An arrhythmia classification method based on selected features of heart rate variability signal and support vector machine-based classifier

F. Yaghouby, A. Ayatollahi

Department of Electrical Engineering, Iran University of Science & Technology, Tehran, Iran
Farid.yaghouby@gmail.com

Abstract— **In this paper we present an arrhythmia classification method using Heart Rate Variability (HRV) signal features and Support Vector Machine (SVM) Classifier. Eight linear and nonlinear features are extracted from the HRV signals and a subset of these features is selected using the Improved Forward Floating Selection (IFFS) method to train the classifier. BSVM is a classification algorithm based on SVM which is able to solve the multi-class classification problems. here, five types of the most life threatening cardiac arrhythmias including normal sinus rhythm, atrial fibrillation, ventricular fibrillation, ventricular bigeminy and sick sinus syndrome can be discriminated by BSVM and selected features with the average accuracy of 99.78%.**

Keywords—**Heart rate variability signal, Arrhythmia classification, Support vector machine, Forward floating feature selection**

I. INTRODUCTION

Nowadays the cardiac arrhythmias are the most famous causes of mortality. Hence, several techniques have been proposed to identify and detect the different types of arrhythmia. These techniques usually extract desired features from ECG or HRV arrhythmic signals to classify them. Since ECG signal processing is time consuming and too sensitive to the amount of the noise, many researchers analyze HRV signals to detect abnormal rhythms. Some examples of these automatic arrhythmia detection and classification techniques are neural networks [1,2], wavelet transforms [3], support vector machines [4], fuzzy logic [5] and the rule-based algorithms [6].

The proposed algorithm in this paper presents an HRV-based arrhythmia classification method which can detect and classify five types of the most famous abnormal cardiac rhythms. These arrhythmias are namely the Normal Sinus Rhythm (NSR), the Atrial Fibrillation (AF), the Ventricular Fibrillation (VF), the Ventricular Bigeminy (B) and the Sick Sinus Syndrome (SSS). This technique is based on the IFFS feature selection method and SVM-based classifier. In the first step, IFFS selects the optimal subset of features from the 8 original features and then SVM separates the arrhythmia classes in the selected feature space. IFFS finds the best subset of features by evaluating a criterion function [7]. It not only reduces the cost of feature extraction methods like Principle Component Analysis (PCA), but also improves the classification accuracy.

SVM, first proposed by Vapnik in 1998 [8], has been used as a powerful tool for classification problems. Here, we propose the multi-class BSVM formulation [9] for arrhythmia classification. As it reported in the past, we can see that SVM provides more accurate results in classification than other methods such as the neural networks.

The details of the mentioned method for arrhythmia classification using HRV features are presented in continue. In the first section, we explain the steps extracting ECG signals and preprocessing it in order to obtain HRV signals. Afterwards, a range of linear and nonlinear features are extracted from HRVs and then by using the IFFS the dimensionality of these original features are reduced to 4. Finally, in the last section the BSVM multi-class classifier is applied to the selected features to detect any types of 5 cardiac arrhythmias.

II. MATERIALS AND METHODS

A. Extracting and preprocessing the signal

The MIT-BIH arrhythmia data base is a standard reference for ECG signal processing which includes 48 ECG recordings each with a length of 30 min [10]. All signals in this database were filtered in the frequency range of 0.1-100Hz and were sampled with a sampling rate of 360Hz. We extract the ECGs relating to NSR, AF, B and SSS arrhythmias from this database. In addition to this database, we use the Creighton University Ventricular Tachyarrhythmia Database to obtain the VF signals after resampling it at a rate of 360 Hz. As the first step it is necessary to extract HRV signals from these ECG signals. For this purpose the interfering signals are eliminated using a 5-15Hz bandpass filter. Then the wave R in the filtered signals is detected by using Hamilton and Tompkins algorithm [11]. For constructing the HRV signal we first measure the time intervals between the successive waves R in each ECG signal and then plot this intervals against the time indices. The obtained HRV signals are divided into the same length segments each

O. Dössel and W.C. Schlegel (Eds.): WC 2009, IFMBE Proceedings 25/IV, pp. 1928–1931, 2009.
www.springerlink.com

containing 64 R-R intervals. Totally, we have 889 HRV segments including 341 NSR segments, 340 AF segments, 142 VF segments, 37 B segments and 24 SSS segments.

B. Extracting features and selecting optimal subset

To illustrate linear and nonlinear behavior of cardiovascular system, it is necessary to consider both linear and nonlinear features of cardiac signals. So we consider a combination of linear and nonlinear features. The linear features which are obtained from time and frequency domains are calculated based on the proposed standard in [12]. The linear features are 5 and include:

- Time domain features: This features which are extracted from the R-R interval time series directly are:
  Mean RR: The mean value of the 64 R-R intervals in each segment.
  STD RR: The standard deviation of the 64 R-R intervals in each segment
  RMSSD: The root mean square successive difference of the 64 R-R intervals in each segment
  pNN50: The number of successive difference of 64 R-R intervals that differs more than 50 ms, respectively, divided by 64.
- Frequency domain features: These features are extracted to discriminate between sympathetic and parasympathetic contests of the HRV signals. In this work we calculate Power Spectral Density (PSD) for the High Frequency (HF) band (0.15-0.4Hz) and Low Frequency (LF) band (0.04-0.15Hz) and the ratio of the LF and HF bands power (LF/HF) as the Frequency domain feature of the HRV signal.

On the other hand, HRV signal analysis by help of methods on nonlinear dynamics leads to very valuable information for physiological interpretation of the heart. So we extract these 3 nonlinear features:

LLE: The Largest Lyapunov Exponent provides useful information about the dependency of system on initial conditions and a positive lyapunov exponent confirms the existence of chaos in the system. For calculating LLE a point is selected in the reconstructed phase space of the system and all neighbor points residing within a predefined radius are determined. As the system evolves, the mean distances between the trajectory of the initial point and the trajectories of the neighbor points are calculated. Then the logarithm of these mean values plots against the time and the slope of the resulting line is considered as *LLE* [13].

D2: The Correlation Dimension is a measure of complexity of the time series and determines the minimum number of dynamic variables which can model the system. We use the algorithm presented in [14] in order to estimate this feature.

ApEn: The Approximate Entropy shows the unpredictability of the fluctuations in a time series. Large values of ApEn show high irregularity and smaller values of it indicate more regular time series. The proposed method in [15] is used to calculate *ApEn* for each segment of HRV signal.

Now, we have 8 features for each of 889 HRV segments. These all features are normalized within the range of [0, 1] initially. We then simplify the proposed method reducing the number of features. For this purpose a new feature selection method named Improved Forward Floating Selection (IFFS) is used to select an optimal subset of original features. This algorithm has a new search strategy to check whether removing any feature in the selected feature set and adding a new one at each sequential step can minimize the criterion function (misclassification rate). The results show that this method compared with other techniques selects the optimal subset of features and requires significantly less computational time.

By applying the IFFS to the original feature space, 4 features are selected from 8. These selected features are *Mean RR*, *RMSSD*, *pNN50* and *D2*.

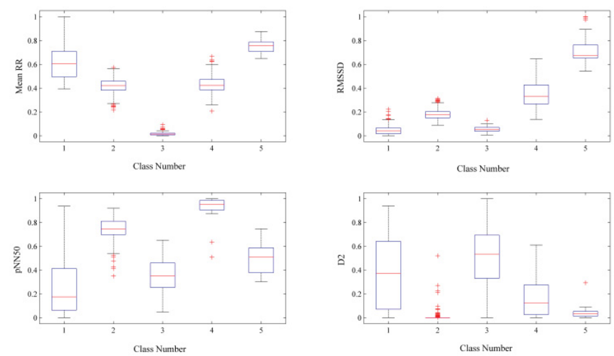

Fig. 1 Box-plots of the four selected features for different arrhythmia classes (1 = NSR, 2 = AF, 3 = VF, 4 = B, 5 = SSS).The values are normalized between 0 and 1.

The box-plots of the four selected features for different arrhythmia classes are presented in Fig. 1. As seen each of the selected features has a value in a range that differs from one class to another. In fact we can say by using these features we have better discrimination between the 5 classes.

Table 1 The Confusion Matrix on the test set. The values are average of 100 train and test procedures

Total number of train/test segments			classification				
			NSR	AF	VF	B	SSS
223.07/117.93	Database annotation	NSR	117.93	0	0	0	0
221.57/118.43		AF	0.22	117.55	0.03	0.63	0
90.01/51.99		VF	0	0.01	51.98	0	0
22.98/14.02		B	0.06	0.77	0.01	13.18	0
14.37/9.63		SSS	0	0	0	0	9.63

C. Classification based on SVM

The last step of the proposed algorithm is classification of arrhythmias using the selected features. As it mentioned we use SVM as a classifier here.SVM is a machine-learning technique which identifies the best separating hyper plane between the two classes [16]. Although SVM can separate the input data into only two classes, the multi-class classification is also possible by BSVM formulation. Suppose the training vectors are:

$$\mathcal{T}_{XY} = \{(x_1, y_1), \dots, (x_l, y_l)\} \tag{1}$$

where

$x_i \in \mathcal{X} \subseteq \mathcal{R}^n$

$y_i \in \mathcal{Y} = \{1,2,\dots,c\}$

The aim of BSVM is training the following classification rule

$$q(x) = arg\max_{y\in\mathcal{Y}} f_y(x) \tag{2}$$

where

$f_y(x) = \langle \alpha_y . k_S(x)\rangle + b_y \;, y \in \mathcal{Y}$

The parameters of the rule above are determined by solving the BSVM formulation:

$$(W^*, b^*, \xi^*) = arg\min_{w,b} \frac{1}{2}\sum_{y\in Y}(\|\omega_y\|^2 + b_y{}^2) + C\sum_{i\in I}\sum_{y\in Y\backslash\{y\}}(\xi_i)^p \tag{3}$$

subject to

$\langle \omega_{y_i} . x_i\rangle + b_{y_i} - (\langle \omega_y . x_i\rangle + b_y) \geq 1 - \xi_i^y \;, i \in I, \in Y\backslash\{y_i\}$

$\xi_i^y \geq 0 \;, i \in I, y \in Y\backslash\{y_i\}$

Considering p=2, we use the Mitchell-Demyanov-Malozemov algorithm [17] to solve (3). Furthermore, we select the Radial Basis Function (RBF) as the kernel in (2). So, we have two free parameters that must be assigned correctly. The first is the width of RBF kernel and the second is the Regularization parameter in (3). We select σ=0.2 and C=10 empirically.

III. RESULTS

Finally, the all 889 segments of different arrhythmias are randomly divided to train and test sets in an approximate ratio of 2/3 and 1/3. After training the SVM classifier we use test set to evaluate the classification performance. This procedure is repeated 100 times and each time we divide whole data into train and test sets and carry out classification using these sets. The average Confusion Matrix obtained from 100 different test sets are presented in Table 1. As seen for the NSR and SSS we have no misclassification to other classes (0%). for the AF in average only 0.88 segments are also misclassified (0.007%), for the VF in average 0.01 segments are misclassified (≅0%) and for B in average 0.84 segments are misclassified (0.06%).

In continue the four famous measures sensitivity, specificity, positive predictivity and accuracy are derived from the proposed algorithm. Furthermore, to compare the efficiency of the proposed method, these parameters are calculated for the SVM classifier which is trained using the 8 original features too. Table 2 shows the average values of these parameters for both mentioned algorithm. As we can see the presented method can discriminate the NSR with an average accuracy of 99.91%, the AF with 99.47%, the VF with 99.98%, the B with 99.53% and SSS with 100%. These results demonstrate the effectiveness of this method in the classification of cardiac arrhythmias. As a comparative study we can see the results of the classification using original features too. Table 2 shows that the classification using the 4 selected features has better results than the classification using the 8 features. The average values of the performance parameters shows that when we use the selected features instead of the original features for classification we have an increment about 2% in the sensitivity, 0.2% in the specificity, 0.5% in positive predictivity and 0.2% in the accuracy. The use of selected features on the other hand decreases the SVM training time significantly. So the proposed classification algorithm based on IFFS and SVM classifier, not only decrease the processing time but also makes a noticeable increase in accuracy of classification.

Table 2 Performance analysis of the BSVM classifier on the original features and the selected features in terms of the average values of the four commonly used measures in %.(number inside the parenthesis are the standard deviations)

	Classification using 4 selected features				Classification using 8 original features			
Arrhythmia classes	Sensitivity (%)	Specificity (%)	Positive Predictivity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive Predictivity (%)	Accuracy (%)
NSR	100	99.85	99.76	99.91	100	99.39	99.03	99.62
AF	99.26	99.57	99.34	99.47	99.14	99.2	98.68	99.25
VF	99.98	99.98	99.92	99.98	99.42	99.95	99.72	99.83
B	94	99.79	95.44	99.53	89.28	99.68	94.32	99.38
SSS	100	100	100	100	93.86	100	100	99.8
Average	98.65(2.61)	99.84(0.16)	98.89(1.95)	99.78(0.26)	96.38(4.68)	99.68(0.36)	98.35(1.02)	99.58(0.25)

IV. Conclusions

In this paper an effective HRV-based arrhythmia classification method has been presented. We first extract 8 features from HRV segments and then in order to reduce the learning time and also to improve the efficiency of the classifier, 4 optimal features are selected from 8 original features using the IFFS algorithm. Then a SVM-based multi-class classifier method named BSVM is used to classify the 5 types of arrhythmias. Comparing the results that have been shown in Table 2 we find that the proposed technique outperforms the same classifier which is applied to the original features producing the classification accuracy of 99.91%, 99.47%, 99.98%, 99.53% and 100% for the arrhythmia classes of NSR, AF, VF, B and SSS respectively.

References

1. Acharya RU, Kumar A, Bhat PS, Lim CM, Iyengar SS, KannathalN, et al.(2004) Classification of cardiac abnormalities using heart rate signals. Med Biol Eng Comput; 42(3):288—93.
2. Tsipouras MG, Fotiadis DI. (2004) Automatic arrhythmia detection based on time and time—frequency analysis of heart rate variability. Comp Meth Prog Biomed; 74(2):95—108.
3. Khadra L, Al-Fahoum AS, Al-Nashash H.(1997) Detection of life threatening cardiac arrhythmias using wavelet transformation. Med Biol Eng Comput; 35(6):626—32.
4. Song MH, Lee J, Cho SP, Lee KJ, Yoo SK. (2005) Support vector machine based arrhythmia classification using reduced features. Int J Control Automat Syst; 3(4):571—9.
5. Tsipouras MG, Goletsis Y, Fotiadis DI. (2004) A method for arrhythmic episode classification in ECGs using fuzzy logic and Markov models. In: Murray A, editor. Proceedings of the computers in cardiology. p. 361—4.
6. Tsipouras MG, Fotiadis DI, Sideris D. (2005) an arrhythmia classification system based on the RR-interval signal. Artif Intell Med; 33(3):237—50.
7. P. Pudil, J. Novovicova, and J. Kittler.(2008) Improved forward floating selection algorithm for feature subset selection, International Conference on Wavelet Analysis and Pattern Recognition., Hong Kong, pp 793-798
8. Vapnik V. (1998) Statistical learning theory. Berlin: Springer.
9. C.W. Hsu and C.J. Lin.(2002) A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks, 13(2)
10. Mark RG, Moody GB. (1997) MIT-BIH Arrhythmia Database [Online]. Available from: http://ecg.mit.edu/dbinfo.html
11. Hamilton PS, Tompkins WJ. (1986) Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. IEEE Trans Biomed Eng;33(12):1157—65.
12. Task force of the European society of cardiology and the North American society of pacing and electrophysiology. (1996) Heart rate variability—standards of measurements, physiological interpretation, and clinical use. Eur Heart J; 17(3):354—81.
13. Wolf A, Swift JB, Swinney HL, Vastano JA. (1985) Determining Lyapunov exponents from a time series. Physica D; 16:285—317.
14. B. Henry, N. Lovell, and F. Camacho. (2001) nonlinear dynamics time series analysis. In M. Akay, editor, Nonlinear Biomedical Signal Processing: Dynamic Analysis and Modeling, IEEE Press, 11(2):1–39.
15. Pincus SM, Goldberger AL. (1994) Physiological time series analysis: what does regularity quantify? Am J Physiol; 266: 1643—56.
16. Mercier G, Lennon M. (2003) Support vector machines for hyperspectral image classification with spectral-based kernels. In: Proceedings of the international geoscience and remote sensing symposium; 288—90.
17. S.S. Keerthi, S.K. Shevade, C. Bhattacharya, and K.R.K. Murthy. (2000) A fast iterative nearest point algorithm for support vector machine classifier design. IEEE Transactions on Neural Networks, 11(1):124–136