Autonomous Detection of Heartbeats and Categorizing them by using Support Vector Machines

Hassan Yazdanian
Department of Biomedical Engineering
University of Isfahan
Isfahan, Iran
Hassanyazdanian@gmail.com

Ashkan Nomani
Department of Biomedical Engineering
University of Isfahan
Isfahan, Iran
ashkan.nomani@gmail.com

Mohammad Reza Yazdchi
Department of Biomedical Engineering
University of Isfahan
Isfahan, Iran
yazdchi@eng.ui.ac.ir

Abstract— **In this paper a new method for categorizing 5 special types of heartbeats has been developed by use of time and apparent properties of the Wavelet Transform of the ECG signal. By using the method in this paper first each heart beat identified autonomously and important points and segments of it, were derived .Then expected features for categorizing the heartbeats are extracted. Finally we categorized the arrhythmias by using the Support Vector Machines. In order to train the SVM and for analyzing its accuracy; arrhythmic signals of MIT-BIH dataset have been used. The results which have been achieved by this method also contain 96.67 percent of accuracy for categorizing five different heartbeats including Normal (N) Left Bundle Branch Block(LBBB), Right Bundle Branch Block(LBBB), Premature Ventricular Contraction (PVC) and Atrial Premature Contraction (APC).**

The advantage of using this method compared to the other ones is that we could achieve the expected precision by using less training attributes respect to the other methods.

Keywords: arrhythmia; categorizing; ECG; segmentation; Support Vector Machine (SVM)

I. INTRODUCTION

ECG signal widely been used as one the most important tools for analyzing the health of heartbeats for different types of heart disease. Autonomous categorization of heartbeats by use of ECG signals is one the most recent fields of research which has been proposed during the latest years [1].

Arrhythmia is defined as total result of perturbation in rhythm, order and place of the origin of each beat or their path of signal conduction [2]. Generally we can divide arrhythmias into two major classes. The first class can be assumed as Ventricular Fibrillation and Tachycardia and which encompass absolute risk factors for patient's life and needed to be cured urgently.

Extreme research has been done for recognizing these types of arrhythmias and many parts of detectors have been designed for classifying them with high rate of accuracy and precision. The second class contains types of arrhythmia that in first look they do not appear to be a major risk factor for patient's life but preventing further problems to happen they need to be cured [3].

These types of arrhythmias seem to be detected scarcely and they need to be recorded by a long record time, using a Holter monitor. Analyzing this massive bank of data consumes a long duration of time. Thus automatic processing of signals can be assumed as an appreciated help for physicists. In this paper a special method for detection and classification of these types of arrhythmias is used.

II. LITERATURE REVIEW

Different methods has been proposed for classification of different types of arrhythmias based on ECG signal processing. For example Chazal et al. [3] have developed one special method for classification of heartbeats using morphology and heartbeat Interval features. They considered different combinations of ECG signal features. Then a linear classifier was trained and a dataset comprising large amounts of attributes was used and by classification they have achieved 84.5% of precision.

In [4] a method for autonomous classification of heartbeats using Wavelet Neural Networks has been proposed. In this method by utilizing the algorithm given in reference [4], first the location of each R peak of the ECG signal is detected , then 5 sets of different features are extracted which uses the algorithm given in [4]. Finally Wavelet Neural Networks was used on a dataset with large extent of attributes and a precision of %98.78 was achieved.

The authors of [5] have autonomously classified ECG signals to 3 categorizes called normal , PVC (Premature Ventricular Contraction) and other contractions. In that paper first , noise of ECG signal has been removed. Then by using the algorithm which has been proposed by [6] important points of ECG signal have been detected and used for feature extraction. Finally the performance of different types of neural networks such as MLP (Multi Layer Perceptron), RBF(Radial Basis Function) ,PNN(Probabilistic Neural Network) and SVM (Support Vector Machines) have been examined for performing classification on these attributes which resulted in a precision of an order of %97.14 .

Reference [7] first detected the ST segment , then used the slope of ST segment and 10 time domain features and apparent features of ECG signal with 191 features of wavelet coefficient feature of each heartbeat for classification. In the aforementioned paper by utilizing support vector machines different types of arrhythmia were classified to 6 types and a precision of %96.36 was reached.

By reviewing literature we can understand that different levels exist in the design procedure of one type of special heartbeat classification system, and if they are designed correctly they can help in developing process of an efficient, highly accurate classifier. Some of these steps can be summarized as: 1-Preprocessing of the signal. 2- Segmentation and derivation of important points of ECG signal. 3- Selection and extraction of appropriate features by the usage of appropriate feature extraction techniques. 4- and then selecting appropriate and fair classifier.

III. PREPROCESSING

Noise reduction and baseline drift is one the most important problems that someone may face in the field of ECG signal processing. Electromyogram noise and non-appropriate skin-electrode contact and 50HZ electrical noise on ECG signal are the most important problems of ECG signal analysis [8]. Different methods such as linear and nonlinear filters, time-frequency analysis, Nero-fuzzy approaches and neural networks and was used for noise removal of the ECG signal .In this paper we use the combination of moving average filter, Butterworth filter and DWT (Discrete Wavelet Transform) in order to remove noise and omitting of baseline drift from the signal. In the first step one moving average filter by the length of 7 was used for filtering the ECG signal and then DWT (Symlets as it Mother Wavelet) is applied to the ECG signal (decomposed in 9 levels). In order to Remove Baseline we set the approximation coefficients to zero for tenth index and above, and for removing high frequency noise we set the second index of approximation coefficient and above to zero.

Then we use Butterworth filter with cutoff frequency of 40HZ in order to filter the signal and after filtering we normalize the output. We will use the derived signal by the operations done for feature extraction and to find the exact location of the P, R and S points.

IV. PROCESSING AND SEGMENTATION OF ECG SIGNAL

For detection of the QRS complex and segmentation of ECG signal different types of methods have been proposed. Some of them could be referred as methods that use direct computation of the ECG signal derivatives, methods which use digital filters for ECG signal processing, methods based on linear predictors and Wavelet transform, methods based on mathematical morphology, EMD transform, Geometrical competition of signals and Neural Networks and finally the methods which use the combinations of the techniques mentioned above [9].

In many of the methods cited above, amplitude and time thresholding on the R peak have used.

In the case of small R amplitude or the case of wide range of QRS time domain, methods based on thresholding face some problems.

In this paper we have used a new method for deriving the exact location of R-peak on the time axis that can lead to better accuracy and precision respect to the other methods used before [9]. For deriving the exact location of S on the time axis a method based on slope and gradient has been used [10] and for derivation of Q location we have used thresholding method on the slope, amplitude and the occurring time of the Q point.

A. *R detection:*

Steps must be taken into account for R peak detection is illustrated in Fig.1. In the first step the signal is preprocessed and then the signal is normalized. In the second step the Shanon Energy Envelope has been extracted and in the third step by calculating the Hilbert transform of the signal derived by the last step the R occurring interval has been estimated. Finally the exact region of R is estimated by using the ±25 samples near the last estimated point. This algorithm has been tested on MIT-BIH database and the precision of %99.80 has been achieved.

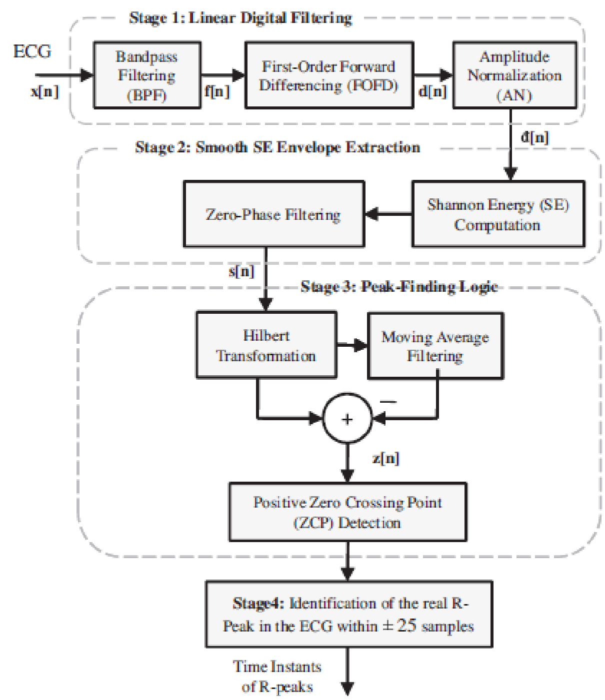

Figure 1 Detection of R peak [10]

B. Detection of S and Q points

Detection of S location is based on the apparent features of the S wave by finding the exact tilt of the points that occur afterward the R point. Fig.2 shows the method used here.

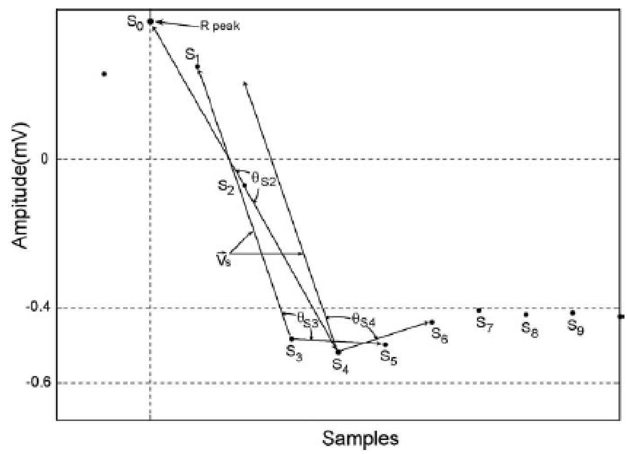

Figure 2 Detection of S peak [11]

Equation (1) is used for deriving the angle between points.

(1)

For deriving the first tilt, vector $\overrightarrow{S_0S_2}$ is used as the initial point of $\overrightarrow{V_S}$ thus in the first step θ_{S_2} is the angle between $\overrightarrow{S_0S_2}$ and $\overrightarrow{S_2S_4}$. For the next step for driving θ_{S_3}, if the slope of $\overrightarrow{S_3S_1}$ is more than $\overrightarrow{V_S}$, then $\overrightarrow{S_3S_1}$ is assumed as $\overrightarrow{V_S}$. Technically speaking, during the process of calculating θ_{S_i}, vector $\overrightarrow{V_S}$ is updated to one vector that has larger slope. The point that for first time θ_{S_i} occurs to be lower than 60°is assumed as S.

For the exact detection of Q point the combination of some requirements has been used. In an interval of 40 samples before of the detected R we are searching for a point that satisfies these following two conditions:

1- The amplitude must be smaller than 0.08

2- Existence in an interval of time which is depending to the heart rate.

If any point couldn't be found in this interval with the required characteristics, again we search for a point that it's variance respect to the point mentioned before is smaller than 0.01 and can satisfy the above 2 requirements. The numbers mentioned above have driven with trial and error.

V. APPORPRIATE FEATURE SELECTION AND EXTRACTION

One of the most important parts of the design process of an efficient classifier is appropriate selection of features that by using them the classifier can operate perfectly. Considering the previous studies based on 4 types of arrhythmia such as PVC, APC, and Normal ECG [2] and analysis of selected features mentioned in the the previous papers, 3 types of features have been selected for analyzing these types of conditions.

1- Features based on morphology.

2- Time-domain related features.

3- Features based on wavelet coefficient of each heartbeat.

These features have been extracted for each heartbeat within the 181 samples intervals nearby the R peak.

Apparent and time-domain related features are:

1- QS width
2- RR distance (Pre RR) R to the R before
3- RR distance (Post RR) R to the next R
4- RR ratio= Pre RR/ Post RR
5- R amplitude
6- QR width
7- RS width
8- Area under the QS region
9- Area under the RS region
10- Average power spectrum of each heartbeat
11- S wave amplitude

Furthermore we have used DWT (Discrete Wavelet Transform), using Daubechies 2 as the mother wavelet , to decompose the signal into 4 levels.178 detail coefficients of each heartbeat have been used as wavelet features for that heartbeat.

Thus each pattern of (each heartbeat) can be shown by using (11+178=189) total feature.

VI. CLASSIFICATION

The classification system based on Support Vector Machines, Belongs to the types of classifiers that include the target and so called the supervised classifier.

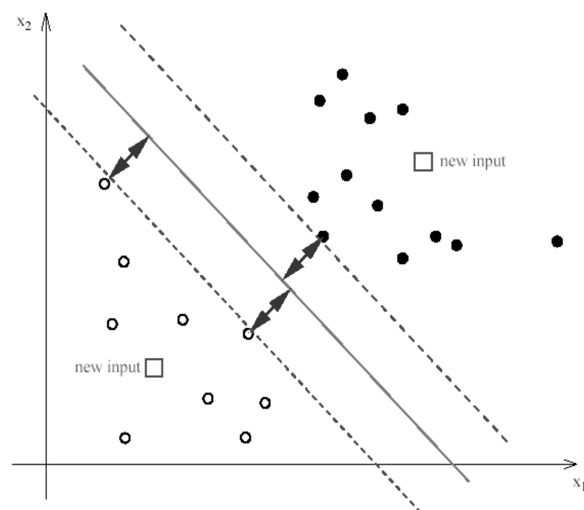

Figure 3 Support Vector Machine

SVM belongs to the group of methods that are still in wide use for classification problems and could be assumed as an efficient alternate for neural networks that were in their ultimate use, decades ahead. The purpose of SVM is to find an optimum hyper-plane to maximize the margin respect to support vectors of points that the margin could be derived by using them, called the support vectors and they are the closest points of each class respect to the hyper-plane.

SVM could be used for separating the type of data that are not linearly separable. For classifying this type of data SVM maps them to a space which has higher dimension but the mapped data can be linearly separable in the new space. This mapping could be done by the aim of kernel functions. Some types of kernel functions that are in still in their wide use could be stated as: 1- Linear Kernels, 2- Polynomial Kernels, 3- RBF (Radial Basis Function) kernels, and 4- Sigmoid kernels [11].

VII. MIT-BIH DATABASE

In this paper we have used MIT-BIH database as mentioned in table 1. This database includes 48 records that each one contains 2 ECG signal recorded from 2 different leads by the length of 30 minutes. For 45 records the first lead is II and for the other records the first lead is V2, V4, and V5.

The second lead for 40 records includes V1 and for other records includes V2, V4, and V5. The first 23 records include the most abundant type of arrhythmia and the next 23 records includes more sophisticated arrhythmia that seem to be rare respect to the first 23ones.

In this study we have tried to include different types of data. Training dataset includes the first 15 min of each signal and the test dataset includes the second 15 min of the signal.

VIII. RESULTS

In order to perform SVM method on our data we have used LIB SVM which is a library based on C++ [12]. After training the SVM network by the training features based on table I, 56 types of independent patterns from training features have been urged to the classifier. SVM performance for different kernels such as RBF and sigmoid with different C (cost arguments) and γ (variance) have been analyzed and the best result was drawn by the parameters c=7 and γ= 0.007, and we could gain the precision of 98.23%.The output results of the classification for two types of different kernels have been summarized in table II.

Table II Performance Comparison for the RBF and the Sigmoid kernel

	Class	RBF	Sigmoid
Specificity	N	100%	100%
Sensitivity	L	100%	100%
	R	96.67%	96.67%
	A	76.92%	23.07%
	V	97.14%	97.14%
Precision	All	96.97%	92.72%

Table III Confusion matrix for RBF kernel

Confusion matrix	N	L	R	A	V
N	57	0	0	0	0
L	0	30	0	0	0
R	1	0	29	0	0
A	2	0	1	10	0
P	0	1	0	0	34

The following parameters, Sensitivity, specificity and precision can be derived as follows respectively:

$$\text{Sensitivity} = \frac{\#\text{of pathologic beats that classified correctly}}{\text{total \# of patholoogic heartbeats}}$$

$$\text{Specificity} = \frac{\text{\# of normal beats that classified correctly}}{\text{\# of norma beats}}$$

$$\text{Precision} = \frac{\text{\# of beats that classified correctly}}{\text{total \# of beats}}$$

By use of the term pathologic heart beats we mean any of the heart beats that may belong to one of these types, 1- LBBB. 2- RBBB. 3- PVC. 4- APC.

Table III shows the confusion matrix of our data and table IV is a comparison between the results derived by the current research and some similar papers which were published before.

Table I MIT-BIH arrhythmia database records included in our dataset

	N	LBBB	RBBB	APC	PVC	Total
#of training attributes	392	476	1061	101	229	2259
# of testing attributes	57	30	30	13	35	165
MIT-BIH data file	101-234-–105–209 221–212–200–119	207-111-109	118–207–212 124–231	100–118–200 201–209	108-–109-203-200 207–223–107-119-233	21

Table IV Comparative results of the ECG beat classification

Ref.	#of training data beats	# of class	Segmentation	Accuracy
[7]	51020	5		85.5%
[8]	13640	5	✓	98.78%
[10]	30873	3	✓	97.14%
[12]	86938	5		96.35%
Our method	2259	5	✓	96.97%

This comparison shows that our method can use lower amounts of training data and also it has better capability of classifying the heart beats respect to other methods.

IX. Conclusion

In this paper details of autonomous classification of heart arrhythmia have been proposed. In our method after the preprocessing of the ECG signal we have detected the QRS complex by using efficient detection algorithms. For feature extraction we have used 3 types of apparent, time-domain and ECG wavelet coefficient related features and finally we have used them for training SVM.

The results which have derived by our method shows that the proposed method can operate perfectly by use of lower amount of training data and also it can classify wider range of arrhythmia respect to the other used techniques which were derived before. Our method could gain the precision of 96.97% for 5 class of data including, 1- Normal heart beats, 2- LBBB 3-RBBB 4- PVC 5- APC and can be used as one efficient tool for different type of heart arrhythmia detection. Our method can be much helpful if the size of the dataset that needs to be processed and classified exceeds in number of attributes.

Acknowledgement

By the end we are going to state our special thanks to all of the honorable persons that helped us a lot for the preparation of results for this paper including: Dr. Miad Faezipour, Dr. Mehdi Dayyan Ardestani(Physician), and Ms Deris, Karimizadeh and Mr. Yeganezad as M.sc students.

Reference

[1] G. D. Clifford, F. Azuaje, and P. McSharry, Advanced methods and tools for ECG data analysis: Artech house London, 2006.

[2] M. S. Thaler, The only EKG book you'll ever need vol. 365: Wolters Kluwer Health, 2010.

[3] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *Biomedical Engineering, IEEE Transactions on,* vol. 51, pp. 1196-1206, 2004.

[4] R. Benali, F. B. Reguig, and Z. H. Slimane, "Automatic Classification of Heartbeats Using Wavelet Neural Network," *Journal of medical systems,* vol. 36, pp. 883-892, 2012.

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

[6] C.-H. Lin, "Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier," *Computers & Mathematics with Applications,* vol. 55, pp. 680-690, 2008.

[7] S. C. Bulusu, M. Faezipour, V. Ng, M. Nourani, L. S. Tamil, and S. Banerjee, "Transient ST-segment episode detection for ECG beat classification," in *Life Science Systems and Applications Workshop (LiSSA), 2011 IEEE/NIH*, 2011, pp. 121-124.

[8] M. Kania, M. Fereniec, and R. Maniewski, "Wavelet denoising for multi-lead high resolution ECG signals," *Measurement Science Review,* vol. 7, pp. 30-33, 2007.

[9] M. S. Manikandan and K. Soman, "A novel method for detecting R-peaks in electrocardiogram (ECG) signal," *Biomedical Signal Processing and Control,* vol. 7, pp. 118-128, 2012.

[10] G.-Y. Jeong, K.-H. Yu, M.-J. Yoon, and E. Inooka, "ST shape classification in ECG by constructing reference ST set," *Medical engineering & physics,* vol. 32, pp. 1025-1031, 2010.

[11] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery,* vol. 2, pp. 121-167, 1998.

[12] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machine, 2001," *Software available at http://www. csie. ntu. edu. tw/~ cjlin/libsvm.*