discriminations: Adapted multiresolution signal decompositions. Applied Signal Processing 1999;6:182–193.

[13] Goldberg DE. Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, MA: Addison–Wesley, 1989.

[14] Wendland H. Piecewise polynomial, positive definite functions and compactly supported functions radial basis functions of minimal degree. Adv in Comp Math 1995; 4:389–396.

[15] Daubechies I. Orthonormal bases of compactly supported wavelets. Comm Pure Appl Math 1988;41:906–966.

Address for correspondence:

Daniel Strauss
Faculty of Mathematics and Computer Science
University of Mannheim
Building D7,27, Room 201
D-68131 Mannheim, Germany
tel./fax: ++49-6831-702247/48
strauss@keynumerics.com