Cardiac arrhythmia detection in ECG signals by feature Extraction and support vector machine

Gorav Kumar Malik
Department of Electronic and
communication Engineering.
F.E.T, Gurukul Kangri University
Haridwar, Uttarakhand-249404
Email: goravmalik@gmail.com

Dr.Yatindra Kumar
Department of Electrical
Engineering
G. B. Pant Engineering College
Pauri-Garhwal
Uttarakhand-246194
Email: kyatindra@gmail.com

Dr. Manoj Panda
Department of Electrical
Engineering
G. B. Pant Engineering College
Pauri-Garhwal
Uttarakhand-246194
Email: pandagbpec@gmail.com

Abstract—Purpose of this work is to develop an automated physiological signal diagnostic tool that can help us to early determination of arrhythmia for proper medical attention. This paper presents a simple automated approach for classification of normal and abnormal ECG based on arrhythmia. The proposed method validated by the data MIT BIH arrhythmia database. The performance in terms of accuracy for clinical decision must be very high. This method uses fourth order wavelet decomposition, wavelet decomposition used for time frequency representation and feature extraction. For classification support vector machine is used for detection kinds of ECG signals.

Index Terms—Arrhythmia; Support vector machine; Wavelet decomposition; Pre-processing; Feature extraction; ECG signals.

I. INTRODUCTION

Presently Heart diseases are one of the various illnesses that cause deaths of many Human beings. Early detection and proper clinical treatment can help to save the life of patients for that purpose features of heart beats are used. Heartbeat is a kind of Physiological signal generated due to electrical activity in the heart. The Electrocardiogram (ECG) generated due to changes in Bioelectric Potential respect to time like the human heartbeats. ECG signal play important role in the diagnostic of arrhythmias [1].

An arrhythmia is abnormal heart beat, the primary and basic classification is two type bradycardia and trchycardia, when heart rate is less than 60 BPM its bracycardia and if heart rate is more than 100 BPM it is trchycardia. And both have different effect on the human being like bradycardia causes a drowsiness, fanting, sleepiness and rare chances of cardiac arrest, but Tachycardic affect the pumping capability of the heart and generate the symptoms chest pain, Problem in breathing and cause of heart attack.

A heart beat can be represented in terms of QRS, T and P wave as shown in figure 1.For the arrhythmia detection we are interested in the beat morphology (Normal and abnormal pattern) of different waves of ECG signals. So perform wavelet decomposition operation, in this process downsampled the signal for reduction in detailed feature of ECG signals, we use fourth level decomposition and choose a pattern similar to the original pattern. Locate value of ECG signal from second order decomposition and get the R peaks and some more feature can be extracted based on location of R, T, S waves and their respective amplitude.
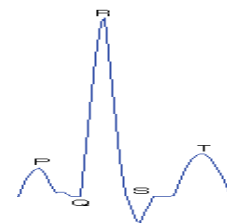

Figure 1: ECG signal

The extracted feature used for classification by the support vector machine. It is the most efficient tool for classification problems. Although there are various method for classification of arrhythmias, but the attention in this work is developing an automated simplest algorithm and arrhythmia detection.

II. DATABASE

The Physiological ECG signal use of this work provided by the MIT BIH arrhythmia database [2]. The MIT BIH arrhythmias data base includes 48 ECG recording; these recordings are half an hour long. Signals are sampled 250 Hz. In the database each signal has an annoted file that includes the beat, rhythm and other information. This database is used by researchers to test their algorithms for detection of arrhythmias and classifications. In present research work we use 32 records of full length to classify into normal or abnormal class, abnormal for arrhythmia ECG signal.

III. METHODOLOGY

Methodology of this work is shown in this section. We use RAW ECG signal of two type normal and abnormal (Arrhythmia signal), these signal are acquired from MIT BIH database as explained previous section, the complete database

Figure 2: Methodology Block Diagram

contain 32 patients data. The process of work carried in this research work is shown in figure.

A.Pre-processing

All kind of physiological signals are weak in nature mixed with different kind of interferences, The aim of Preprocessing is just use to remove the signals interferences. To achieve the highest accuracy. It is necessary to remove Interferences from ECG signal. Otherwise wrong classification leads to wrong clinical treatment. There are generally four main categories of interferences low frequency interferences, baseline interferences, power line interferences and high frequency interferences. To removal of these noises we use three kinds of filters.

Adaptive filter:

Adaptive filter is a kind of filter in which changing of filter parameter give optimum results for power line interference reduction [3], Adaptive filters 'learn' with initial input signal for track them to remove the signal interferences by changing the filter parameters.

Butterworth Low Pass and High Pass Filter:

Interference of low frequency and base line drift is major problem, if there is large drift in the ECG signal then it change heart rate and location of different waves in ECG. So the proper attention required correcting the baseline wander inference, to remove the low frequency noise and baseline wander we use a low pass Butterworth filter.

Butterworth filter also best for removal of high frequency interferences due to its simplicity and maximum flat magnitude response, the filter function is represented as

$$H(Z) = \frac{g'(1 - Z^{-1})^p}{\sum_{p=0}^{N} b_p Z^{-p}}$$

Where bp series are filter coefficients and g' is gain factor

B.Feature extraction:

Classification of arrhythmia by support vector machine requires generation of input vector; an input vector to classification should include the morphology and rhythm feature of ECG signal as shown in figure [3]. Feature extraction step is very influential for arrhythmia detection. The objective of feature extraction step is to derive a set of extracted parameter. Those signalize the ECG signal. These set should contain maximal information about ECG signal, so the selection criteria of these feature parameter is very important for Arrhythmia classification.
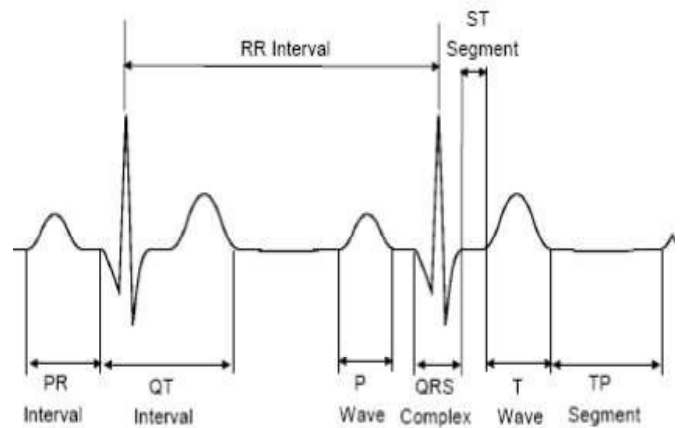

Figure.3: ECG Waveform [4]

In this work input vector for classifier contain two kind of feature Morphological and Frequency domain feature as listed in table no.1. Discrete wavelet Transform (DWT) use for feature extraction, Discrete Wavelet transform is one of the newly used techniques for feature extraction. Signals decomposing into elementary building blocks that are well localized both in time and frequency [5]. Its application to Physiological signal processing has been at the cutting edge of its developments where it has been found particularly useful in the study of these, often problematic, signal like ECG [6]

Selection of appropriate decomposition levels for wavelet levels is play crucial role in the analysis of signals using the discrete wavelet Transform (DWT) [7]. Dominant frequency components are use to choose the decomposition levels of the signal. In the present work, the decomposition levels were chosen to be 4. Therefore, the ECG signals were broken up into the details D1 − D4 and one final estimate, DWT design can take many forms. Sherlock and Monro [8] developed a way to determine the coefficients of a filter bank. They derive an orthonormal perfect-reconstruction finite impulse response (FIR) filter of arbitrary length. The z-domain filter coefficients are represented by a low pass filter

$$H(Z)\,H(Z^{-1}) + H(-Z)\,H(-Z^{-1}) = 1 \qquad (1)$$

The z transform of high-pass filter h is H (z) and represented as

$$G(Z) = z\,H(-Z^{-1}) \qquad (2)$$

Increasing sequence length of filters (indexed by j) can be obtained:

$$H_{j+1}(Z) = H(Z^{2j})H_j(Z)$$

$$G_{j+1}(Z) = G(Z^{2j})H_j(Z) \quad (3)$$

Where j=0, 1.............................K-1

With initial condition $H_0(z)=1$, we can express its relationship in time domain

$$h_{j+1}(p) = [h]_{\uparrow 2^j} * h_j(p)$$

$$g_{j+1}(p) = [g]_{\uparrow 2^j} * h_j(p)$$

The subscript $[.]_{\uparrow q}$ show q factor up-sampling of and p is the equally sampled discrete time. The normalized wavelet and scale basis functions $\psi_{il}(p)$ $\varphi_{il}(p)$, can be defined as.

$$\varphi_{i,l}(p) = 2^{j/2} \; h_j\left(p - 2^j \; l\right)$$
$$\psi_{i,l}(p) = 2^{j/2} \; g_j\left(p - 2^j \; l\right)$$

(Scale parameter j, translation parameterl)

Factor $2^{j/2}$ is inner product normalization, the discrete WT decomposition can be described as.

$$s_{(j)}(l) = x(k) * \varphi_{i,l}(p)$$
$$d_{(j)}(l) = x(k) * \psi_{i,l}(p)$$

Where $s_{(j)}(1)$ is approximation coefficients at resolution.

$d_{(j)}(1)$ detail coefficients at resolution.

In the present work smoothing of the Daubechies wavelet of order 4 uses, here we chose order according to our need. Decomposition of a signal $x^{(N)}$ can be represented as.

$$x^{(N)} = g^{(N-1)} + g^{(N-2)} + \cdots + g^{(N-M)} + x^{(N)}$$

(N=level of decomposition)

The ECG feature described in Fig. 2, were obtained from the MITBIH database signals using Matlab. The R peak of heartbeat of the ECG signals was obtained by determining the peak having higher amplitude than a threshold value. A Similar method was employed to obtain other wave location like P and T, after finding the location we also calculate the distance between them. Extracted features shown in table 1, total 18 features are extracted and used for classification purpose.

Table 1: Feature of ECG

S.No	Feature Symbol	Feature Description
1.	P wave	Duration of P wave in second
2.	S wave	Duration of S wave in second
3.	T wave	Duration of T wave in second
4.	Q wave	Duration of Q wave in second
5.	RS Slop	Slope Between R and S
6.	QS Slop	Slope Between Q and S
7.	PR Interval	Duration between P and R wave in second
8.	ST Interval	Duration between S and T wave in second
9.	QT Interval	Duration between Q and T wave in second
10.	TP Interval	Duration between T and P wave in second
11.	QR Height	Amplitude of QR wave
12.	RS Height	Amplitude of RS wave
13	HRV	Heart Rate Variability
14	Energy	Signal Energy
15	Area	Area Of QRS
16	Non linear Energy	Energy of signal in sum of sampled energy form
17	Spectral Entropy	∑ PSD. log(PSD)
18	Peak Power	Max(Power spectral density)

C. Classification by SVM:

SVM is a well-known supervised learning method [9],For the classification, we will adopt in this work support vector machine (SVM) approach which has shown particularly effective in numerous application fields including the classification of ECG signals [10]. The SVM proposed by Vapnik (1995) has been most widely used for classification.

To understand SVM lets consider for simplicity a supervised two class (Binary) classification problem. Here we can apply a maximal margin classifier; it is a simplest version of SVM applicable to linearly separable data [11]

Now we have a dataset $\{(Y_j, d_j)\}$, Where Y_j is the input pattern for jth example and d_j is the corresponding desired output {-1, 1}.

For the decision surface a hyper plane working. We can represent:

$$W^T Y_j + b \geq 0 \qquad \text{then } d_j = +1$$
$$W^T Y_j + b \leq 0 \qquad \text{then } d_j = -1$$

Where $W^T Y_j + b$ is the output function. The minimum distance between points to hyper plane called geometric margin. For the good classification we want to maximize the geometric margin. For that first we introduce functional margin $W^T Y_j + b$ for the linearly separable data we can write;

$$W^T Y^+ + b = +1$$
$$W^T Y^- + b = -1$$

The closest data point on the hyperplane is Y^+ (Y^-), Y^+ for positive side and Y^- for negative side. Now we can compute the geometric margin

$$\gamma = \frac{1}{2}\left(\frac{W^T Y^+ + b}{|W|} - \frac{W^T Y^- + b}{|W|}\right)$$
$$\gamma = \frac{1}{2|W|}(W^T Y^+ + b - W^T Y^- - b)$$
$$\gamma = \frac{1}{2|W|}(1 - (-1))$$
$$\gamma = \frac{1}{|W|}$$

Geometric margin can be maximizing by fix the functional margin value to one and minimize the weight vector $|W|$. By the Lagrange multipliers ($\alpha_j \geq 0$) we can transform in to dual problem.

$$\alpha_j[d_j(W^T Y_j + b) - 1] = 0$$

Which means only the point with functional margin unity contribute to the output function [5]

The problem solution by SVM done by mapping the training data into a higher dimension and transform them to feature vector. Then transform function used for linear separation.

IV. Results and discussion

The experimental Work carried out in MATLAB and for this work MITBIH arrhythmia database is used. We have used 32 recorded signals divided into normal and abnormal categories, (15 normal, 17 abnormal) each recording is 30 minutes long [12].

The DWT transform used for feature extraction where total 18 features are extracted. We combine these features for training and testing vector of SVM classifier. For the training purpose, we use these different inputs and text matrix to meet the best accuracy of the classification as shown in table 2.

Table: 2 Value of SVM parameters

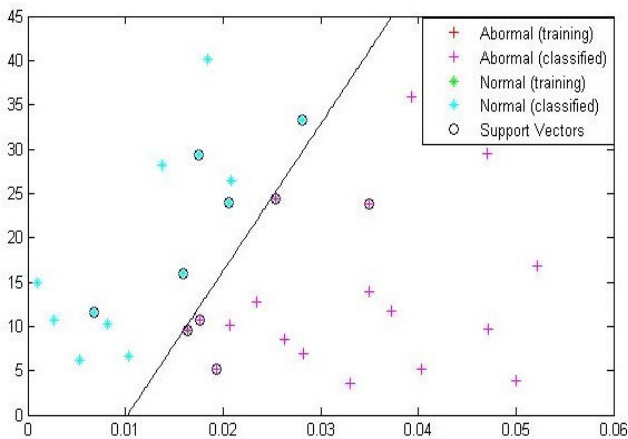| Training data	Specificity	sensitivity	Classification accuracy
24%	78.33	82.35	78.2
30%	86.67	88.24	87.5
36%	98.84	97.12	98.2

Figure 4: SVM classifier output

The SVM classifier trained with input data, initially we take only 24% of the total data for training purpose, and the classifier result is shown in figure 4. Then we achieve the 78.2 % accuracy. But if we increase the training data to 36%, then we achieve highest accuracy up to 98.2 %. So the higher training data give optimal classified result. We can also increase the accuracy if we increase the total number of record used, in this situation less percentage of training data also effective.

The result obtained by proposed method detects arrhythmia with 98.2% accuracy. Clearly indicates that proposed method is better for arrhythmia detection.

V. CONCLUSION

In this paper, we proposed an automated arrhythmia classification system using different feature of ECG signals. We discussed the classification result of SVM shows the ECG signal feature can be used as a reliable indication of cardiac problems. Generally results are not achieved with 100 percent accuracy. The accuracy of the suggested system depends upon several factors like quality and size of training data and also used parameters. However the present method shows the result with 98 percent accuracy.

REFERENCES

[1] Bhyri, Channappa, S. T. Hamde, and L. M. Waghmare. "ECG feature extraction and disease diagnosis." Journal of medical engineering & technology 35.6-7 (2011): 354-361..

[2] Goldberger, Ary L., et al. "Physiobank, physiotoolkit, and physionet." Circulation 101.23 (2000): e215-e220.

[3] Kumar, Yatindra, and Gorav Kumar Malik. "Performance analysis of different filters for power line interface reduction in ECG signal." International Journal of Computer Applications 3.7 (2010): 1-6.

[4] Singh, Yogendra Narain. "Human recognition using Fisher's discriminant analysis of heartbeat interval features and ECG morphology." Neurocomputing 167 (2015): 322-335.

[5] Li, Cuiwei, Chongxun Zheng, and Changfeng Tai. "Detection of ECG characteristic points using wavelet transforms." IEEE Transactions on biomedical Engineering 42.1 (1995): 21-28.

[6] Übeyli, Elif Derya. "Support vector machines for detection of electrocardiographic changes in partial epileptic patients." Engineering Applications of Artificial Intelligence 21.8 (2008): 1196-1203.

[7] Rai, Hari Mohan, Anurag Trivedi, and Shailja Shukla. "ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier." Measurement 46.9 (2013): 3238-3246

[8] Sherlock, Barry G., and Donald M. Monro. "On the space of orthonormal wavelets." IEEE Transactions on Signal Processing 46.6 (1998): 1716-1720.

[9] Semwal, Vijay Bhaskar, et al. "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification." Multimedia Tools and Applications: 1-19.

[10] Melgani, Farid, and Yakoub Bazi. "Classification of electrocardiogram signals with support vector machines and particle swarm optimization." IEEE transactions on information technology in biomedicine 12.5 (2008): 667-677.

[11] Abe, Shigeo. Support vector machines for pattern classification. Vol. 2. London: Springer, 2005.

[12] Rai, Hari Mohan, Anurag Trivedi, and Shailja Shukla. "ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier." Measurement 46.9 (2013): 3238-3246.