DENOISING OF ECG SIGNALS USING WAVELETS AND CLASSIFICATION USING SVM

[1]SHIVANESAN S M., [2]PRADHEEP M., [3]SHARATH K., [4]ARAVIND PRASAD., [5]MANOJ M., [6]GANESAN M.

[1,2,3,4,5]Students of ECE department, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India.
[6](Assistant Professor)is with Department of ECE, Amrita, Vishwa Vidyapeetham, Coimbatore 641112, India.
E-mail: shiva.rathna@gmail.com, meetganesan@gmail.com

Abstract- Electrocardiogram is the recording of the electrical potential of heart versus time. The analysis of ECG signal has great importance in the detection of cardiac abnormalities. In this paper we have dealt about the removal of noises in ECG signals and arrhythmia classification of the signal.The inputs for our analysis is taken from MIT-BIH database (Massachusetts Institute of Technology Beth Israel Hospital database). The denoising is done through wavelet transform and thresholding. Confirmatory tools such as Poincare plot and Detrended Fluctuation Analysis (DFA) are used to find out the healthiness of the signal. Then Support Vector Machine (SVM) is used to find out what type of arrhythmia is present in the signal.

Keywords- Classification, DFAElectrocardiogram, MIT-BIH database, Poincare, SVM , Wavelets.

I. INTRODUCTION

In today's environment there has been lot of threats due to heart disease and no proper diagnosisWith the recent developments in technology, physicians have powerful tools to observe the working of the heart muscle and thus to establish their diagnosis. Among cardiovascular examinations, electrocardiogram (ECG) analysis is themost commonly used and very effective too. This is due to the fact that ECGpresents useful information about the rhythm and the electricalactivity of the heart. Thus, it is used for the diagnosisof cardiac arrhythmias worldwide. For effective diagnostics, the study ofthe ECG signal must be carried out for several hours. For thisreason, researchers have been interested in enablingcomputers to classify the abnormal ECG signals. During the last five decades the analysis of ECG signals evolved from simple visual examinations to totally automated analysis [1, 2, 3].

A noise free signal is necessary in any type of signal analysis and classification [4]. Several algorithms have been proposed for denoising of the biomedical signals, especially ECG signals. Advancement havebeen done to various fields for denoising and classification of ECG signals.ECG signal is one of the bio signals that is considered as a non-stationary signaland needs a hard work to denoise [5, 6]. An efficienttechnique for such a non-stationary signal processing isthe wavelet transform. The wavelet transform can beused as adecomposition of a signal in the time frequency scale plane.

There are many applications of wavelet transform such as sub-band coding datacompression, characteristic point's detection and noisereduction. In order to reduce the noise of ECG signalmany techniques are available like digital filters (FIR orIIR), adaptive method and wavelet transform thresholding methods. However, digital filters andadaptive methods can be applied to signal whosestatistical characteristics are stationary in many cases.Recently thewavelet transform has been proven to beuseful tool for non-stationary signalanalysis.
Thresholding is used in wavelet domain to smooth out or to remove coefficients of wavelet transform.The denoising method thatapplies thresholding in wavelet domain has beenproposed by Donoho as a powerful method [7].

Before classification of these signals to find out the arrhythmia, the need to know whether the signals is abnormal or not becomes necessary. To find out the healthiness of the signal we used two methods Poincare plot and Detrended Fluctuation analysis (DFA) [8, 9, 10]. After confirming with the abnormality of the signal simple Support Vector Machine (SVM) is used to train the signals and find out the exact arrhythmia in the ECG signal.In [11], two classification systems based on the support vector machines (SVM) approach areimplemented.

The signals for analysis are taken from the MIT-BIH database [12].We have discussed about how wavelet can efficiently denoise the obtained from the database. The Poincare plot and Detrended fluctuation analysis are done in order to confirm the abnormality of the signal. The classification of the ECG signal is based on the major arrhythmias that causes threats to human life viz. Bradycardia Tachycardia, cardiac and ventricular. SVM trains, classifies and finds out what type of abnormality is present.
The article is organised as follows: section II tells about the principle behind wavelet transform and

thresholding. Section III describes about the Poincare plot and Detrended fluctuation Analysis (DFA). Section IV tells about the SVM classification of arrhythmia. Section V discusses about the results obtained and tabulation. We present our conclusions and future work in the section VI. The overall block diagram is depicted in Fig. 1.

II. WAVELET TRANSFORM

A. Discrete Wavelet Transform:

The wavelet transform is similar to the Fourier transform. For the FFT, the basis functions are sines and cosines. For the wavelet transform, the basis functions are more complicated called wavelets, mother wavelets or analysing wavelets and scaling function. In wavelet analysis, the signal is broken into shifted and scaled versions of the original (or mother) wavelet. The fact that wavelet transform is a multiresolution analysis makes it very suitable for analysis of non-stationary signals such as the ECG signal [13].

The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules. In other words, this transform decomposes the signal into mutually orthogonal set of wavelets.The DWT can be realized in terms of high pass and low pass filters. The approximation properties of filter banks and their relation to wavelets are presented in the paper [14]. The output of the low pass filter gives the approximation coefficients and the output of the high pass filter gives the detailed coefficients.



Fig. 1. Overall block diagram

The DWT of an discrete signal x(n) of length M-1 is given in the below equation.

$$x(n) = \sum_k W\varphi(j_o, k)\varphi_{j0,k}(n) + \sum_{j=j0}^{\infty} W_\varphi(j_o, k)\varphi_{j,k}(n) \quad (1)$$

Here $W\varphi(j_o, k)$ and $W\psi(j_o, k)$ are called the wavelet coefficients.$\varphi_{j,k(n)}$and$\varphi_{j0,k(n)}$are orthogonal to each other. Hence we can simply take the inner product to obtain the wavelet coefficients.

$$W_\varphi[j_0, k] = \frac{1}{\sqrt{M}}\sum_n x(n)\varphi_{j_0,k}[n] \quad (2)$$

$$W_\varphi[j, k] = \frac{1}{\sqrt{M}}\sum_n x(n)\varphi_{j_0,k}[n] \quad (3)$$

The coefficients $W\varphi(j_o,k)$ are called the approximation coefficients and the coefficients $W\varphi(j_o, k)$ are called the detailed coefficients. The DWT can be realized in terms of high pass and low pass filters. The approximation properties of filter banks and their relation to wavelets are presented in the paper [14]. The output of the low pass filter gives the approximation coefficients and the output of the high pass filter gives the detailed coefficients.Computation of the wavelet coefficients at every possible scale is a fair amount of work and it generates an awful lot of data. Selection of a subset of scales and positions based on powers of two (dyadic scales and positions) results in a more efficient and accurate analysis.

B. Wavelet Decomposition and Thresholding:

The DWT decomposes the signal into approximate and detail information as discussed.The wavelet decomposition process can be iterated, so that one signal is broken down into many lower resolution components. This is called the wavelet decomposition tree.

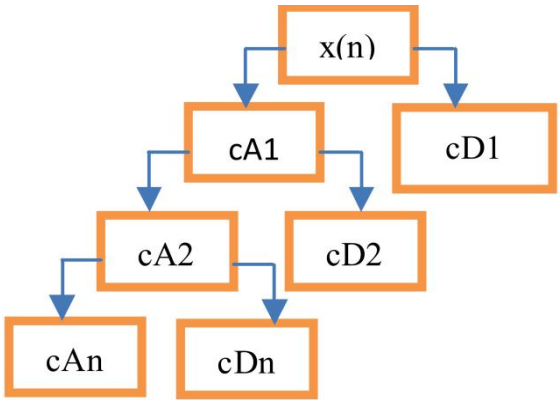


Fig. 2. Decomposition tree

In this proposed method, the corrupted ECG signal x(n) is denoised by taking the DWT of raw and noisy ECG signal. A family of the mother wavelet is available having the energy spectrum concentrated around the low frequencies like the ECG signal as well as better resembling the QRS complex of the ECG signal. We have used daubacheswavelet, which resembles the ECG wave.In discrete wavelet transform (DWT), the low and high frequency components in x(n) is analysed by passing it through a series of low-pass and high-pass filters with different cut-off frequencies.

This process results in a set of approximate coefficients (cA) and detail coefficients (cD). To remove the power line interference and the high frequency noise, the DWT is computed to level 4 using daubaches16 (DB16)mother wavelet function and scaling function. Then the approximate coefficients at level 5 (cA5) are set to zero. The residue of the raw signal and the approximate noise is obtained to get noise free ECG signal.The method is based on taking the discrete wavelet transform (DWT) of a signal, passing this transformthrough a threshold, which removes the coefficientsbelow a certain value.

$$T = \sigma\sqrt{2 * \log(n)} \quad (4)$$

Where, T is threshold, n is the number of samples and σ is the noise standard deviation. Thresholding is applied at every loop to smooth out the signals and denoise the raw data

.

III. FEATURE EXTRACTION

Classification to find out the arrhythmia becomes unnecessary when the ECG signal is normal and does not contain any sort of abnormalities. To visually verify and find out whether the signals are normal or not we use two highly accurate methods. The Poincare plot and the Detrended Fluctuation analysis (DFA).

A. Poincare Plot:

The Poincare plot of RR intervals is one of the techniques used in heart rate variability (HRV) analysis. It is both a useful visual tool which is capable of summarizing an entire RR timeseries derived from an electrocardiogram in one picture, and a quantitative technique whichgives information on the long- and short-term HRV. A Poincare plot of RR intervals is composed of points (RR_i,RR_{i+1}), that is each pointin the plot corresponds to two consecutive RR intervals [15]. The resulting cloud of points isusually characterized by its length (SD2) along the line of identity and its breadth across this line (SD1).The visual inspection of the formed shapes of the Poincare plot of RR intervals is a widely used method for analysing the quality of recorded ECG signals and to identify premature andectopic beats, as well as technical artefacts. The plot of a healthy person will have a comet shaped structure along the line of identity [16]. Also the ratio of the Poincare descriptors (SD1, SD2) should be high for a healthy person.

Let RR, x and y vectors be defined as

$$RR = (RR_1, RR_2 \dots\dots RR_n, RR_{n+1}),$$

$$x = (x_1, x_2, \dots\dots, x_n) = (RR_1, RR_2, \dots\dots, RR_n)$$

$$y = (y_1, y_2, \dots\dots, y_n) = (RR_2, RR_3, \dots\dots, RR_{n+1}), \quad (5)$$

$$SD1 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{i1})^2} \quad (6)$$

$$SD2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{i2})^2} \quad (7)$$

$$d_{i1} = \frac{(x_i - \bar{x}_i) - (y_i - \bar{y}_i)}{\sqrt{2}}, \quad d_{i2} = \frac{(x_i - \bar{x}_i) + (y_i - \bar{y}_i)}{\sqrt{2}} \quad (8)$$

The over bar stands for mean.

By calculating SD1 and SD2 from the eq. (6) & (7) and finding the ratio of the descriptors we could comment on the abnormality.

B. Detrended Fluctuation Analysis (DFA):

This is also a plot analysis by which we could visually see and comment on the abnormality of the given ECG signal. The plot here is a double logarithmic plot. Its logF(n) versus log(n).The main objective of DFA is to extract the extrinsic fluctuations in order to allow the analysisof the signal's variability associated exclusively withautonomic control.The integrated signal y(k) is then segmented into multiple windows of length n. For each of these windows,a least-squares first-order approximation (a linesegment) is calculated, representing the "trend" ofthat segment of the signal that has been found out. The trend signal yn(k) in Eq.(10),formed by the line segments, is an approximation tothe integrated signal y(k).

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} e(k)^2}, \quad (9)$$

$$e(k) = y(k) - y_n(k). \quad (10)$$

$$y(k) = \sum_{i=1}^{k}[RR(i) - RR_{ave}] \quad (11)$$

Where $RR_{(i)}$ is the duration of the i-th RR interval,RR_{ave} is the mean interval, and k is the current outputsample time-index. The slope of the line (log F(n)& log (n)) gives us a coefficient called α which gives us the HRV fluctuations. The normal plot of DFA is shown in the Fig3.

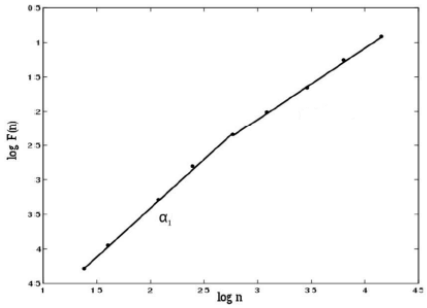


Fig3. Normal DFA plot.

IV. SUPPORT VECTOR MACHINE

The support vector (SV) machine is a new type of learning machine. It is based on statistical learning theory.Support vector machines (SVMs) are becoming popular in a wide variety of biological applications. The main use of SVM is classification. They were used in cancer cells classification and now SVM's are used widely in biomedical signal analysis. Here we have employed SVM to classify the ECG signal and to find out the arrhythmia present in it.SVM finds the optimal separating hyper plane (OSH) with the minimum errors. The linear separation hyperplane is inthe form of

$$f(x) = w^T x + b. \quad (12)$$

But the ECG signals involve more than a b than two classes, so we need a classifier that is more than a binary classifier [16, 17]. The widely practised ones are one-against-all (OAA) and the one-against-one (OAO) strategies.The one against one constructs $\frac{n(n-1)}{2}$decision functions for all the combinations of class pairs. Experimental results in [18] indicate that

the one-against-one is more suitable forpractical use. (More details appeared in [18]). We use OAOfor ECG multi class classification.

V. RESULTS AND TABULATION

A. Wavelet denoising:

The denoised signal is evaluated on the basis of SNR and correlation factor. Records or samples from MIT-BIH database were used. The tabulation below shows the records used and the SNR of the filters.DB16 was found to denoise effectively.

$$SNR = \log_{10} \frac{\sum_{n=0}^{N} V_R{}^2(n)}{\sum_{n=0}^{N} S_R{}^2(n)} (13) S_R(n):$$

the deformation in reconstructed ECG signal.

V_R (n): the original signal.

Table1. SNR values of various records and filters

SAMPLES MIT-BIH	DB4	DB6	DB16	Sym8
100	12.37	11.86	13.06	7.37
105	13.15	9.47	11.65	4.75
112	12.47	11.9	12.92	6.84
228	18.05	16.5	17.57	11.72

Table2. Correlation values

SAMPLES MIT-BIH	DB4	DB6	DB16	Sym8
100	0.8279	0.4599	0.9325	0.1120
105	0.3146	0.2935	0.5848	0.0662
112	0.6291	0.4878	0.6661	0.4522
228	0.6222	0.4826	0.7929	0.4882

From the Table 1 and Table2 we found out that DB16 denoised the signal efficiently and can be used for effective denoising of ECG signals.

B. Poincare Plot:

The Fig4 below shows the Poincare plot of the record 228 of the MIT-BIH database. From the figure we could see thatthe shape is not comet shaped along the line of incidence (LOI) as discussed in section three. So just by looking at graph of the Poincare plot we can tell that the signal is abnormal and contains some kind of artefact.

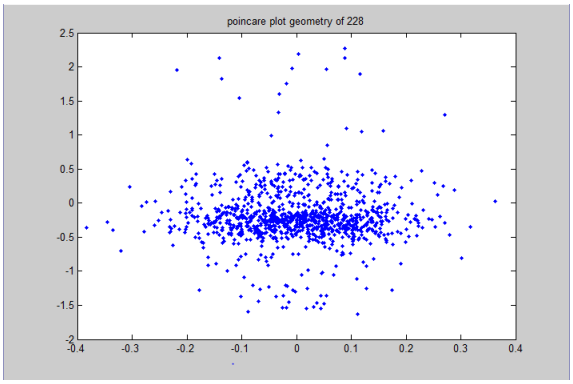


Fig4. Poincare plot of record 228

SD1= 317.2

SD2= 317.0

$$\frac{SD1}{SD2} = 1.0062$$

The ratio of SD1 and SD2 is very low and hence the signal is abnormal.

C. Detrended Fluctuation Analysis:

The fig5 below shows the DFA plot of the record 228 of the MIT-BIH database. Comparing the Fig5 with the normal plot from the Fig3 we see the plot is disfigured and the abnormality is confirmed.

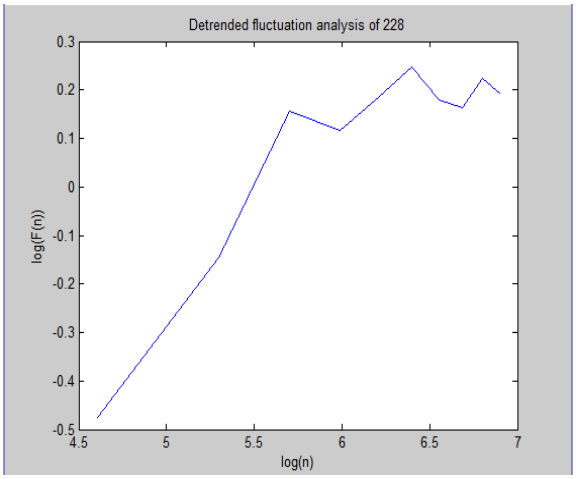


Fig5. DFA plot of record 228

The alpha value is found from the slope of the line formed in the plot.

α = 0.2783 (Anti-correlated)

D. Support Vector Machine:

The signals are trained and classified by the SVM algorithm as discussed earlier. The table below shows the arrhythmia present in the record 228.

Record	Normal/abnormal	Type of arrhythmia
112	Abnormal	Bradycardia
219	Abnormal	Tachycardia
228	Abnormal	Bradycardia

Table 3. SVM arrhythmia classification

CONCLUSION AND FUTURE WORK

The results shown above are the outputs of the record 228. Similarly other records from the MIT-BIH database were run and the outputs were verified and type of arrhythmia present in it was found and analysed. The Poincare plot and Detrended Fluctuation Analysis (DFA) are highly accurate in determining the abnormality of the ECG signal. It's fast and just by looking at the output plot of the ECG signal we can comment on the abnormality of the signal. This paper throws light on the method of denoising that can efficiently remove various noises like baseline wander, powerline interference etc.from the ECG signal. We have also seen that the best filter for denoising ECG signals was DB16 which has got the maximum number of vanishing moments. The higher the vanishing moments higher will be the amount of denoising but not at the cost of losing the original data from the signal. SVM has always been an accurate classifier and one against one (OAO) strategy is used for multiclass classification of ECG signals. Keeping in mind that ECG being a non-stationary signal, we have designed the entire block that can handle the non-stationary ECG signal.

The inputs to the whole setup was taken from the MIT-BIH database and not real time ECG signals. In the future ECG signals can be acquired from the human and can be sent as the inputs and the type of arrhythmia can be found. Since it is going to be real time acquisition of data, the amount of noise present will be more, so care must be taken while denoising the real time signals.

REFERENCES

[1] Henzel N., and J. Leski, 1999. "Efectywnaobliczeniowometodaanalizyacyklicznychzdarzaprzypomocytechnikwieloczstotliwociowych," XI konferencjaBiocybernetykaiInzynieriaBiomedyczna, pp 188-122.

[2] Henzel N., and J. L ski, 1999. "AnalizasygnaluHRV w podpasmachwidmowych. BiocybernetykaiInynieriaBiomedyczna," PAN, pp 418-422.

[3] Merri M., M. Alberto M. and A. J. Moss, 1993."Dynamic analysis of ventricular representationDuration from 24-hour recording". IEEETransactions on biomedical engineering, Vol. 40,No. 12.

[4] Luca Mainadri T., A. M. Bianchi and S. Baselli,Cerutti, 1995. "Pole-tracking algorithms for the extraction of time-variant heart rate variability spectral parameters," IEEE Transactions on biomedical engineering, Vol. 42, No.3. pp: 20-31.

[5] Leski J., 1991, "Detectjazespolów QRS dlazaklóconychsignalów EKG," Post. Fiz. Mid., 26, 3-4 PL ISSN 0137-8465.

[6] Shrouf A. 1994. "The lineal prediction methodsanalysis and compression", PhD thesis of lskTechnical University in Gliwice.

[7] D. L. Donoho, 1991. "De-noising by softthresholding", IEEE Transaction on Information Theory, Vol. 41, pp. 613–627, May 1995

[8] M. Brennan, M. Palaniswami and P. Kamen, IEEE Trans.Biomed. Eng. 48, 1342 (2001).

[9] M. Brennan, M. Palaniswami and P. Kamen, Am. J. Physiol. Heart Circ. Physiol. 283, H1873 (2002).

[10] K. Hu, P.C. Ivanov, Z. Chen, P. Carpena and H.E. Stanley, "Effects of trends on detrended fluctuation analysis," Physical Review E, 2001,vol. 64, 011114.

[11] S. Osowski, T. Linh, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," IEEE Transactions OnBiomedical Engineering, vol. 51, no. 4, pp. 582–589, Avril 2004.

[12] www.physionet.org

[13] Mahmoodabadi and S. Ahmadian, "ECG feature extraction based on multiresolution wavelet transform", Proceedings of the IEEE 27th Annual Conference on Engineering in Medicine and Biology, pp. 3902–3905, Shanghai, China, January 2005.

[14] M. Vettereli, "Wavelets, approximation and compression", IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 59–73, August 2001.

[15] Brennan M, Palaniswami M and Kamen P 2001 "Do existing measures of Poincar´eplot geometry reflect nonlinear features of heart rate variability?" IEEE Trans. Biomed. Eng. 48 1342–47

[16] F. Melgani and L. Bruzzone, "Classification of hyper spectral remote sensingimages with support vector machine," IEEE Trans. Geosci. Remote Sens.,vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[17] W.Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," IEEE Trans. Neural Networks., vol. 13, no. 2, pp.415–425, Mar. 2002.

[18] Hsu, C.-W., Lin, C.-J., "A comparison of methods for multiclass support vector machines", IEEE Trans. Neural Networks 13(2), 415-425,2002

★★★