Classification of heart diseases from ECG signals using wavelet transform and kNN classifier

Ridhi Saini*, Namita Bindal‡, Puneet Bansal§
Electronics and Communication Engineering Department
University Institute of Engineering and Technology, Kurukshetra University
Kurukshetra, India
*ridhisaini9@gmail.com, ‡namitabindal7@gmail.com, §puneet4u@gmail.com

Abstract— **Heart is the most vital organ which circulates blood along with nutrients and oxygen throughout the body. There are number of reasons which may affect its normal working. In this paper ten heart diseases, as well as normal, have been classified by extracting features from original ECG (electrocardiogram) signals and sixth level wavelet transformed ECG signals. The results have been compared and improved accuracy has been obtained using wavelet transformed signals.**

Keywords—Classification algorithm; Discrete wavelet transforms; Electrocardiography; Principal component analysis; Statistical analysis.

I. INTRODUCTION

Heart is the most sophisticated organ of human body and the diseases related to it need to be identified with much accuracy and in short time, failure of which can lead to sudden death. For diagnosis of heart, ECG signals are used. Since there is variation in physiological structures of humans, they have different shapes of ECG. Therefore, the wave shape of the same disease may vary little bit from person to person. For some diseases, the clinical symptoms along with ECG will be sufficient to diagnose the disease at early stages; whereas, many cardiac diseases develop over a long time. So, some specialised instruments are used to record the signals for a few days, such as Holter ambulatory monitor [1]. This gets cumbersome for the cardiologist to study the waves and interpret them in limited time. Nowadays, ECG interpretation with the aid of software is used by cardiologists to support their diagnosis results; but, the concept is not new [2]. Automatic diagnosing systems are a need of the day which will analyze ECG waveforms of several hours and give results in a few seconds. In this paper, ECG signals of normal and diseased subjects have been obtained from MIT-BIH arrhythmia database and MIT-BIH malignant ventricular arrhythmia database [3]. Signal analysis methods extract features of the signal such as position and amplitudes of different peaks of ECG, etc. either in time or frequency domain [4]. To better interpret the ECG signals, their morphology has been studied by [5-10]. Detection based on QRS wave shape has been done by [6], [11]. For detection of heart disease, besides historical records, physical examination also needs to be done [12]. In elderly patients, making the diagnosis is more difficult; because, of a relative absence of typical signs and symptoms and the possibility of attributing heart failure symptoms to other diseases [13].

II. METHODOLOGY

The objective of the paper is to offer an ECG diagnosing technique to improve the performance of the diagnosis of patients with heart diseases and distinguish them from the normal subjects on the basis of their ECG waveforms especially for diagnosing the people in remote villages where expert medical facilities are not available or are far to reach within specific time. The techniques used comprise of the possible use of wavelet entropy based measures in improving the feature extraction from ECG signals and the effects of Principal Component analysis and statistical parameters on analysis of ECG signals. Finally k-nearest neighbour (kNN) classifier has been used to classify the diseases. Though many classification techniques are available such as artificial neural network (ANN) and support vector machine (SVM) classifiers, the structure of the kNN classifier imposes lower computational burden [14]. The MIT-BIH arrhythmia database consists of forty eight half-hour excerpts of two-channel ambulatory ECG recordings, obtained from forty seven subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979, who suffered from right bundle branch block, left bundle branch block, paced rhythm, atrial premature beat and also of normal patients [3]. The MIT-BIH malignant ventricular arrhythmia database contains twenty-two half-hour ECG recordings of subjects who suffered from ventricular flutter, sustained ventricular tachycardia, ventricular bigeminy, ventricular fibrillation, high grade ventricular ectopic activity and asystole. The discrete data contains a variable, 'val', having 2 rows (signals) and 525000 columns (samples/signal) and duration of 35:00 min with sampling frequency of 250 Hz [3]. All the signals were divided in duration of five minutes; because, normally these signals have been recorded by the practitioner for this much time only. The statistical features computed were kurtosis, skewness, entropy, mean, standard deviation, variance, root mean square, minimum, maximum, range, harmonic mean, median and energy. These features were extracted from ECG waveforms to interpret them more precisely and accurately. These functions were implemented on waveforms with the help of MATLAB software [15]. Two methods were used to classify the ECG signals. In first method, the statistical features were computed directly from raw ECG signals and then principal component analysis was

ISBN:978-1-4799-8890-7/15/$31.00 ©2015 IEEE

applied before classification. In second method, the ECG signal were decomposed using wavelet transform with Daubechies order four wavelet up to level six. Then wavelet coefficients were extracted at detailed level four. The statistical features were extracted using wavelet coefficients which were then classified using kNN classifier after application of principal component analysis. The algorithms used for these two methods are:

A. Method 1: Algorithm for feature extraction and classification using original ECG signals

1. Divide the signal into intervals of 5 min
2. Extract statistical features
3. Normalize the data for zero mean and unity standard deviation
4. Apply Principal component analysis
5. Normalize to make data range from -1 to +1
6. Divide the data for training (95%) and testing (5%)
7. Train kNN classifier on training data
8. Calculate percentage error for training data
9. Test testing data using kNN classifier trained in step 7
10. Calculate percentage error for testing data

B. Method 2: Algorithm for feature extraction and classification using wavelet transformed ECG signals

The algorithm steps are same as that of method 1 with addition of following two steps after step 1

i. Decompose the ECG signal using wavelet transform using Daubechies order 4 wavelet up to level 6

ii. Extract the wavelet coefficients at detailed level 4

Fig. 1 and Fig. 2 show that the extraction at level 6 (approximation and details) contain more details and db4 is better than db5 and other wavelets as its shape is more similar to the used ECG signal. Also the waveform at detailed level 4 in level 6 extractions is quite similar to the original waveform with some improvements after removal of noise which is a result of passing the waveform through a series of high and low pass filters. This de-noising has resulted in selecting level 6 extraction using db4 wavelet and from this the wavelet detailed coefficients have been extracted at level 4.

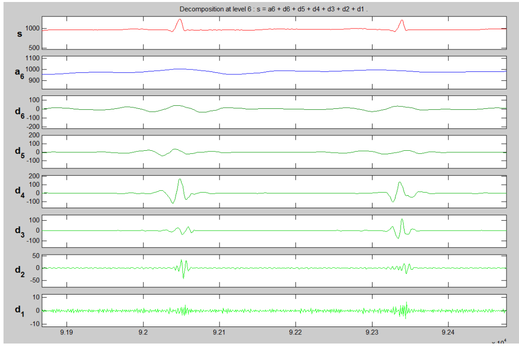


Fig. 1. Feature extraction using db4 wavelet at level 6 for ECG waveform of a normal person.

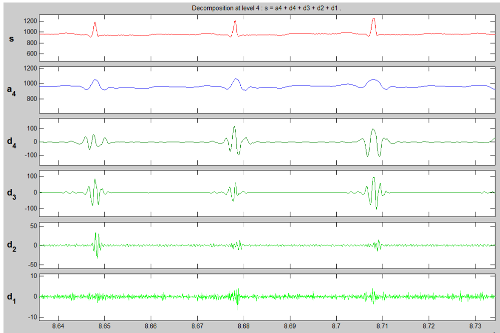


Fig. 2. Feature extraction using db5 wavelet at level 4 for ECG waveform of a normal person.

The following figures (Fig. 3 to Fig. 12) show the feature extraction using db4 wavelet at level 6 for ECG waveforms of diseased patients with right bundle branch block, left bundle branch block, paced rhythm, atrial premature beat, Ventricular tachycardia, Ventricular flutter, Ventricular bigeminy, Ventricular fibrillation, high grade ventricular ectopic activity and Asystole.

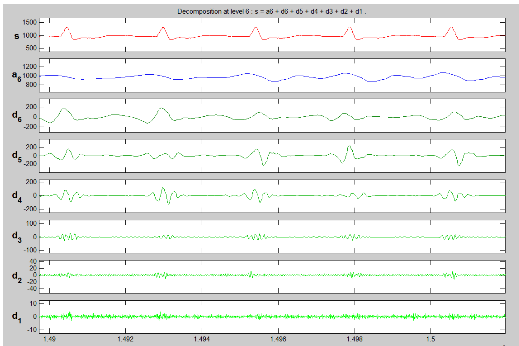


Fig. 3. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Left bundle branch block.

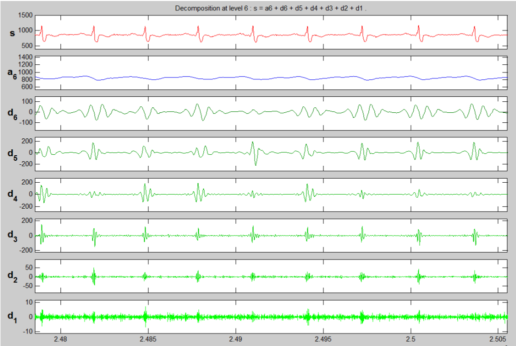


Fig. 4. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Right bundle branch block.

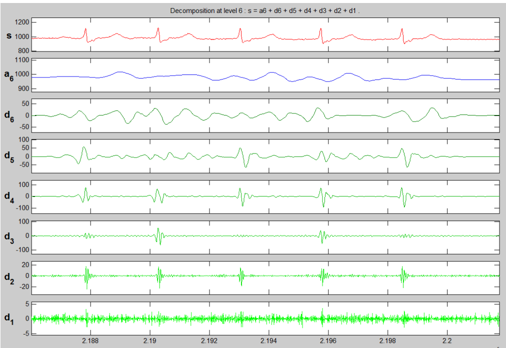


Fig. 5. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Atrial premature beat.

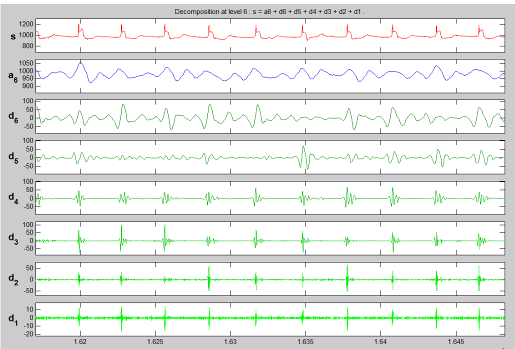


Fig. 6. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with paced rhythm.

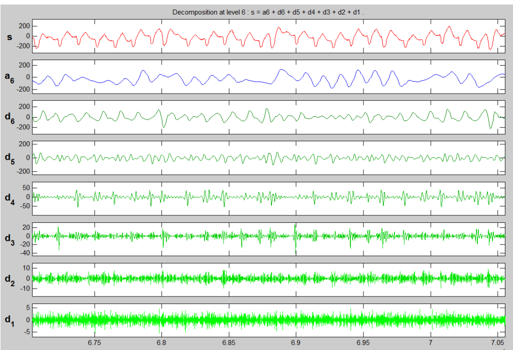


Fig. 7. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Ventricular flutter.

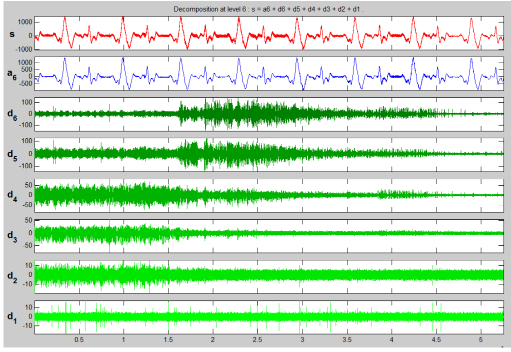


Fig. 8. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Ventricular tachycardia.

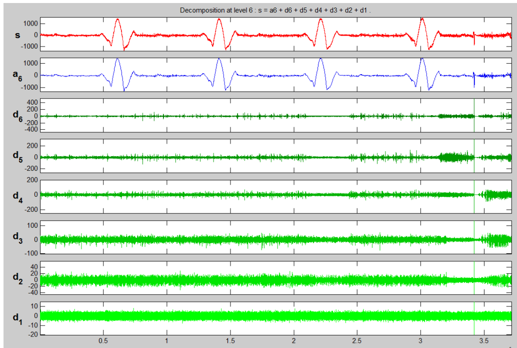


Fig. 9. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Ventricular fibrillation.

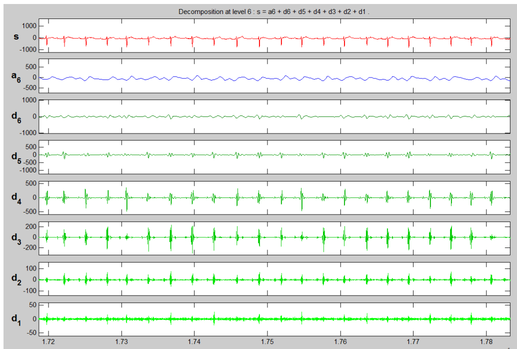


Fig. 10. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Ventricular bigeminy.

Fig. 11. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with Asystole.

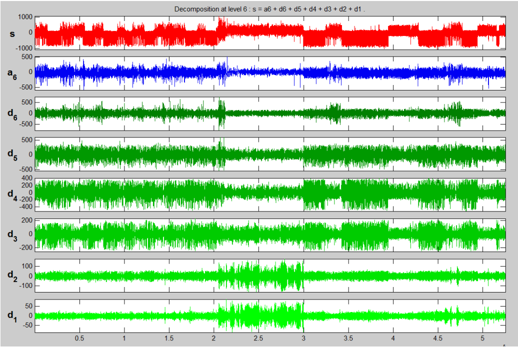
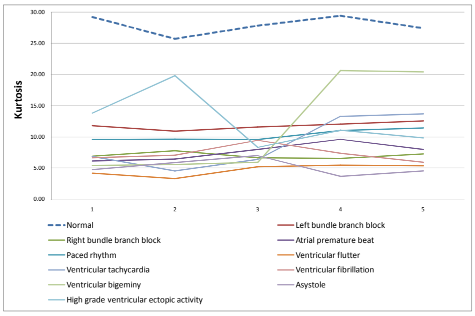


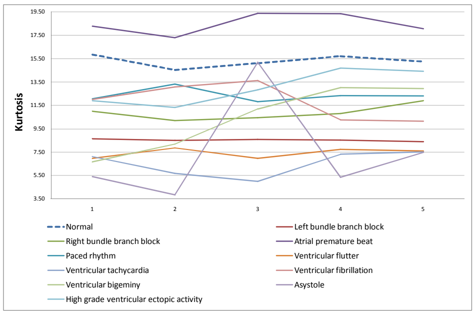
Fig. 12. Feature extraction using db4 wavelet at level 6 for ECG waveform of a patient with High grade ventricular ectopic activity.

III. Results

The features extracted with and without wavelet transform has been plotted in Fig. 13 to Fig. 25 (for some data).

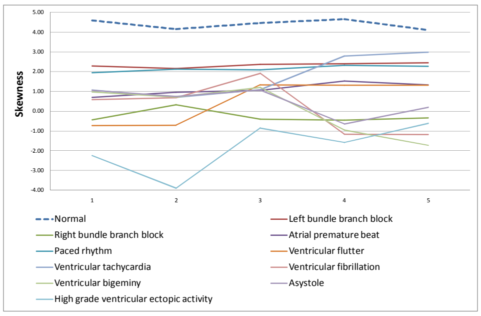


(a)

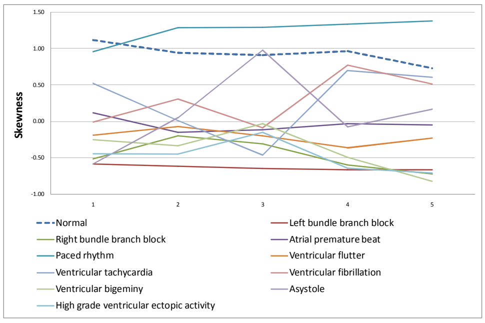


(b)

Fig. 13. Variations in kurtosis for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

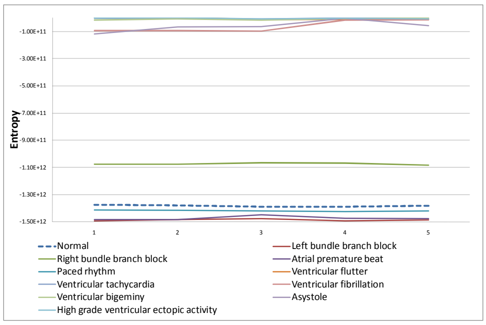


(a)

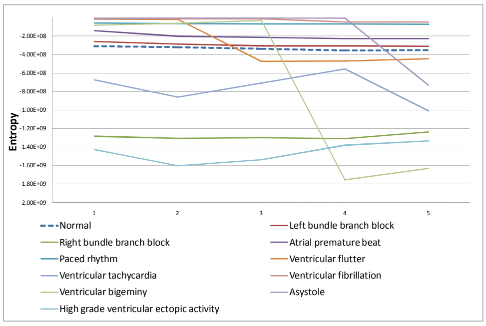


(b)

Fig. 14. Variations in skewness for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

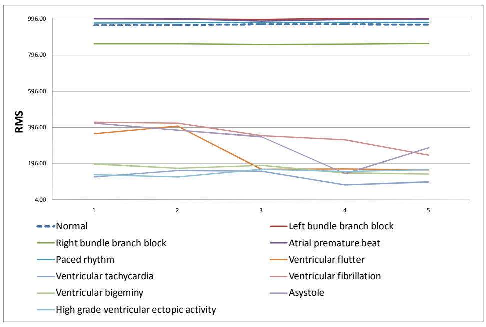


(a)

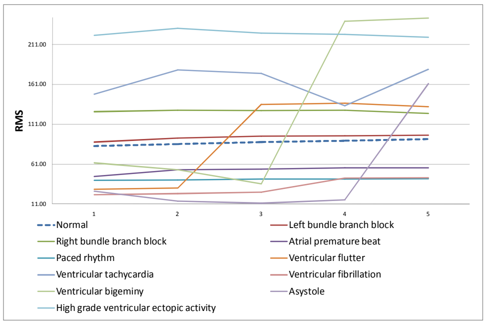


(b)

Fig. 15. Variations in entropy for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

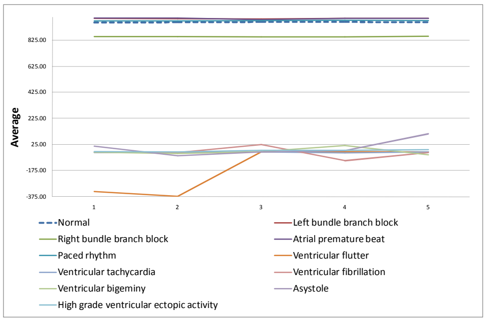


(a)

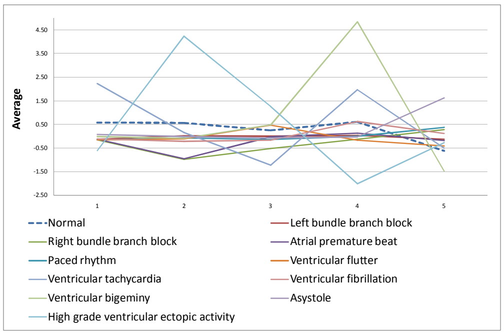


(b)

Fig. 16. Variations in root mean square for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

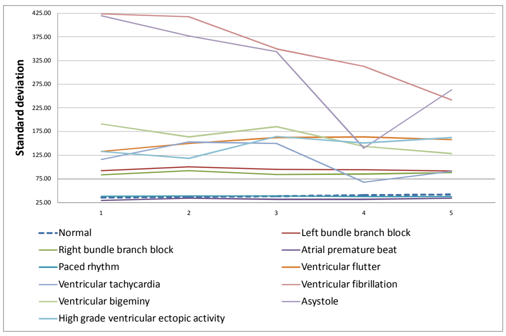


(a)

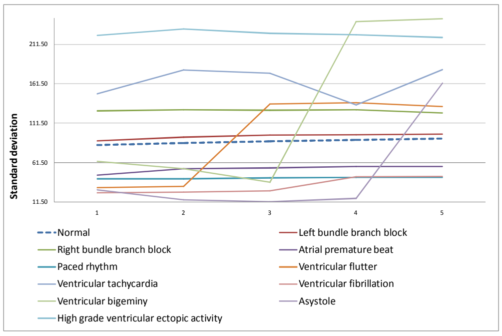


(b)

Fig. 17. Variations in average for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

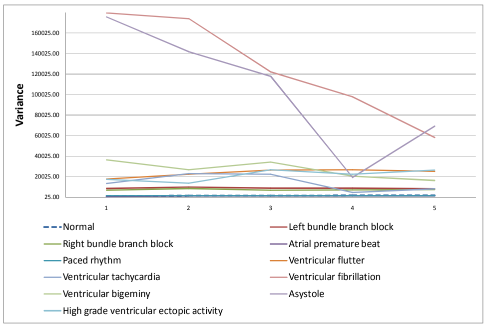


(a)

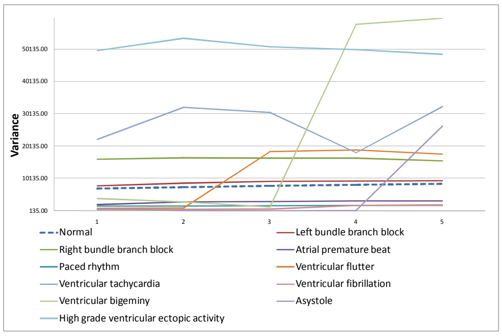


(b)

Fig. 18. Variations in standard deviation for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

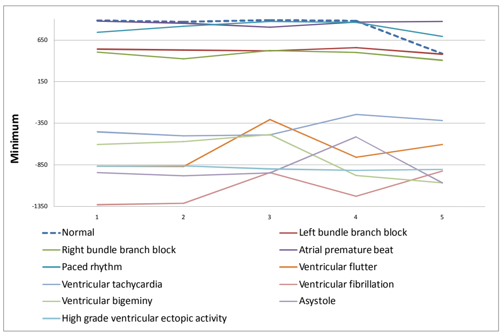


(a)

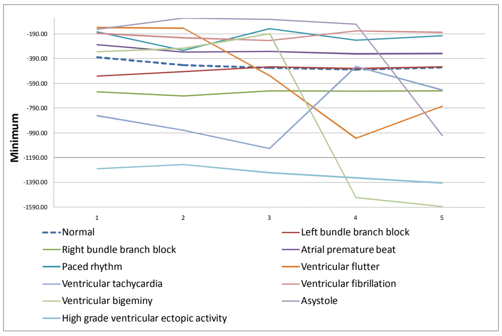


(b)

Fig. 19. Variations in variance for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

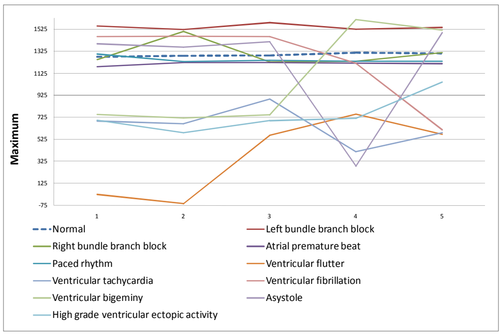


(a)

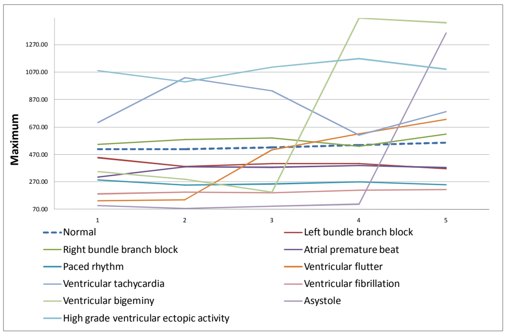


(b)

Fig. 20. Variations in minimum for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

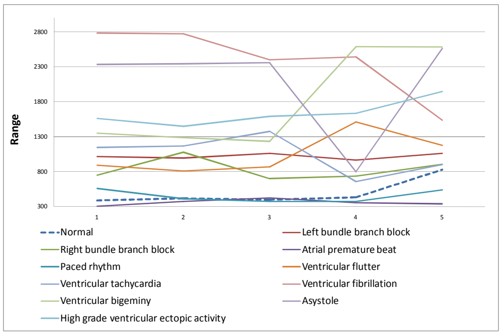


(a)

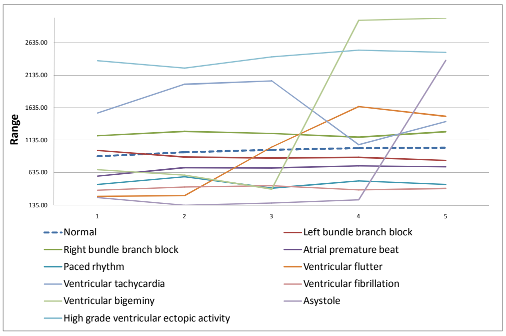


(b)

Fig. 21. Variations in maximum for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

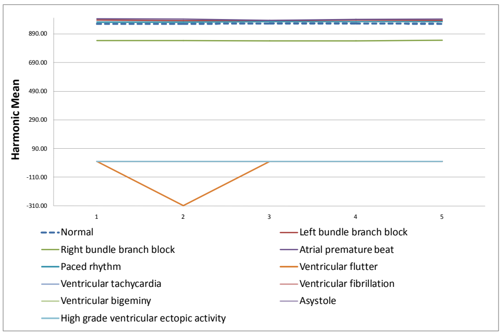


(a)

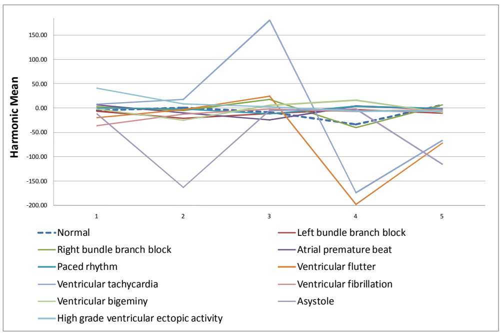


(b)

Fig. 22. Variations in range for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

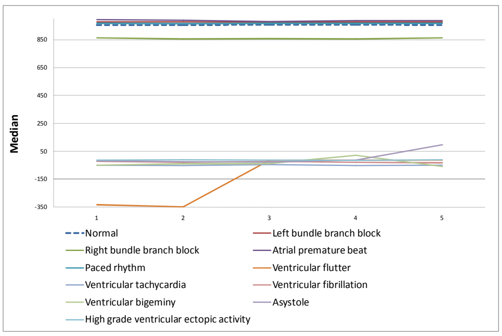


(a)

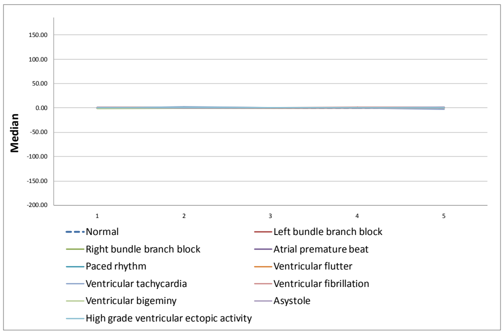


(b)

Fig. 23. Variations in harmonic mean for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.



(a)



(b)

Fig. 24. Variations in median for some ECG signals (a) without wavelet transformation, (b) with wavelet transformation.

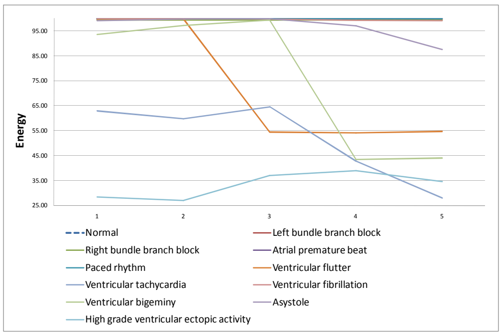


Fig. 25. Variations in energy for some ECG signals with wavelet transformation.

Twelve statistical features from raw ECG signals have been evaluated. The training data consisted of 372 patterns, thus making the training matrix of size 372 x 12. Principal component analysis has been used to reduce the training matrix by extracting orthogonal components from input features, thus reducing the training matrix to 372 x 8. This reduces memory requirement by 38.46 % as well as will increase the speed of executing the algorithm. The resulting orthogonal feature matrix has been classified using kNN classifier. Same procedure was followed for testing matrix. Classification efficiency of 56.25 % has been achieved. To improve the efficiency, the ECG signal was decomposed using wavelet transform and thirteen (including energy feature) statistical features were evaluated from these decomposed signals. After classification of statistical features of decomposed ECG signal, a classification efficiency of 87.5 % has been obtained which is an improvement of 31.25 %. An accuracy of 31.25 % was achieved using ANN with 50-50 neurons in two hidden layers trained by Levenberg-Marquardt back-propagation algorithm. This showed the supremacy of kNN over ANN.

IV. Conclusion

To help medical practitioners diagnose subject's heart diseases, ECG signal can be interpreted by software algorithms in the direction of achieving automated heart disease diagnosing system. In this paper classification is done from original ECG signals and wavelet transformed ECG signals. Classification efficiency of 87.5% has been obtained using wavelet transformed ECG signals. Classification is done using kNN classifier. As a future scope of this paper, an automated hardware system can be developed which may record and diagnose heart disease automatically and can be used at remote places where it is difficult for medical practitioners to service the people. Also other classifiers like Support vector machines, Gaussian Mixture Model, Hidden markov model can be used for classification and their performance can be compared with kNN.

References

[1] W.J. Tompkins, Biomedical Digital Signal Processing. Prentice Hall, Upper Saddle River, New Jersey, 2000.

[2] P.W. Macfarlane, "Recent developments in computer analysis of ECGs," Clinical Physiology and Functional Imaging, vol 12, no 3, pp. 313–317, 1992.

[3] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, and H.E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet, "Components of a new research resource for complex physiologic signals," Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; June 13, 2000.

[4] V.X. Afonoso, and W.J. Tompkins, "Detecting ventricular fibrillation: Selecting the appropriate time-frequency analysis tool for the application," IEEE Eng. in Medicine and Biology, pp. 152-159, 1995.

[5] S. Osowski, L.T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," IEEE Transactions on Biomedical Engineering, vol 51, no 4, pp. 582–589, 2004.

[6] B.U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection: Reviewing and comparing algorithms for detecting this important ECG waveform," IEEE Engineering in Medicine and Biology Magazine, vol 21, no 1, pp. 42–57, 2002.

[7] P.S. Gokhale, "ECG signal de-noising using discrete wavelet transform for removal of 50Hz PLI noise," International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, vol 2, Issue 5,pp 81-85, May 2012.

[8] S.S. Mehta, and N.S. Lingayat, "Biomedical signal processing using SVM," IET-UK International Conference on Information and Communication Technology in Electrical Sciences (ICTES 2007), Dr. M.G.R. University, Chennai, Tamil Nadu, India, pp.527-532, Dec. 20-22, 2007.

[9] Abed Al Raoof Bsoul, Kevin Ward, Kayvan Najarian, and Soo-Yeon Ji, "A unified signal processing and machine learning method for Detection of abnormal heart beats using electrocardiogram," IEEE International Conference on Bioinformatics and Biomedicine Workshops, pp. 453-460, 2010.

[10] Meng-Chou Chang, Zong-Xin Lin, Che-Wei Chang, Hsiao-Lung Chan, and Wu-Shiung Feng, "Design of a system-on-chip for ecg signal processing," the 2004 IEEE Asia-Pacific Conference on Circuits and Systems, pp. 441-444, Dec 6-9, 2004.

[11] O Pahlm, and L. Sornmo, "Software QRS detection in ambulatory monitoring – a review," Medical & Biological Engineering & Computing, vol 22, pp.289–297, 1984.

[12] J. Wilbur, and P. James, "Diagnosis and management of heart failure in the outpatient setting," Primary Care: Clinics in Office Practice, vol 32, pp. 1115–1129, 2005.

[13] N.D. Gillespie, "The diagnosis and management of chronic heart failure in the older patient," British Medical Bulletin, vol 75–76, no 1, pp. 49–62, 2006.
[14] C.M. Bishop, Pattern Recognition and Machine Learning, Springer publishing, 2006.
[15] MATLAB Wavelet toolbox user guide, ver. 2012a.