Higher order statistics for automated classification of ECG beats

Ataollah Ebrahimzadeh
Faculty of Electrical and Computer Engineering
Babol University of Technology
Babol, IRAN
ataebrahim@yahoo.com

Ali Khazaee
Faculty of Electrical and Computer Engineering
Babol University of Technology
Babol, IRAN
Khazaeeali@yahoo.com

***Abstract*— This work describes a Support Vector Machine (SVM) method used to analyze ECG signals for diagnosing cardiac arrhythmias effectively. The proposed method can accurately classify and differentiate normal (Normal) and abnormal heartbeats. Abnormal heartbeats include left bundle branch block (LBBB), right bundle branch block (RBBB), atrial premature contractions (APC) and premature ventricular contractions (PVC). This paper proposes a two stage, feature extraction and classification method for the detection of ECG beat types. Feature extraction module extracts higher order statistics of ECG signals in conjunction with three timing interval features. Then a number of support vector machine (SVM) classifiers with different value parameters are designed. These parameters are: Gaussian radial basis function (GRBF) kernel parameter and C penalty parameter of SVM classifier. We compared the classification ability of five different classes of ECG signals that were achieved over eight files from the MIT/BIH arrhythmia database.**

Keywords- ECG beat classification, SVM, Higher order statistics, Cumulants

I. INTRODUCTION

The analysis of ECG has been widely used for diagnosing many cardiac diseases. The development of accurate and quick methods for automatic ECG classification is vital for clinical diagnosis of heart diseases.

In the literature, several methods have been proposed for the automatic classification of ECG signals. Among the most recently published work are those presented in [1-4].

In this paper, we have proposed an automated method for ECG heartbeats classification into five different classes. For feature extraction module, we have used a suitable set of features that consists in both statistical and temporal features, to include both of the shaping and timing information of signal. Then, we investigated the different support vector machine (SVM) classifiers and varied the parameters values for those SVMs. Then we have prepared some experiments to measure their performances and compare them.

The paper is organized as follows. Section 2 explains the feature extraction. Section 3 presents the classifier. Section 4, describes the database and performance metrics. Section 5 shows some simulation results. Section 6 discusses the results and finally Section 7 concludes the paper.

II. FEATURE EXTRACTION

A. Higher Order Cumulants

Probability distribution moments are the generalization of concept of the expected value, and can be used to define the characteristics of a probability density function [5]. The auto-moment of the random variable may be defined as follows:

$$M_{pq} = E[s^{p-q}(s^*)^q] \qquad (1)$$

where p called moment order and s^* stands for complex conjugation. Now, consider a zero-mean discrete based-band signal sequence of the form $s_k = a_k + jb_k$. Using the definition of the auto-moment, the expressions for different order may be easily derived. For example:

$$\begin{aligned} M_{83} &= E[s^5(s^*)^3] = E[(a+jb)^5(a-jb)^3] \Rightarrow \\ M_{83} &= E[(a^5 + j5a^4b + j^2 10a^3b^2 + j^3 10a^2b^3 + j^4 5ab^4 \\ &\quad + j^5 b^5)(a^3 - j3a^2b + j^2 3ab^2 - j^3 b^3)] \Rightarrow \qquad (2) \\ M_{83} &= E[a^8 + j2a^7b - j^2 2a^6b^2 - j^3 6a^5b^3 + j^5 60a^3b^5 \\ &\quad + j^6 2a^2b^6 - j^7 2ab^7 - j^8 b^8] \Rightarrow \\ M_{83} &= E[a^8 + 2a^6b^2 - 2a^2b^6 - b^8] \end{aligned}$$

The symbolism for p^{th} order of cumulant is:

$$C_{pq} = Cum[\underbrace{s,\ldots,s}_{(p-q)terms}, \underbrace{s^*,\ldots,s^*}_{(q)terms}] \qquad (3)$$

It can be computed the relation between moments and cumulants. e.g.:

$$C_{61} = M_{61} - 10M_{20}M_{41} - 5M_{21}M_{40} + 30M_{20}{}^2M_{21} \qquad (4)$$

These values are computed under the constraint of unit variance and noise free. We have considered the second, third and fourth order of the cumulant at six points distributed evenly within the signal length, as the features. Therefore the total number of the statistical features is equal to 18. We have computed the all of these features for the considered ECG signals.

978-1-4244-8165-1/11/$26.00 ©2011 IEEE

B. Temporal Features

Feature extraction plays an important role in any classification task. In this work, based on extensive research, we have used a balanced combination of statistical and timing features. For a 91-element vector representation of the QRS complex, the cumulants corresponding to the time lags of 15, 30, 45, 60, 75 and 90 have been chosen. Additionally, we have added three temporal features. In this way each beat has been represented by a 21-element feature vector, with the first 18 elements corresponding to the higher-order statistics of QRS complex (the second-, third-, and fourth-order cumulants, each represented by six values) and the last three are the temporal features of the actual QRS signal. The application of the cumulant characterization of QRS complexes reduces the relative spread of the ECG characteristics belonging to the same type of heart rhythm and in this way makes the classification relatively easier.

As stated, in addition to the higher order statistical features, we extracted three local timing features. They are an R-R time interval ratio (IR) and two R-R time intervals. The IR ratio feature reflects the deviation from a constant beat rate and was defined as:

$$IR_i = \frac{T_i - T_{i-1}}{T_{i+1} - T_i} \tag{5}$$

where T_i represents the time at which the R-wave for beat i occurs. The local RR-interval ratio provides a convenient differentiator between normal beats $(IR_i \sim 1)$ and PVC beats $(IR_i < 1)$, and is normalized by definition $(IR_i = 1)$ at constant rate. Two other timing features are the next and previous R-R time intervals for each heartbeat. Each feature vector is then normalized to have a zero-mean and a unity variance. The length of the final feature vectors is 21.

III. CLASSIFER

SVM performs classification tasks by constructing optimal separating hyper-planes (OSH). OSH maximizes the margin between the two nearest data points belonging to two separate classes. So the following inequality is valid for all input data:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \text{for all } \mathbf{x}_i \quad i = 1,2,...,l \tag{6}$$

This is a convex quadratic programming (QP) problem and Lagrange multipliers are used to solve it. Those training points, for which the equality in (6) holds, are called support vectors (SV). After solving problem, the optimal bias is given by:

$$b^* = y_i - \mathbf{w}^{*T}\mathbf{x}_i \tag{7}$$

for any support vector $\mathbf{x}_i$. The optimal decision function (ODF) is then given by:

$$f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{l} y_i \alpha_i^* \mathbf{x}^T \mathbf{x}_i + b^*) \tag{8}$$

where α_i^* 's are optimal Lagrange multipliers.

For input data with a high noise level, SVM uses soft margins can be expressed as follows with the introduction of the non-negative slack variables $\xi_i, i = 1,...,l$

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad for \quad i = 1,2,...,l \tag{9}$$

To obtain the OSH, it should be minimizing the $\Phi = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i^k$ subject to (9), where C is the penalty parameter, which controls the tradeoff between the complexity of the decision function and the number of training examples, misclassified.

In the nonlinearly separable cases, the SVM map the training points, nonlinearly, to a high dimensional feature space using kernel function $K(\vec{x}_i, \vec{x}_j)$, where linear separation may be possible. The kernel functions of SVMs are Linear, Polynomial, Sigmoid and Gaussian radial basis function (GRBF) with the following equation:

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \tag{10}$$

where σ is the parameter of the GRBF kernel function. After a kernel function is selected, the QP problem will change and After training, the following, the decision function, becomes:

$$f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{l} y_i \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b^*) \tag{11}$$

The performance of SVM can be controlled through the term C and the kernel parameter which are called hyper-parameters. These parameters influence on the number of the support vectors and the maximization margin of the SVM [6].

IV. DATABASE AND PERFORMANCE METRICS

The MIT–BIH arrhythmia database [7] was used as the data source in this study. The database contains 48 recordings. The sampling frequency is 360 Hz, the data are bandpass filtered at 0.1–100 Hz and the resolution is 200 samples per mV. The database is indexed both in timing information and beat classification. For more details about MIT–BIH Arrhythmia database see [8].We used a total of 8 records (see Table 2 for list) from the database. We used the database index files from database to locate beats in ECG signals.

Various approaches are adopted to evaluate the classifier performances in literature. In this study, we have considered four statistical indices: Accuracy (Acc), Sensitivity (Se), Specificity (Sp), and Positive Predictivity (Pp), which are defined in the following equations (12-15), respectively.

The most crucial metric for determining overall system performance is usually accuracy. We defined the overall accuracy of the classifier for each file as follows:

$$Acc = \frac{N_T - N_E}{N_T} \times 100 \tag{12}$$

In this equation, Acc is the accuracy, and the variables, N_E and N_T , represent the total number of classification errors and beats in the file, respectively. Sensitivity, Se, the ratio of the number of correctly detected events, TP (true positives), to the total number of events is given by:

$$Se = \frac{TP}{TP + FN} \times 100 \tag{13}$$

where FN (false negatives) is the number of missed events. The specificity, Sp, the ratio of the number of correctly

$$Sp = \frac{TN}{TN + FP} \times 100 \tag{14}$$

where FP (false positives) is the number of falsely detected events. Positive predictivity, Pp, is the ratio of the number of correctly detected events, TP, to the total number of events detected by the analyzer and is given by:

$$Pp = \frac{TP}{TP + FP} \times 100 \tag{15}$$

V. Simulation Results

We randomly selected 100 beats from each class, and used these 500 beats for training of classifiers. Total number of beats in our database was 18,290. Clearly, the number of our training set is less than 3% of all beats. Thus, our study is well generalized. After performing classification with various kernels of SVM, we empirically found that, the SVM with Gaussian kernel has better results than the other .

kernels such as linear and polynomial. In table I, we investigated the accuracy of Gaussian SVM classifier with multiple values of C and sigma parameters for the classification of ECG signals into five classes. Then, we used the best C and sigma for the next experiments. Table II shows the file-by-file comprehensive results for a sample SVM classifier (The SVM with C and sigma parameters of 470000 and 0.01, respectively). Table III, proposes the confusion matrix for classification of ECG signals using the sample SVM that produces the best overall accuracy of 95.70%.

VI. Discussion

In table I, it can be seen that the best performance for GRBF support vector machines, is achieved by the C and σ parameters with values of 470000 and 0.01, respectively (Best performance is bolded in Table I). We attain a high overall accuracy of 95.70% using support vector machine.

Two innovations of SVMs are responsible for the success of these methods, namely, the ability to find a hyper plane that divides samples in to two classes with the widest margin between them, and the extension of this concept to a higher dimensional setting using kernel function to represent a similarity measure on that setting. This makes SVMs a practical and and effective solution for many pattern recognition and classification problems in bioinformatics.

In table II the classification results of beats for each record can be seen. Table III is the confusion matrix of the all data classification results. As it can be seen most wrong classified normal beats are those classified as APC (322 beats). If you see APC beats in table III, you find that most misclassified APC beats, 51 samples, are classified as normal. The mainspring is that, Normal and APC patterns are morphologically similar to each other. This evident the importance of three temporal features that we have used in feature set. These features improve the discriminating ability of the classifier, especially in discriminating morphologically similar heartbeat patterns (i.e. Normal and APC beats). About PVC beats this is inversely. Number of PVC beats that classified as normal beats and number of normal beats that classified as PVC beats are 32 and 28, respectively. Small value of misclassified beats between normal and PVC beats evident the morphological dissimilarities between Normal and PVC beats.

VII. Conclusion

We have proposed a number of efficient methods for accurate classification of ECG beat for a relatively large set of data. These methods include two modules: The feature extraction module and classifier module. In the feature extraction module we have extracted the 18 higher order statistics (cumulants) and three R-R interval based temporal as the effective features for differentiating normal beats and four abnormal beats. Then a number of support vector machine (SVM) classifiers with different values of C (penalty factor) and sigma parameters are designed and compared their ability for classification of five different classes of ECG signals. A classification overall accuracy of 95.70% were achieved over eight files from the MIT/BIH arrhythmia database.

References

[1] M. Faezipour, et al., "A Patient-Adaptive Profiling Scheme for ECG Beat Classification," *IEEE Trans. Inf. Technol. Biomed.,*vol. 14, no. 5, pp. 1153-1165, Sept. 2010.

[2] E. D. Ubeyli, "Eigenvector methods for automated detection of electrocardiographic changes in partial epileptic patients," *IEEE Trans. Inf. Technol. Biomed.,* vol. 13, no. 4, pp. 478–485, Jul. 2009.

[3] A.Khazaee, A. Ebrahimzadeh, "Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features," *Biomedical Signal Processing and Control* 5, pp. 252–263, 2010.

[4] A. Ebrahim Zadeh, A. Khazaee, and V. Ranaee, "Classification of the electrocardiogram signals using supervised classifiers and efficient features," *computer methods and programs in biomedicine* 99, pp. 179–194, 2010.

[5] McCullagh, P. (1987). Tensor methods in statistics. Chapman & Hall.

[6] C. Burges A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, vol. 2, pp 121-167,1998.

[7] R.G. Mark, G.B. Moody, MIT–BIH Arrhythmia Database 1997 [Online]. Available: http://ecg.mit.edu/dbinfo.html.

[8] G.B. Moody, R.G. Mark, The impact of the MIT/BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.* 20 (May–June (3)) (2001) 45–50.

TABLE I. COMPARATIVE RESULTS FOR STUDYING THE ACCURACIES OF SVMS WITH DIFFERENT VALUES OF C AND SIGMA PARAMETERS

C \ σ	0.01	0.02	0.03	0.04	0.045	0.5	1	2	3	4	5	10
0.1	55.45	55.45	55.47	55.21	55.00	59.32	59.65	59.99	60.3351	60.98	61.42	60.84
1	54.73	55.50	57.34	58.32	58.42	61.48	62.42	70.64	71.1426	70.04	70.38	75.39
10	59.53	60.56	61.21	61.33	61.51	74.97	75.57	77.48	84.6154	85.23	85.75	86.32
100	63.68	73.12	73.90	74.74	74.76	87.39	89.07	90.08	90.4915	90.13	89.32	89.04
1000	77.04	79.07	79.89	86.52	87.91	91.26	92.08	92.03	91.3799	91.23	90.76	89.24
10000	88.33	90.22	91.14	92.25	92.63	93.37	92.38	90.71	90.2497	90.38	90.35	89.21
100000	94.51	94.89	94.73	94.77	94.79	92.81	91.56	90.47	90.2497	90.38	90.35	89.21
150000	95.11	94.93	94.91	94.87	94.70	92.81	91.56	90.47	90.2497	90.38	90.35	89.21
200000	95.30	94.95	95.03	94.73	94.41	92.81	91.56	90.47	90.2497	90.38	90.35	89.21
470000	**95.70**	95.07	94.38	94.21	94.26	92.81	91.56	90.47	90.2497	90.38	90.35	89.21
500000	95.68	95.07	94.32	94.19	94.21	92.81	91.56	90.47	90.2497	90.38	90.35	89.21
800000	95.38	94.79	94.05	93.75	93.98	92.81	91.56	90.47	90.2497	90.38	90.35	89.21
1000000	95.29	94.57	93.78	93.69	93.86	92.81	91.56	90.47	90.2497	90.38	90.35	89.21

TABLE II. DETAILED RESULTS FOR FILE-BY-FILE INCLUDING BOTH TEST AND TRAIN DATASET

Record	**Accuracy**	**Normal**			**LBBB**			**RBBB**			**APC**			**PVC**		
		Se	***Sp***	***Pp***	***Se***	***Sp***	***Pp***	***Se***	***Sp***	***Pp***	***Se***	***Sp***	***Pp***	***Se***	***Sp***	***Pp***
118	98.55	-	99.47	0	-	99.91	0	99.31	90.17	99.49	84.38	99.82	95.29	81.25	99.82	76.47
124	97.53	-	99.87	0	-	98.04	0	99.48	89.80	99.67	0	100	-	38.30	99.94	94.74
207	97.83	-	99.77	0	99.73	91.22	98.24	65.88	99.94	98.25	98.11	99.94	99.05	97.14	99.64	94.44
208	97.98	96.91	99.80	99.87	-	99.81	0	-	99.22	0	-	99.07	0	99.70	99.94	99.90
209	98.17	98.51	96.09	99.42	-	100	-	-	100	-	95.82	98.59	90.84	100	99.90	25
214	96.59	-	96.99	0	96.35	100	100	-	99.82	0	-	99.91	0	98.44	99.85	98.82
222	84.66	85.10	86.06	98.37	-	98.90	0	-	99.82	0	80.29	87.48	39.29	-	98.59	0
223	94.98	98.27	92.84	98.08	-	97.47	0	-	99.53	0	81.94	99.68	88.06	82.88	99.76	98.74

TABLE III. CONFUSION MATRIX OF EXPERIMENT 4

Confusion Matrix	**Normal**	**LBBB**	**RBBB**	**APC**	**PVC**
Normal	7762	46	33	322	28
LBBB	67	3281	4	0	6
RBBB	16	33	3626	2	1
APC	51	2	11	683	20
PVC	32	72	9	10	1667