© 2016 IJSRST | Volume 2 | Issue 5 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Engineeringand Technology

Classification of Arrhythmia using KNN-Classifier

Kandala. S. S. V. V. Ramesh, Ch. Nagabhushana Rao

Computer Science and Engineering, Dadi Institute of Engineering and Technology, Anakapalli, Andhra Pradesh, India

ABSTRACT

In this paper we are introducing a new methodology called discrete wavelet transform along with the higher order statistics (HOS). The feature vectors give the information about the original signal, these feature vectors are a compressed version of the original signal. In this paper we are using the discrete wavelet transform (DWT) for the calculation of feature vectors. In this methodology we have mainly three stages. In the first stage data is collected from the MIT/BIH database, after collecting the data segmentation is done. In the second stage, we perform DWT on the segmented data. In this process the data is mainly divided into approximated coefficients and detailed coefficients. In the third stage, we are calculating HOS (cumulants) for the detailed coefficients. Finally the signal is given to the decision making device, according to the threshold value the beats are classified faithfully. In this process we are getting Precision 96.30%, Recall 96.30% and by using K-NN classifier.
Keywords : Cumulants, Discrete Wavelet Transform, K-NN classifier

I. INTRODUCTION

In the Today's world large numbers of human beings are frequently affected by the cardiac diseases. So it is essential to know the functioning of heart. ECG is such an essential bio-electric signal which gives the information about the heart functioning. The ECG signal is characterized by the waves P, QRS and T. Among these QRS complex is very important [1]. The ECG signal is generated by the heart muscles on depolarization and repolarization of the atria and ventricles. ECG is the primary tool to know the status of the heart. The ECG signal is a non-stationary signal, the amplitude values are changed at every instant in the rapid manner. That‟s why it is very difficult to analyze the state of the heart accurately. Minute deviations in the ECG beat leads to different arrhythmia, we should careful about these changes. Due to this non-stationary nature of the ECG signal it is easily effected by the noises like muscular noise, baseband line noise and power line noise etc. [2]. To eliminate such noises we are introducing the DWT along with the higher order statistics. In the DWT we are getting approximated and detailed coefficients. Approximated coefficients have the information about the low frequency signal, detailed coefficients gives the information about the high frequency signal [3]. In those coefficients we are using the detailed coefficients to extract the information. In the section II the proposed method details are given, section III provides the information about the results and section IV gives the conclusion of the proposed method.

II. METHODS AND MATERIAL

Proposed Method

A. ECG Signal Extraction

This process is also known as the pre-processing of input ECG signal. The ECG signal is extracted from the MIT/BIH database [4]. The database provides nearly 48 recordings. Mostly we are taking each record with 30 minutes duration. The ECG signal is mainly composed of five types of beats [5].

They are listed below,

1. N (Normal Sinus Rhythm)
2. APC (Atrial Premature Contraction)
3. PVC (Premature Ventricle Contraction)
4. V (Ventricular Contraction)
5. RBBB (Right Bundle Branch Block)
6. LBBB (Left Bundle Branch Block)

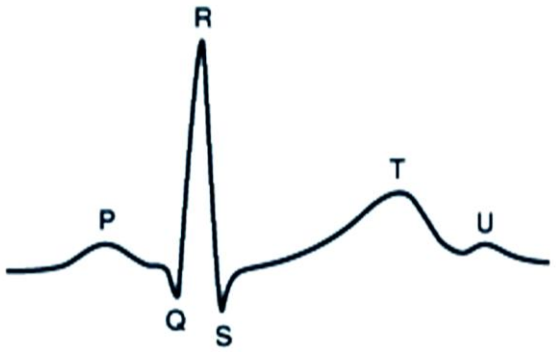

Figure 1. ECG signal for healthy human

The ECG signal is composed with P, QRS and T waves. The P wave is the first bump and normally an upward bump, it measure the atrial depolarization. The QRS complex starts with a negative deflection Q, then a large positive measurement R and next a negative movement S. The QRS complex indicates ventricular depolarization and contraction [6]. The T wave is the normally a modest upward waveform indicating the repolarization of the ventricles.

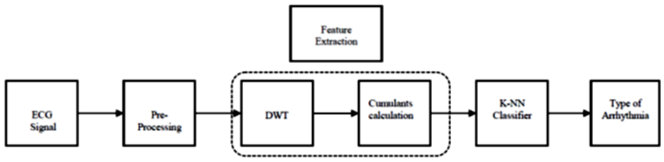

Figure 2. Proposed method for better ECG beats classification

B. **Higher Order Statistics**

As discussed earlier the ECG signal is a non-stationary signal with large data. The variations occurred in the ECG signal are in rapid manner. We can"t get the exact variations for the analysis due to the non-stationary nature of the ECG signal. HOS based methodologies are effective in suppressing such signal variations in the presence of noise [7]. Finally we are getting the signal with less noisy components these signals are used for the classification of beats.

C. Feature Extraction

The ECG signal is composed with large amount of data, for the analysis it is very difficult to get the information about the signal. So, we should compress the data as much as possible. The feature vectors are those kinds of attributes to give the information about the original signal.

a. Discrete Wavelet Transform

The Discrete Wavelet Transform is used to extract the wavelet coefficients of discrete time signals. In other methodologies like Fourier Transform we are getting the frequency related information only. But in this DWT we are getting both time and frequency domain analysis of a signal, it gives the multi-resolution analysis [8]. In the DWT analysis the signal is decomposed based on the level. In this methodology we are using the level-3 scheme. In the DWT analysis we are using the two types of filters namely Low Pass Filter (LPF) and High Pass Filter (HPF) along with the down conversion. The LPF gives the approximation coefficients and the HPF gives the detailed coefficients.

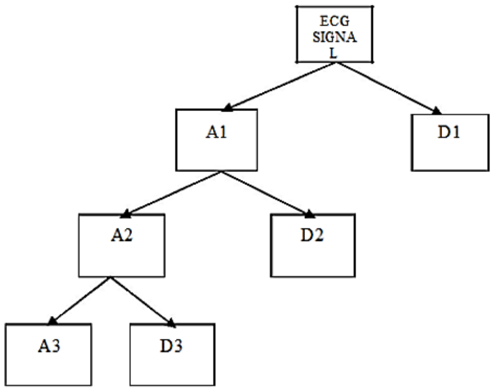

Figure 3. Discrete wavelet transformation for level-3

b. **Cumulants Calculation**

The cumulants gives the nature of the original signal. The first and second order cumulants are having the significance in the signal processing [9]. If the signal having non-linearity nature, then we can"t get the exact nature of the signal by using first and second order cumulants.That"s why in this paper we are going for third and fourth order cumulants along with the second order Cumulants.

The second moment generating function is given by

$$\phi_y(\omega) = \ln\left(\int_{-\infty}^{\infty} f_y(y)\, e^{j\omega y}\, dy \right)$$

Cumulants are the coefficients of the Taylor series expansion of the second generating function.

$$c_n = (-j)^n \left(\frac{d^n}{d\omega^n} \phi_y(\omega) \right) \text{at } \omega = 0$$

The n^{th} cumulants of the signal are calculated as

$$c_n^y(\tau_1, \tau_2, \ldots\ldots, \tau_{n-1}) = m_n^y(\tau_1, \tau_2, \ldots\ldots, \tau_{n-1}) - m_n^G(\tau_1, \tau_2, \ldots\ldots, \tau_{n-1})$$

Here, τ_i represents the time difference and m_n^y represents n^{th} moments of the random variable y

The second order cumulants defined as

$$cum(y_a, y_b) = E\{y_a y_b\}$$

The third order cumulants defined as

$$cum(y_a, y_b, y_c) = E\{y_a y_b y_c\}$$

The fourth order cumulants defined as

$$cum(y_a, y_b, y_c, y_d) = E\{y_a y_b y_c y_d\} - E\{y_a y_b\}E\{y_c y_d\} - E\{y_d y_c\}E\{y_b y_d\} - E\{y_a y_d\}E\{y_b y_c\}$$

E. K-NN classifier

The K-NN classifier is one of the most basic classifiers for data classification; at the same time it is quite simple to implement [10]. It achieves consistently high performance without a priori assumptions. In the K-NN classifier we are mainly having the variables namely sample, training and group. A new sample is classified by calculating the distance to the nearest training signal. We can extend the K-NN classifier by changing the K value; it means that increasing the number of nearest samples. There are many techniques available for improving the performance and speed of a nearest neighbor classification, one solution is choose a training data such that classification by the 1-NN rule. This can result in significant speed improvement as K can now be limited to '1' and redundant data points have been removed from the training set [11].

III. RESULT AND DISCUSSION

Precision and recall are two important parameters to demonstrate the classification performance [12]

$$precision = \frac{t_p}{t_p + f_p}$$

$$recall = \frac{t_p}{t_p + f_n},$$

Where t_p, t_n, f_n, f_p represents the true positive, true negative, false negative, false positive details.

Extracted features are partitioned into train and test data using ten-fold cross validation. Later, test data fed k-nn classifier and parameters are calculated using confusion matrix. Table 1 presents confusion matrix for one fold. This process repeated ten times with different test sets. Average results are presented in Table 2.

Table 1. Confusion matrix of one fold.

a	b	c	d	e	f	< ------- Classified as
1971	1	0	19	5	4	a = normal
2	1987	0	16	1	4	b = paced
0	0	1999	1	0	0	c = PVC
25	8	0	1781	73	113	d = APC
4	1	0	81	1913	1	e = RBBB
1	21	0	73	1	1904	f = LBBB

Table 2. Tenfold cross validation results for K=1

NATURE OF BEAT	NUMBER OF BEATS	Precision (in %)	Recall (in %)
normal	2000	0.984	0.986
paced	2000	0.985	0.994
PVC	2000	1.000	1.000
APC	2000	0.908	0.891
RBBB	2000	0.960	0.957
LBBB	2000	0.940	0.952
Weighted Average	**12000**	**0.963**	**0.963**

From the table it is evident that of our methodology shows the superiority in classification.

IV. CONCLUSION

The proposed method provides the better classification of ECG beats. In this methodology we are using three steps for the better classification of ECG beats. In the first step the data is collected from the MIT/BIH database. In the second step we are applying the DWT technique to the ECG signal, the signal is divided into detailed and approximate coefficients. The detailed coefficients having the important information, we are calculating the feature vectors for these detailed coefficients. For the classification of ECG beats we are applying these feature vectors to the K-NN classifier. In this process we are getting the Precision of 96.30%, Recall of 96.30% for 12,000 beats which are calculated from the MIT/BIH database.

V. REFERENCES

[1] Nimunkar A.J, Tompkins WJ: R-peak Detection and Signal Averaging for Simulated Stress ECG using EMD. Engineering in Medicine and Biology Society, 2007, EMBS 2007. 29th Annual International Conference of the IEEE, 22-26 2007, 1261-1264.

[2] Thakor NV, Zhu YS. "Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection," Biomedical Engineering, 1991; 38(8): 785-794.

[3] Sung-NienYu,Ying-Hsiang Chen, "Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network," periodical journal Pattern Recognition Letters, Elsevier Science Inc New York, NY, USA, Volume 28 , Issue 10, July 2007, pp. 1142-1150.
4http://www.physionet.org/physiobank/database/ mitdb/. 5http://www.biomedical-engineering-online.com/content/1/1/5.

[4] Nimunkar A.J, Tompkins WJ: R-peak Detection and Signal Averaging for Simulated Stress ECG using EMD. Engineering in Medicine and Biology Society, 2007, EMBS 2007. 29th Annual International Conference of the IEEE, 22-26 2007, 1261-1264.

[5] De Chazal P, Reilly RB: A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng 2006, 53(1):2535-2543, No. 12. 17. Engine M: ECG beat classification using neuro-fuzzy network. Elsevier Science Inc Pattern Recognition Letters 2004, 25(15):1715-1722.

[6] Nimunkar A.J, Tompkins WJ: R-peak Detection and Signal Averaging for Simulated Stress ECG using EMD. Engineering in Medicine and Biology Society, 2007, EMBS 2007. 29th Annual International Conference of the IEEE, 22-26 2007, 1261-1264.

[7] Nikias CL, Petropulu AP: Higher Order Spectra Analysis: A Nonlinear Signal Processing Framework. Prentice Hall, Englewood Cliffs, NJ 1993.

[8] G.Selvakumar, K.BoopathyBagan, "Wavelet Decomposition for Detection and Classification of Critical ECG Arrhythmias," Proceeding of the 8th WSEAS International Conference on Mathematics and computers in Biology and Chemistry, June 2007, pp. 80-84.

[9] LabibKhadra, Amjed S. Al-Fahoum, and SaedBinajjaj, "A quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques," IEEE Transactions on Biomedical Engineering, Vol. 52, No. 11, November 2005, pp. 1840-184
Raut, R.D.; Dudul, S.V. "Arrhythmias Classification with MLP Neural Network and Statistical Analysis," First IEEE International Conference on Emerging Trends in Engineering and Technology, 2008, Page(s): 553 – 558

[10] Karimifard S, Ahmadian A: Morphological Heart Arrhythmia Classification Using Hermitian Model of Higher-Order Statistics. 29th IEEE EMBS Annual International Conference 2007.

[11] Karimifard S, Ahmadian A, Khoshnevisan M: Morphological Heart Arrhythmia Detection Using Hermitian Basis Functions and kNN Classifier. 28th IEEE EMBS Annual International Conference 2006.

[12] ShivajiraoJadhav, Sanjay Nalbalwar and Ashok Ghatol, "ECG Arrhythmia Classification using Modular Neural Network Model", in Proc. 2010 IEEE EMBS Conference on Biomedical Engineering & Sciences IECBES 2010