Journal of Advanced Biomedical Sciences

Journal of Advanced Biomedical Sciences

Homepage: http://jabs.fums.ac.ir

Online ISSN: 2783-1523

Fasa University of Medical Sciences

Intelligent Diagnosis of Heart Diseases Based on Electrocardiographic Signal

Hosseinpoor Mohammadjavad[1, 2]

1. Department of Computer Engineering, Estahban Branch, Islamic Azad University, Estahban, Iran
2. Member of Young Researchers and Elite Club, Estahban Branch, Islamic Azad University, Estahban, Iran

Article Info

Article Type:
Original Article

Article History:
Received
01 Oct 2023
Received in revised form
28Oct 2023
Accepted
15 Nov 2023
Published online
15 Jan 2024

Publisher:
Fasa University of Medical Sciences

Abstract

Background & Objectives: Cardiovascular disease is a leading cause of death worldwide. ECG signals are used to diagnose it. This study aims to eliminate signal noise by converting available wavelets and extracting existing waves. The location-related properties and amplitude of these waves will be extracted to develop a model based on the random forest algorithm for training and evaluating the algorithm.

Materials & Methods: This study uses the MIT-BIH dataset, which contains digital ECG signals extracted from Holter bands for different patients at Arrhythmia Hospital from 1975 to 1979. The study applies signal processing and machine learning techniques to classify ECG signals and identify heart patients. The MATLAB software implemented the algorithm, which was evaluated based on accuracy, error rate, TP, FP, Precision, Recall, F-Measure, and ROC criteria. These criteria were determined by a confusion matrix.

Results: The study results and comparisons demonstrate that the proposed method is highly effective in detecting heart patients. The proposed method's accuracy was found to be 99%, which is higher than other machine learning methods.

Conclusion: The proposed method achieved an accuracy of 99.1957%, surpassing other machine learning methods like support vector machine, neural network, and Bayes.

Keywords: Heart disease, MIT-BIH dataset, Random Forest algorithm, Wavelet transform

Cite this article: Hosseinpoor M. Intelligent Diagnosis of Heart Diseases Based on Electrocardiograph Signal. JABS.2024; 14(1): 47-55.

DOI: 10.18502/jabs.v14i1.14806

Introduction

The heart is a vital organ in the body, and cardiovascular diseases (CVD) are a leading cause of death. To prevent deaths caused by CVD, doctors use tools and methods such as ECG. Electrocardiography is a graphic representation that shows the electrical behavior of the heart muscles during contraction and relaxation (1). It can help diagnose heart irregularities. However, analyzing electrocardiography traditionally can be complicated and time-consuming for doctors. Expertise is required for accurate signal analysis (2). Due to these challenges and the large volume of electrocardiography data, computer-based methods have proven highly effective in diagnosing heart abnormalities (3-5). Long-term electrocardiograph records, each comprising

Corresponding Author: Hosseinpoor Mohammadjavad, Department of Computer Engineering, Estahban Branch, Islamic Azad University, Estahban, Iran
Email: mj.hosseinpoor@iau.ac.ir

100,000 beats, can be challenging to analyze and prone to human error. Therefore, machine learning methods have been used for automatic diagnosis, resulting in good efficiency and accuracy (6).

Electrocardiography data is produced by electrodes placed on the skin and chest. Each electrode's output is called a lead, and together they form electrocardiographic signals, including P, Q, R, S, and T waves (7-9). Identifying significant elements from raw electrocardiography data, which may appear irregular and unstructured at first glance, can be challenging. Researchers have proposed several methods for classification and pattern recognition to overcome this challenge. Typically, these methods involve preprocessing the raw electrocardiography data and extracting features from the processed data. Feature extraction methods include wavelet transform, DCT discrete cosine transform, and other techniques. Machine learning algorithms, such as neural networks (NN), K-NN, and decision trees, are used to classify extracted features (10).

In recent years, several studies have focused on diagnosing coronary artery disease (CAD) using ECG signals. Adyasha et al. developed four deep learning (DL) models, including Autoencoder (AE), Radial Basis Function Network (RBFN), Self-Organizing Map (SOM), and Restricted Boltzmann Machine (RBM). In their article, they used two publicly available arrhythmia datasets. The authors used PTB-ECG and MIT-BIH datasets to train and validate their proposed models. They also created an ensemble classification model by combining AE and SOM models using the majority voting principle, which achieved the best performance. Simulation-based experiments using the two standard datasets demonstrated that the AE model provided high accuracy, F1 score, and area under the curve (AUC) values. Specifically, for the MIT-BIH dataset, the AE model achieved accuracy, F1 score, and AUC values of 0.974, 0.932, and 0.922, respectively. For the PTB-ECG dataset, the AE model achieved accuracy, F1 score, and AUC values of 0.984, 0.967, and 0.932, respectively. The AE model outperformed the other three models (11).

Chazal et al. used perceptron multilayer neural networks with different numbers of layers, basic radial function neural network, and probabilistic neural network (PNN). The results of this study showed the highest accuracy of 97.14%, using the MIT-BIH arrhythmia dataset (12). Guixiang et al. proposed a new method in their article for detecting CW (Cardiac Waveform) and the onset and offset of ECG (Electrocardiogram). The proposed method consists of three stages. The signal is filtered using Wavelet Transform (WT). Then, CW is detected using a combined WT operation with adaptive thresholding in a fixed window. The onset and offset of CW are identified using a method based on maximizing the slope of the region with peak adoption. In the last two stages, a search window and threshold correction are performed. The MIT-BIH arrhythmia database and QT database were used to evaluate sensitivity (Se), positive predictivity (PP), accuracy (Acc), and computation time (CT). CW peak detection achieved Se%, PP%, and Acc% values of >= 98.64, 98.64, 97.30, respectively, with a computation time of <= 5.77 seconds for a 15-minute ECG. Onset and offset detection achieved values of >= 98.48, 98.48, 97.84, 97.94, and 97.00, respectively. The method successfully extracts CW and their onsets, particularly for the P-wave. These features are used to obtain clinical diagnosis parameters such as heart rate, amplitude, and duration. This method provides a foundation for cardiac disease diagnosis and serves as a basis for intelligent diagnostic systems.

Abo-zahhad et al. used neural networks based on electrocardiography signals to diagnose ischemic heart disease in healthy individuals. Their approach utilized a feed-forward multilayer (FFMLP) neural network and the error backpropagation learning algorithm (14). Manimegalai et al. employed the discrete wavelet transform to extract P, Q, R, S, and T waves and differentiate normal electrocardiography samples from abnormal ones (15). Raman and his colleagues classified electrocardiography signals using a support

[DOI: 10.18502/jabs.v14i1.14806]

Journal of Advanced Biomedical Sciences

Fasa University of Medical Sciences

vector machine. They used features such as QRS, R peak, time interval, gradient, and polarity during the feature extraction stage. The result of the principal component analysis (PCA) method was used as input to the support vector machine. They reported a final accuracy of 90% for this method (16). Kumar et al. used a combination of random forest and discrete cosine transform (DCT) methods to classify heart disease. They first calculated the DCT of the RR interval and used it as an input feature for the random forest algorithm. The method was tested on the MIT/BIH dataset, and the final accuracy reported was 92.16% for 30 trees (best result) (17). Our research aims to classify electrocardiograph data using therandom forest method. We will also implement the discrete wavelet transform to extract features.

Materials and Methods

This section presents a method for classifying electrocardiograph signals. The method consists of several steps, including feature extraction, training, and model testing. To classify the signals, we will use the random forest algorithm. Below, we will provide details on each step. The proposed method's general flowchart includes loading the dataset, preprocessing the dataset, extracting signal features, training the model with the random forest algorithm, and evaluating the model. Figure 1 provides an overview of the steps involved. The first step is to load the data sets that contain time signals. The next step is to preprocess the signal, which involves removing background deviation and signal noise.

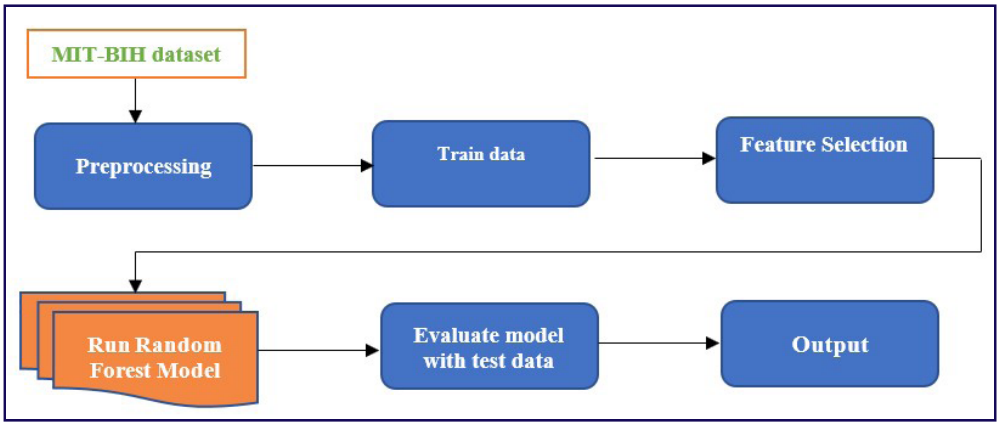

Figure 1. Flowchart of the proposed method

The signal's characteristics are extracted after pre-processing. These include pulse length, position of ECG waves, and length of each wave. The model is then trained using these features, combined with signal type, through random forest. Test signals are evaluated to assess the model's performance. The proposed method consists of the following general steps:

1- Loading the dataset.
2- Preprocessing the dataset.
3- Feature extraction from the signal.
4- Training the model using the Random Forest algorithm.
5- Evaluating the model.

The first step is to load the dataset containing signals over time. Then, preprocess the signals by removing background noise and denoising them. After preprocessing, extract features such as pulse duration, ECG wave positions, and duration of each wave. Finally, train the model using these features and the signal type, based on Random Forest. Evaluate the model on the test signals. More details on these steps will be explained.

Removing baseline drift and signal noise

The ECG signal requires preprocessing. We will address power cable noise and baseline drift using

[DOI: 10.18502/jabs.v14i1.14806]

a notch filter and discrete wavelet transform, respectively. Additionally, we will use discrete wavelet transform to remove baseline drift in ECG signals. Baseline drift indicates signal deviation from the horizontal line (0 volts). At second 30 in the figure below, the drift reaches its maximum value of approximately 0.5 volts. Removing this drift significantly improves the detection of waves in electrocardiography signals. This research utilizes discrete wavelet transform to remove baseline drift from the electrocardiography (ECG) signal. To achieve this, we begin by breaking down the signal into its detail and approximation components. By removing certain components, we can eliminate the drift. The discrete wavelet transform decomposes the original signal into approximation and detail coefficients. This process is repeated at each subsequent level until the desired number of levels is reached.

Detection of waves in ECG signals

To detect waves in ECG signals, specific features such as the P wave, QRS complex, and T wave must be identified. These waves represent different electrical activities in the heart during each cardiac cycle. A general approach to detecting these waves in ECG signals is as follows:

1. Preprocessing: As mentioned earlier, preprocess the ECG signal to remove noise and baseline drift using techniques like wavelet transform or filtering.

2. Peak Detection:

- P Wave Detection: Identify the first prominent positive peak after a period of relatively lower amplitude as the P wave.
- QRS Complex Detection: Locate the Q and S waves as the negative peaks preceding and following the largest positive peak (R wave) within a certain window.
- T Wave Detection: Find the first prominent positive peak following the QRS complex as the T wave.

3. Segmentation: Once the peaks are detected, segment the signal into individual waves using the detected peak locations.

4. Feature Extraction: Extract relevant features from each segmented wave, such as duration, amplitude, slope, and area under the curve.

5. Classification: Classify each segment based on the extracted features to determine the type of wave (P, QRS, or T).

6. Post-processing: Refine the detection results if necessary, considering physiological constraints and possible artifacts.

7. Evaluation: Assess the accuracy of wave detection by comparing the detected waves with annotated ground truth or clinical standards.

Results

This section presents the data set used and the results of the proposed method's different stages on ECG signals. It also provides a detailed description of the results obtained by applying the random forest method to the data set.

Dataset used

The research uses the MIT-BIH dataset, a commonly used source for classification purposes, which contains ECG information. This dataset consists of digitized ECG signals from Holter strips of various patients at the Arrhythmia Hospital between 1975 and 1979. Out of the 4,000 records of Holter tapes, 48 were interpreted and classified into two groups. The dataset consists of two groups: the first group has 23 records with examples of heart failure, while the second group has 25 records with examples of complex failures. The dataset was extracted from 25 men aged between 32 and 89 years old and 22 women aged between 22 and 89 years old. The signals are sampled at a frequency of 360 Hz and are 30 minutes long. The signal range is [2047, 0], where 1024 represents 0 volts. The MIT-BIH dataset is suitable for classification tasks because it includes patients of various ages and physical conditions.

[DOI: 10.18502/jabs.v14i1.14806]

Preprocess ECG signals

We used two slot filter methods to remove power cable noise and background deviation. First, we applied a slot filter to the initial signal to eliminate the power cable noise. Then, we analyzed the signal using 6db wavelet transformation up to 10 levels. Finally, we removed the approximate signal of the last level to successfully eliminate the background deviation.

To evaluate the proposed method, we will first separate the different heartbeats from the dataset and extract their characteristics as features. These features will serve as input for the dataset, while the output will be the heartbeat label. The dataset will have two classes: normal and non-normal. The non-normal class will be further categorized into four sub-classes, including Normal (N), Paced (P), LBBB (LB), and RBBB (RB). To create the dataset, we will select records 100, 105, 107, 217, 111, 214, 118, and 212. Table 1 outlines the characteristics and respective class designation of these records.

Table 1. Records used and their class

Description	Number of beats used	Records used	Class name
Normal (N)	212	100,105	1
Paced (P)	174	107, 217	2
LBBB (LB)	125	111, 214	3
RBBB (RB)	235	118, 211	4

We evaluated the proposed method using the K-Fold method with a value of 10 for K. Figure 2 presents the results of the model predictions. Out of 746 records, six were predicted incorrectly, while 740 records were predicted accurately.

		Predict class			
		N	P	LB	RB
Actual class	N	211	0	1	0
	P	0	171	1	2
	LB	0	0	123	2
	RB	0	0	0	235

Figure 2. Confusion matrix resulting from random forest model

Table 2 provides the calculated accuracy, error, TP, FP, precision, recall, F-Measure, and ROC parameters based on the confusion matrix.

[DOI: 10.18502/jabs.v14i1.14806]

Table 2. Parameters calculated based on the confusion matrix

Value	Parameter
99.19%	accuracy
0.8%	Error
99.2%	TP Rate
0.3%	FP Rate
99.2%	Precision
99.2%	Recall
98.9%	F-Measure
1	ROC Area

The results show that the Random Forest method had significantly higher accuracy compared to the SVM, Naïve Bayes, and MLP methods when implemented on the target dataset. The results show that the Random Forest method had significantly higher accuracy compared to the SVM, Naïve Bayes, and MLP methods when implemented on the target dataset. Chart 1 and 2 demonstrate this. Additionally, the Random Forest method had the lowest prediction error.

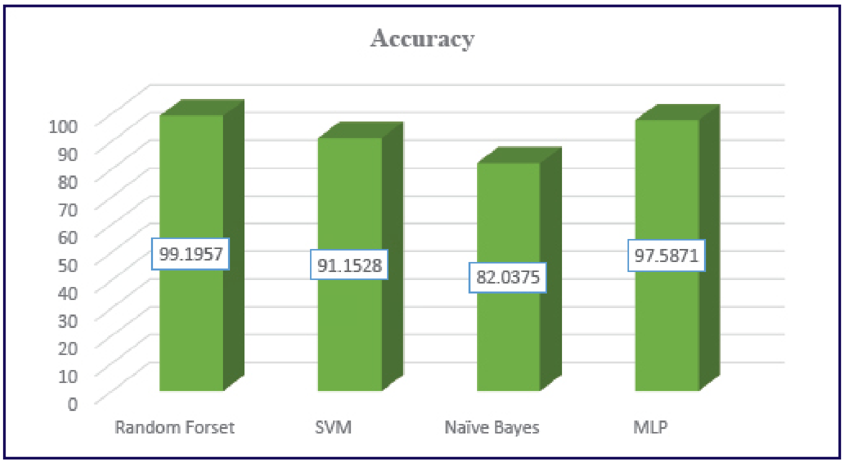

Chart 1. Comparison of random forest prediction accuracy with other machine learning methods

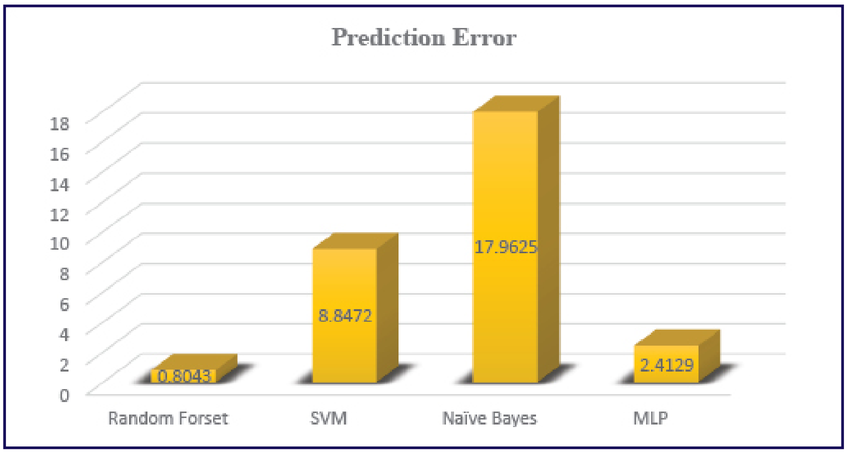

Chart 2. Comparison of prediction error of random forest compared to other machine learning methods

[DOI: 10.18502/jabs.v14i1.14806]

Discussions

Cardiovascular diseases are still a leading cause of death worldwide. They are often diagnosed using electrocardiogram (ECG) signals, which reveal the heart's electrical activity during contraction and expansion. The cardiac muscles' activity is shown through graphic diagrams that display distinct waves, including P, Q, R, S, and T. Healthcare providers use these diagrams to diagnose heart failure. Machine learning and signal processing methods can help extract and classify ECG signal characteristics, leading to more accurate signal categorization and ultimately improved diagnosis.

We used signal processing and machine learning to classify ECG signals. First, we pre-processed the ECG records by removing noise, such as background deviations, using discrete wavelet transformation. Then, we extracted the waves of the ECG signal and created features based on their position and amplitude to create a structured dataset. We used the random forest method to classify the structured dataset. The accuracy achieved was 99.19% using the K-Fold method with K equal to 10 for evaluation. This method outperforms other machine learning methods, such as support vector machine, neural network, and simple Bayes. Table 3 compares the proposed method with other methods using the accuracy metric. Table 3 shows that Chazal et al (12). achieved 97.14% accuracy using a multi-layer perceptron neural network. Raman and Gash (16) also achieved 90% accuracy by classifying electrocardiograph signals with the support vector machine method. Kumar et al (17). Combined the random forest and discrete cosine transform methods to diagnose heart disease on the MIT/BIH dataset with 92.16% accuracy. Our proposed method is more efficient than similar approaches in diagnosing heart diseases.

Table 3. Comparison of previous methods and the proposed method

Algorithm	Accuracy
Chazal Method	97.14
Raman Method	90
Kumar Method	92.16
Proposed Method	99.19

Various methods can be used to extract existing waves in ECG signals, extract their features, remove noise, and classify signal types. Future work can focus on these suggestions. To accomplish this, it is suggested to utilize the spatial position of ECG signal waves and their amplitudes. For future work, consider incorporating not only the spatial position of the waves but also statistical parameters such as standard deviation and means into the structured dataset fields. To enhance the performance and accuracy of the Random Forest algorithm used in this study, consider combining it with methods like Bagging and Boosting. However, it is important to note that while these methods can improve accuracy, they may also decrease speed. This approach can be useful in applications where speed is not a top priority. Techniques such as Principal Component Analysis (PCA) or selecting subsets of features using evolutionary algorithms like Genetic Algorithms can be employed to examine the performance of this method when utilizing a reduced set of features. Algorithms like Genetic Algorithms can be employed to examine the performance of this method when utilizing a reduced set of features.

Conclusion

The research shows that the proposed method is highly effective in detecting cardiac patients with an accuracy rate of 99%, surpassing previous methods. Additionally, the proposed model is easy to use in any medical center for diagnosing heart patients.

Acknowledgment

The author expresses gratitude to the Islamic Azad University, Estehban branch for their cooperation and assistance during the research process. The study's information has been registered with the ethics code IR.IAUESTAHBAN.REC.1401.123 by the Research Ethics Committee of the Islamic Azad University of Estehban Branch.

Conflict of Interests

No conflicts of interest have been reported by the authors.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Ethical Considerations

Any exploitation of this research is allowed with the permission of the author.

Code of Ethics

The upcoming research has been registered under the ethics code IR.IAUESTAHBAN.REC.1401.123 with the research of Islamic Azad University, Estehban branch.

Authors' Contributions

All the research implementation and analysis of the results and writing of the article were done by Mohammad Javad Hosseinpour.

Data Availability Statement

The dataset is available on the Keggel website according to the link below:

https://www.kaggle.com/datasets/mondejar/mitbih-database

References

1.Kannan R,Vasanthi V. Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease. Soft Computing and Medical Bioinformatics Journal (SCMBJ). 2019; 63-2.

2.Yu-Xia G, Ying A, Feng-Y G, Wei-Bai P, Jian-Xin W. Intelligent Electrocardiogram Analysis in Medicine: Data, Methods, and Applications. Chinese Medical Sciences J. 2023; 38(1):38-4.

3.Xieqi C, Shubin Z, Lele P, Qianwen Z, Liu H. A novel method based on shifted rank-1 reconstruction for removing EMG artifacts in ECG signals. Biomedical Signal Processing and Control. 2023; 85(1).63-9.

4.Kalyakulina A I, Yusipov VA, Moskalenko M, Nikolskiy A V, Kozlov A A, Zolotykh N Y, et al. Finding Morphology Points of Electrocardiographic-Signal Waves Using Wavelet Analysis. Radiophysics and Quantum Electronics. 2019; 61(8): 689–3.

5.Amelia EM, Andrew W, Alexandra LF, Michelle M, Gregory J, Hanoch P, et al. Mental health diagnoses in hospitalized adults with congenital heart disease in the state of Texas: A 10-year review. I J of Cardiology Congenital Heart Disease. 2022;7:100340.

6.Addison G, Nicholas D, Pei-Ni J. Tificial intelligence in echocardiography to diagnose congenital heart disease and fetal echocardiography. Intelligence-Based Medicine. 2022; 6:100082.

7.Paulino AA, Alexandros B, Aaqib HM. Frequency and Impact of Infectious Disease Conditions Diagnosed During Decompensated Heart Failure Hospitalizations in the United States. The American J of Cardiology. 2023; 191:1-7.

8.Jason C, Geoff S, David B, Rachael C, Leeanne G, David SC, Calum N. Algorithmic complexity stratification for congenital heart disease patients. I J of Cardiology Congenital Heart Disease. 2023; 12(11):100430.

9.Tianrui C, Yancong Q, Ding L, Xiaorui H, Le Y, Anzhi Y, et al. Multifunctional, breathable MXene-PU mesh electronic skin for wearable intelligent 12-lead ECG monitoring system. Chemical Engineering J. 2023; 455.

10.Pendela K, Suja A A, Bhaskar PVR. Heart disease prediction using hybrid optimization enabled deep learning network with spark architecture. Biomedical Signal Processing and Control. 2023; 84: 104707.

11.Adyasha R, Debahuti M, Ganapati P, Suresh C, Satapathy K X. Improved heart disease detection from ECG signal using deep learning based ensemble model,Sustainable Computing. Informatics and Systems. 2022;35(1):74-6.

12.Chazal P, Reilly R B. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering. 2006; 53(12):2535-3.

13.Guixiang L, Dequn H, Lei W, Jing Z, Jun C, Kai W, et al. A new method of detecting the characteristic waves and their onset and end in electrocardiogram. Signal Processing and Control. 2022; 75(1):21-5.

14.Abo-Zahhad M M, Abdel-Hamid T K, Mohamed, A

Journal of Advanced Biomedical Sciences

Fasa University of Medical Sciences

M. Compression of ECG signals based on DWT and exploiting the correlation between ECG signal samples. I J of Communications. 2014; 7(01): 53-2.

15. Manimegalai P, Bharathi P, Thanushkodi K. Real Time Implementation of Analysis of ECG Characteristic Points Using Discrete Wavelets. Global J of researches in engineering Electrical and electronics engineering. 2012;12 (1):83-9.

16. Raman P, Ghosh S. Classification of Heart Diseases based on ECG analysis using FCM and SVM Methods. I J of Engineering Science. 2016:6739(1):47-8.

17. Kumar R G, Kumaraswamy Y S. Investigating cardiac arrhythmia in ECG using random forest classification. Int. J Comput. Appl. 2012;3(2):1-4.

[DOI: 10.18502/jabs.v14i1.14806]