Hybrid Wavelet Packet – Support Vector Classification of Atrial Activation Patterns

DJ Strauss[1,2,3], J Jung[3]

[1]Key Numerics, Saarbrücken, Germany
[2]Institute for New Materials, Saarbrücken, Germany
[3]Saarland University Hospital, Homburg/Saar, Germany

Abstract

The discrimination of ventricular tachycardia (VT) with 1:1 retrograde conduction from sinus tachycardia often remains a challenge for rate–based arrhythmia recognition algorithms commonly used in dual–chamber implantable cardioverter defibrillators (ICDs). In this study, we propose hybrid wavelet packet–support vector classifiers for the recognition of atrial activation patters as rate independent approach.

Consecutive beats representing antegrade (AA) and retrograde atrial (RA) activation patterns within data segments of 10s duration were supplied to a binary hybrid wavelet packet–support vector classifier with radial kernel and morphologically adapted wavelet packets followed by a regularization scheme to exclude ectopic beats and artifacts. This system utilizes an optimal representation of the EE waveforms and inherently minimizes an upper bound on the generalization error.

All data segments of an independent test set with AA and RA sequences were classified correctly by the proposed scheme. The possible use of this scheme in dual-chamber ICDs may increase the specificity of the VT detection by these devices.

1. Introduction

The implantable cardioverter–defibrillator (ICD) is accepted to be the most effective therapy for preventing sudden cardiac death due to ventricular tachycardias (VT) [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 and despite the use of detection enhancements like *rate stability* or *sudden onset* in modern ICD–systems, inappropriate ICD–therapy occur in up to 13% of the patients who received such a device [2]. Whether the use of the recently introduced dual–chamber ICDs [3] will contribute to a reduction of inadequate therapies is currently under examination.

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 [4] or neural networks [5] 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.

Wavelet decompositions have proven to be suitable for dimensionality reduction, e.g., see [6, 7] and recently the the multilevel concentrations of adapted wavelet decompositions have been used to provide robust feature vectors for arrhythmia detection [8, 9].

In this paper, we present a new hybrid wavelet packet–support classification the antegrade atrial (AA) activation and retrograde atrial (RA) activation. For this, we combine the wavelet packet based feature extraction proposed in [8] with the very recently introduced hybrid wavelet–support vector classifiers [10, 11]. This approach is theoretically well founded and effective in practice.

2. Methods

2.1. Data segments

The data segments used in this study were obtained during a clinically indicated electrophysiological examination. A written consent was obtained from all patients. The patient population was composed as follows: One group with typical atrio–ventricular (AV) nodal tachycardia, studied as a model for patients having a spontaneous retrograde activation, and another group with clinical monomorphic VT.

In both groups, bipolar endocardial signals were obtained from the high right atrium using the distal pair of a 6–French quadripolar electrode catheter with an inter–electrode spacing of 0.5 cm (USCI, Bard, Billerica, MA, USA). These EEs were recorded during sinus rhythm and induced or spontaneously occurring AV nodal tachycardia or during induced monomorphic VT with 1:1 retrograde conduction. The endocardial recordings were amplified (HBV 20, Biotronik, Berlin, Germany), bandpass filtered (40–500 Hz) and digitized at 2 kHz with a 12 bit resolution (DT 2824–PGH, Data Translation, Marlboro, MA, USA). Data segments of 10 seconds duration were stored for subsequent analysis.

2.2. RKHS regularization

The SVM is a novel type of learning machine and very promising for pattern recognition. In the following, we restrict our interest to the hard margin SVM and refer to [12, 13] for more general discussions. We will leave out the term 'hard margin' unless there is room for confusion.

Basically, the SVM relies on the well known optimal hyperplane classification, i.e., the separation of two classes of points by a hyperplane such that the distance of distinct points from the hyperplane, the so–called *margin*, see Fig. 1 (bottom), is maximal. SVMs utilize this linear separation method in very high dimensional feature spaces induced by reproducing kernels to obtain a nonlinear separation of original patterns, see Fig. 1 (top).

Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^d$ containing the data to be classified. We suppose that there exists an underlying unknown function t, the so–called *target function*, which maps $\mathcal{X}$ to the binary set $\{-1, 1\}$. Given a training set

$$\mathcal{A} := \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \{-1, 1\} : i = 1, \ldots, M\} \qquad (1)$$

of M associations we are interested in in the construction of a real valued function f defined on $\mathcal{X}$ such that sgn(f) is a 'good approximation' of t. If f classifies the training data correctly, then we have that $\mathrm{sgn}(f(\mathbf{x}_i)) = t(\mathbf{x}_i) = y_i$ for all $i = 1, \ldots, M$. Here

$$\mathrm{sgn}(f(\mathbf{x})) := \begin{cases} 1 & \text{if } f(\mathbf{x}) \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

We will search for f in some reproducing kernel Hilbert spaces (RKHSs) $\mathcal{H}_K$ and a regularization problem in RKHSs arises. For a given training set (1) we intend to construct a function $f \in \mathcal{H}_K$ which minimizes

$$\lambda \sum_{i=1}^{M} (1 - y_i f(\mathbf{x}_i))_+ + \frac{1}{2} ||f||^2_{\mathcal{H}_K}, \qquad (2)$$

where

$$(\tau)_+ = \begin{cases} \tau & \text{if } \tau \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

By the *Representer Theorem* ([14, 15]), the minimizer of (2) has the form

$$f(\mathbf{x}) = \sum_{j=1}^{M} c_j K(\mathbf{x}, \mathbf{x}_j). \qquad (3)$$

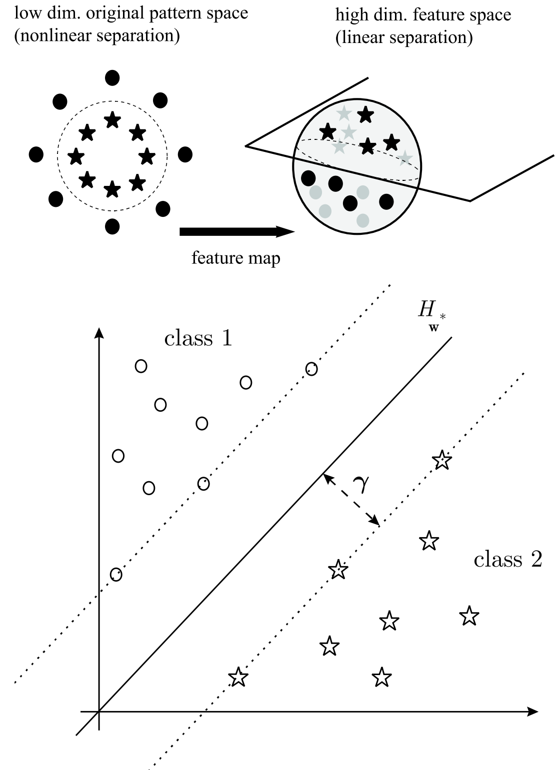

Figure 1. The principle of the SVM. Top: a linear separation by a hyperplane in a high dimensional space leads to nonlinear separation in the original space. Bottom: the optimal hyperplane which separates two classes of points with the margin γ.

Setting $\mathbf{f} := (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_M))^T$, $\mathbf{K} := (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^M$ and $\mathbf{c} := (c_1, \ldots, c_M)^T$ we obtain that

$$\mathbf{f} = \mathbf{K}\mathbf{c}.$$

Note that $\mathbf{K}$ is positive definite. Further, let $\mathbf{Y} := \mathrm{diag}(y_1, \ldots, y_M)$ and $\mathbf{u} := (u_1, \ldots, u_M)^T$. By $\mathbf{0}$ and $\mathbf{e}$ we denote the vectors with M entries 0 and 1, respectively. Now the expansion coefficients

$$\mathbf{c} = \mathbf{Y}\boldsymbol{\alpha}$$

can be obtained by the following quadratic programming (QP) problem

$$\max_{\boldsymbol{\alpha}} \left(-\frac{1}{2} \boldsymbol{\alpha}^T \mathbf{Y}\mathbf{K}\mathbf{Y}\boldsymbol{\alpha} + \mathbf{e}^T \boldsymbol{\alpha} \right) \qquad (4)$$

subject to

$$\mathbf{0} \leq \boldsymbol{\alpha} \leq \lambda \mathbf{e},$$

see [11] for details. Such QP problems provide a global solution which is a major advantage of SVMs in contrast to other learning schemes, e.g., feed forward backpropagation networks which can be trapped into local minima. Moreover, the complexity of a SVM is automatically adapted to the data, depending on the number of expansion coefficients $\mathbf{c}$ which do not vanish. In general, there are only a few parameters to adjust, see [12] and [15] for detailed discussions.

3. Adapted decompositions

Consecutive beats within each recorded EE were selected by threshold application. A reliable detection threshold was defined by 27% of the largest sample amplitude of each data segment.

Let P and Q be decomposition operators which are associated with the decimators of a two–channel normalized paraunitary analysis bank, P: lowpass, Q: highpass. By the described preprocessing of our data, the following decomposition turned out to be well suited for our feature extraction task since it discards all the irrelevant information, e.g., noise, and retains discriminating signal features (see [8]):

$$(\boldsymbol{y}_1, \dots, \boldsymbol{y}_6) = (Q^2P^2\mathbf{x}, PQP^2\mathbf{x}, Q^2P^3\mathbf{x}, PQP^3\mathbf{x}, QP^4\mathbf{x}, P^5\mathbf{x}). \quad (5)$$

All paraunitary filter banks can be parameterized by the lattice structure [16]. 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 [17, 8] for details. We use $(\boldsymbol{\vartheta})$ to denote this dependence. Let a set of M endocardial waveforms $\mathbf{x}_i \in \mathcal{X} \subset \mathbb{R}^d$ be given which belong to two distinct classes with corresponding labels $y_i \in \{-1, 1\}$ $(i = 1, \dots, M)$. By M_+ and M_- we denote the sets of indices $i \in \{1, \dots, M\}$ with $y_i = 1$ and $y_i = -1$, respectively. We label antegrade atrial activation by 1 and retrograde atrial activations by -1.

For a fixed EE activation $\mathbf{x}$ we define the function $\boldsymbol{\xi}_{\mathbf{x}} : \mathcal{P} \to \mathbb{R}^6$

$$\boldsymbol{\xi}_{\mathbf{x}}(\boldsymbol{\vartheta}) = (\xi_1(\boldsymbol{\vartheta}), \dots, \xi_6(\boldsymbol{\vartheta})) = (||\boldsymbol{y}_1(\boldsymbol{\vartheta})||_{\ell^1}, \dots, ||\boldsymbol{y}_6(\boldsymbol{\vartheta})||_{\ell^1})$$

and set $\boldsymbol{\xi}_i(\boldsymbol{\vartheta}) = \boldsymbol{\xi}_{\mathbf{x}_i}(\boldsymbol{\vartheta})$ $(i = 1, \dots, M)$. 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

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

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 [12, 13].

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

$$\hat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}\in\mathcal{P}} \left\{ \min_{i \in M_+, j \in M_-} \left|\left|\Phi(\boldsymbol{\xi}_i(\boldsymbol{\vartheta})) - \Phi(\boldsymbol{\xi}_j(\boldsymbol{\vartheta}))\right|\right|^2_{\mathcal{F}_K} \right\}, \quad (6)$$

where $\Phi(\cdot)$ denotes the SVM feature map. After some manipulations and simplifications, see [11], it remains the following problem

$$\hat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}\in\mathcal{P}} \left\{ \left|\left|\boldsymbol{\xi}_+(\boldsymbol{\vartheta}) - \boldsymbol{\xi}_-(\boldsymbol{\vartheta})\right|\right|_2 \right\},$$

where $\boldsymbol{\xi}_+(\boldsymbol{\vartheta})$ and $\boldsymbol{\xi}_-(\boldsymbol{\vartheta})$ is the center of the antegrade and retrograde atrial activations, respectively. See [11] how to solve problems of this type.

4. Results

In our experiments, we apply the Gaussian RBF $k(t) = e^{-\frac{t^2}{2s^2}}$ as kernel of the SVM. The whole hybrid EE classification scheme was applied to seven consecutive beats of each data segment which were not used for the adaptation and the RKHS regularization problem, respectively. Three–hundred and one data segments (AA: 156; RA: 145) were classified by the scheme and twenty eight data segments were discarded by a *X out of Y* detector which serves as additional regularization scheme to exclude ectopic beats and artifacts. The results for the classified data segments are shown in Fig. 2 for various scalings of the SVM. We see that the error rate equals zero for most scalings.

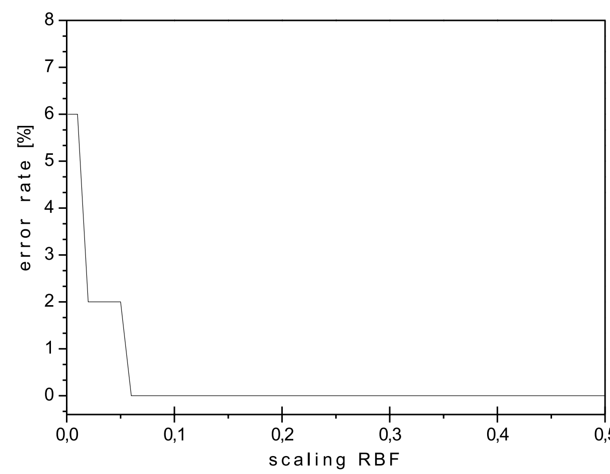

Figure 2. The results of the hybrid wavelet packet–support vector classifier.

When applying the well accepted correlation waveform analysis (CWA) [4] with best fit alignment to the same data, the results in Fig. 3 are obtained. It is noticeable that we have strong overlaps for the antegrade and retrograde atrial activations such that no proper X out of Y criterion can be defined.

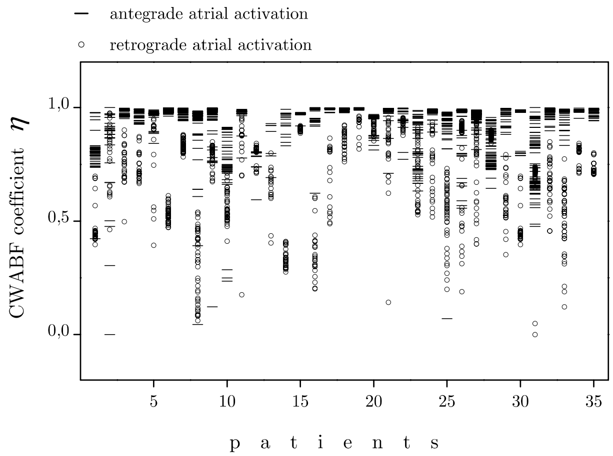

Figure 3. The results for the correlation waveform analysis with best fit alignment.

5. Conclusion

We have proposed a wavelet packet–support vector classification of atrial activation patterns which may be suitable to discriminate between sinus tachycardia and ventricular tachycardia with 1:1 retrograde conduction. We have shown that this method is effective and outperforms a well known detection scheme of endocardial signal analysis, namely the CWA with best fit alignment. The possible use of this scheme in dual-chamber ICDs may increase the VT detection rate by these 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] Schaumann A, von zur Mühlen F, Gonska BD, Kreuzer H. Enhanced detection criteria in implantable cardioverter–defibrillators to avoid inappropriate therapy. Am J Cardiol 1996;78:42–50.

[3] Nair M, Saoudi N, Kroiss D, Letac B. Automatic arrhythmia identification using analysis of the atrioventricular association. Application to a new generation of implantable defibrillators. Participating centers of the Automatic Recognition of Arrhythmia Study Group. Circulation 1997; 95:967–973.

[4] 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.

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

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

[7] Strauss DJ, Steidl G, Delb W. Feature extraction by shape–adapted local discriminant bases. Signal Processing 2002;To appear.

[8] Strauss DJ, 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.

[9] Jung J, Strauss DJ, 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 IEEE Computers in Cardiology (26); .

[10] Strauss DJ, Steidl G, Jung J. Arrhythmia detection using signal–adapted wavelet preprocessing for support vector machines. In Murray A, Swiryn S (eds.), IEEE Computers in Cardiology (28). 2001; 497–501.

[11] Strauss DJ, Steidl G. Hybrid wavelet–support vector classification of waveforms. J of Computational and Applied Mathematics ;To appear.

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

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

[14] Kimeldorf GS, Wahba G. Some results on tchebycheffian spline functions. J Anal Applic 1971;33:82–95.

[15] Wahba G. Support vector machines, reproducing kernel hilbert spaces and the randomized GACV. In Schölkopf B, Burges C, Smola AJ (eds.), Advanced in Kernel Methods – Support Vector Learning. Cambridge, MA, 1999; 293–306.

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

[17] Strauss DJ, 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.

Address for correspondence:

Dr. Daniel J. Strauss
Head of Modelling and Scientific Computing
Institute for New Materials
Saarbrücken, Germany
tel./fax: ++49-681-9300155/242
strauss@keynumerics.com