Automated Prognosis of ECG for Arrhythmia Classification using DWT, PCA framework and SVM Methodologies

Shweta Bahade[1], Deepak Sharma[2]

[1]*Department of Electronics & Telecommunication Engineering, Chhatrapati Shivaji Institute of Technology, Durg (C.G.), India, 491001*
[2]*Department of Electronics & Telecommunication Engineering, Chhatrapati Shivaji Institute of Technology, Durg (C.G.), India, 491001*

---***---

Abstract - *An electrocardiogram (ECG or EKG) is a test that shows the nature of heart conditions by measuring electrical activity of the heart. This has been used as all effective method to identify any heart malfunction prior to a cardiac arrest. Machine learning techniques are a good tool for diagnosis since they are able to observe things that are not seen by naked eyes. The approach adopted for exploiting machine intelligence for automatic heart arrhythmia recognition has been designed, developed and implemented. The proposed system is comprised of four active stages namely: Pre-processing stage, Feature extraction stage, feature reduction stage and finally classification stage for final detection of the arrhythmia. DWT has been extensively used for de-noising purpose in the pre-processing stage as well as for feature extraction. PCA is used in the feature reduction stage and SVM has been employed for classification. SVM is one of the most effective techniques for learning from the provided data. Its classification is vital to saving life. In this project, the records were selected from the open source MIT-BIH arrhythmia and the data is divided into training and testing sets after extraction of the features of the given data. Normal and five different types of abnormalities generally observed were under consideration for detection purpose. Cardiac arrhythmia under consideration were left bundle branch block (LBBB), right bundle branch block (RBBB), atrial premature contractions (APC), premature ventricular contractions (PVC) and paced rhythm beat (PRB). A set containing 800 features were used for training and testing purposes. The proposed system has exhibited an overall efficiency of 97.44% in classifying the abnormalities.*

Key Words: **Electrocardiogram, Arrhythmia, Discrete Wavelet Transform, Principal Component Analysis Support Vector Machine, Dimensionality, De-noising.**

1. INTRODUCTION

The ECG morphology reflects the heart status. It is used to look for pathological patterns among the heartbeats and also used to measure the beats' regularity and other conditions like mental stress. In general, ECG provides two primary types of information. It measures time intervals on ECG, so that a cardiologist can determine the duration in which the electrical wave passes through electrical conduction system of the heart. The regularity or irregularity of the electrical activity is found out with this signal, fast or slow. Secondly, the strength of electrical activity is measured, which tells a cardiologist if parts of the heart are too large or are overworked. Any disorder in electrical functioning of the heart neural cells affects the ECG signals, termed as arrhythmia [1]. The techniques applied to execute the task are DWT, PCA and SVM.

For the feature extraction and classification task we will be using discrete wavelet transform (DWT) [2] as wavelet transform which is a two-dimensional timescale processing method. In pre-processing stage, de-noising the ECG signal is done. In further processing, dimensionality reduction [3] is performed based upon feature selection, Principal Component Analysis (PCA) is utilized for this purpose.PCA is a multivariate technique for identifying patterns in data and represents the data to highlight their redundancies. These identified patterns of data can be compressed by reducing the number of dimensions, without much loss of information. Further for classification purpose Support Vector Machine (SVM) has been used.

The contribution lies in the fact that the combination made in the work works excellently as compared other works reported in this paper. In addition, the work may lead to many biomedical-based applications including biometric systems, health status identification, heart disease risk management etc.

Paper is organized as follows: Data description and methodology are explained in section II. Obtained results and discussion are presented in section III. Section IV presents the conclusion and Section V talks about the future scope.

2. METHODOLOGY

2.1 Block Diagram of Work Flow

The machine intelligence based cardiac arrhythmia detection involves multiple operations on the ECG signal. These involve pre-processing of the ECG signal, features extraction stage, feature reduction stage and classification stage. Processing phase involves the readiness of the signal by removing the DC baseline, removal of various types of noise [4] by suitable methods. Feature extraction stage is used to acquire the features of the signal helpful for the decision making process, feature reduction process helps in diminishing the feature vectors to discard the redundant features from the feature database for speedy execution and finally the classification process provides the decisions regarding the status of the heart. A schematic view of the overall methodology adopted in the present work is represented in Fig-1 to understand the methods used along with their purposes.

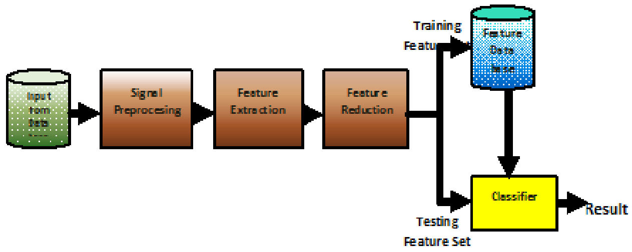

Fig -1: Block diagram of work flow

2.2 Discrete Wavelet Transform (DWT)

DWT has been used for feature extraction from ECG signal. In DWT, a time-scale representation of the digital signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cut-off frequencies at different scales. Wavelets can be realized by iteration of filters with rescaling. The resolution of the signal, which is a measure of the amount of detail information in the signal, is determined by the filtering operations, and the scale is determined by up-sampling and down-sampling (sub-sampling) operations. After Removal of DC gain and Baseline Wander DWT based De-noising process is done.DWT de-noising are performed in three basic steps, first one is called analysis decomposition DWT filter bank, second one is called thresholding and the last one is called synthesis reconstruction IDWT filter bank. Now the signal would be reconstructed and ready for further sampling.

2.3 Principle Component Analysis (PCA)

Higher number of features leads to higher computational cost and execution time required will be high. Some of the features extracted may be correlated that sum up to a huge number of redundant data. This significantly affects the classification process. In order to enhance the accuracy and efficiency of the classification process, it is important to discard the correlated features. This is known as feature reduction which is achieved by Principal Component Analysis (PCA) [5]. The goal is to extract the important information from the table of observations, represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity or dissimilarity of the observations and of the variables. PCA, being a simple, non-parametric method of extracting relevant information from confusing data sets, is used abundantly in all forms of analysis.PCA [6] shows the way for reducing a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlie it. With this dimensional reduction, this technique looks for simplifying a statistical problem with the minimal loss of information.

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) has been used for classification purpose. Support vector machine is primarily the classifier method that performs classification tasks by constructing hyper-plane in multidimensional space that separates cases of different class labels. SVM [7] supports both regression and classification tasks and can handle multiple continues variable.

2.5 Dataset Used

The database for the ECG signals is obtained from MIT – BIH database. The DWT used is sym8 with a decomposition level 9. Thresholding applied is minimax type.

Table -1: Data set used for training and testing

Type of beat	MIT-BIH Record No.	Total number of R peaks detected	Number of beats used for training	Number of beats used for testing
N (Normal)	100, 101, 103, 105	160	100	60
LBBB (Left Bundle Branch Block)	109, 111, 207, 214	160	100	60
RBBB (Right Bundle Branch Block)	118, 124, 212, 231	160	100	60
PVC (Premature Ventricular Contraction)	106, 119, 200, 203	160	100	60
APC (Atrial Premature Contraction)	209, 222	80	50	30
PRB (Paced Rhythmic Beat)	107, 217	80	50	30
TOTAL	**24**	**800**	**500**	**300**

2.6 Algorithm

The algorithm (Fig-2) paints the stepwise picture of the complete process involved.

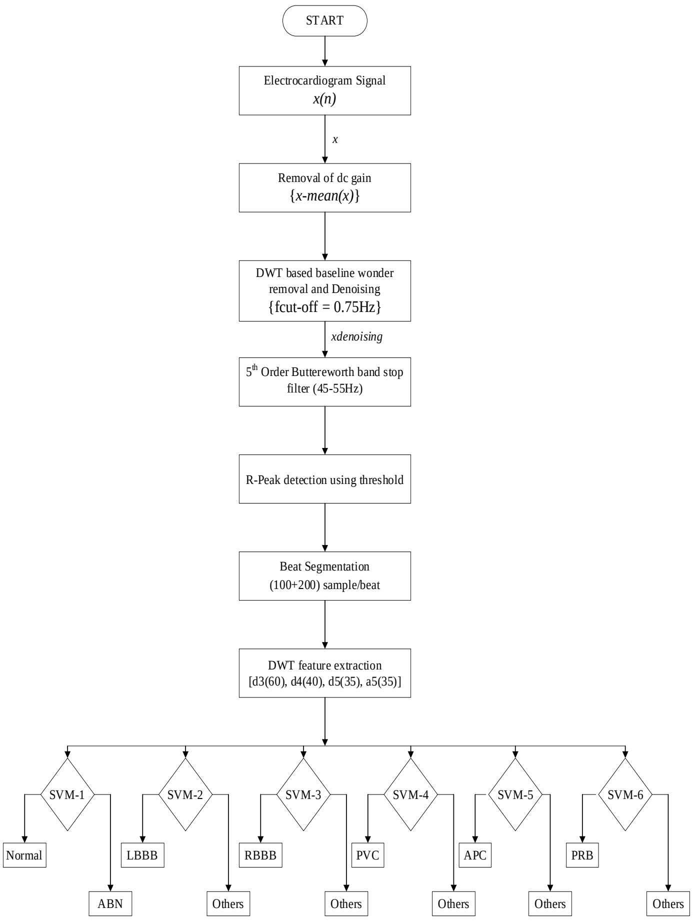

Fig -2: Work Flow Algorithm Diagram

It describes the whole work has been divided into four phases: Preprocessing of the ECG, Feature Extraction, Feature Reduction and Classification. Preprocessing stage includes: DC Gain removal, Normalization, Baseline wander removal, Power Line Interference removal and de-noising. Feature extraction phase includes forming of a feature set from the database. Feature reduction stage performs the deductions of redundant features, which are not required or somehow correlated to other features. Classification stage actually detects the arrhythmia present and its type.

3. RESULTS AND DISCUSSION

With the help of data set (Table-1) we perform the classification in six classes. In first class we classified a signal in between normal and abnormal ECG. In this case we got beat-wise testing accuracy of 98.67%. Likewise, we classified Left Bundle Branch Block (LBBB) and others and got 97% of beat-wise testing accuracy. Similarly, we tested four more ECG signals namely Right Bundle Branch Block (RBBB),[8],Premature Ventricular Contractions (PVC), Atrial Premature Contractions (APC) [9] [10] and Paced Rhythm beat (PRB) and got the accuracy as shown in the Table-2.

Table -2: Performance metric of the classification method

Type of beat	MIT-BIH Record No.	Total number of R peaks detected	Number of beats used for training	Number of beats used for testing
N (Normal)	100, 101, 103, 105	160	100	60
LBBB (Left Bundle Branch Block)	109, 111, 207, 214	160	100	60
RBBB (Right Bundle Branch Block)	118, 124, 212, 231	160	100	60
PVC (Premature Ventricular Contraction)	106, 119, 200, 203	160	100	60
APC (Atrial Premature Contraction)	209, 222	80	50	30
PRB (Paced Rhythmic Beat)	107, 217	80	50	30
TOTAL	**24**	**800**	**500**	**300**

The average accuracy of the classification system [11][12] is coming out to be 97.44 % and this shows that the proposed method for the detection of arrhythmia is very effective and efficient and supersedes many other techniques reported in the literature. A visual demonstration has also been provided in the following self-explanatory figures for better understanding the effectiveness of various stages employed in the work.

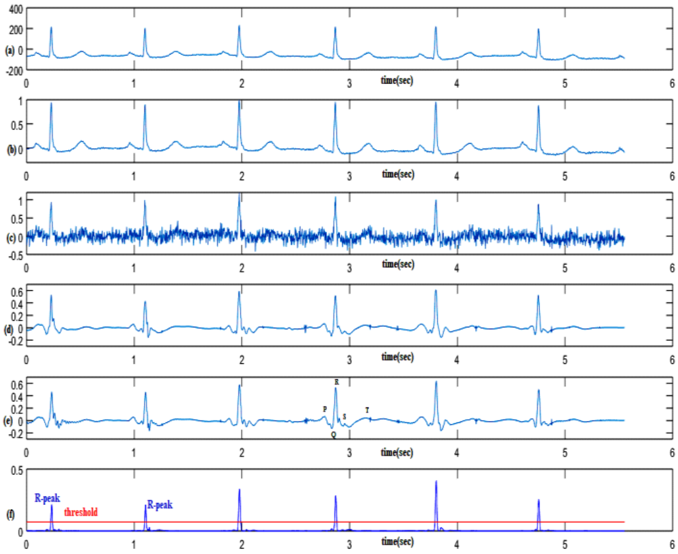

Fig - 3 : ECG for healthy heart (6sec) (a) Original signal (b)Normalization with dc gain removed (c) Addition of white Gaussian noise (d) DWT based baseline wonder and noise removal (e) Removal of power frequency carrier (45-55hz) using Butterworth filter (f) R-peak detection using thresholding (th=0.18*max(abs(x))^2) for segmentation (100+ R-peak + 200 samples).

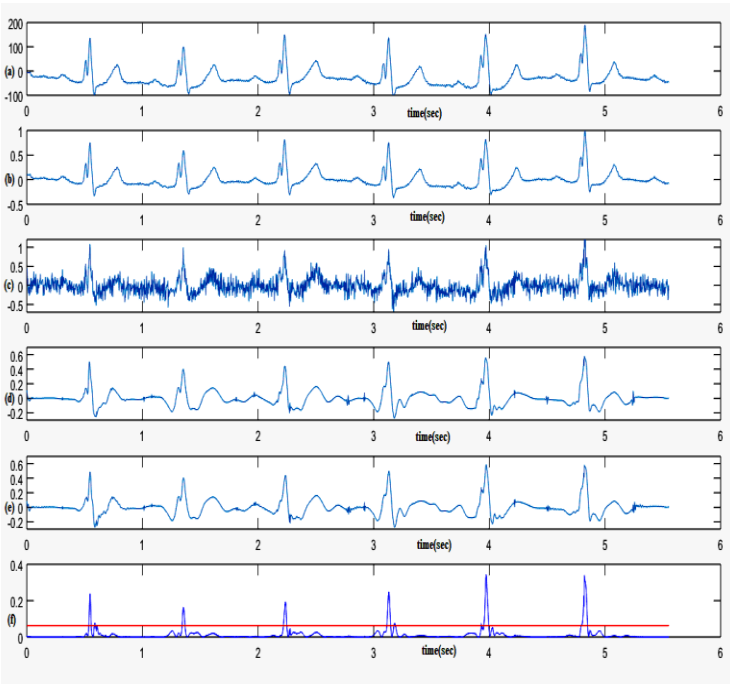

Fig - 4 : Left bundle branch block (LBBB) (6sec) (a) Original signal (b)Normalization with dc gain removed (c) Addition of white Gaussian noise (d) DWT based baseline wonder and noise removal (e) Removal of power frequency carrier (45-55hz) using Butterworth filter (f) R-peak detection using thresholding (th=0.18*max(abs(x))^2) for segmentation (100+ R-peak + 200 samples).

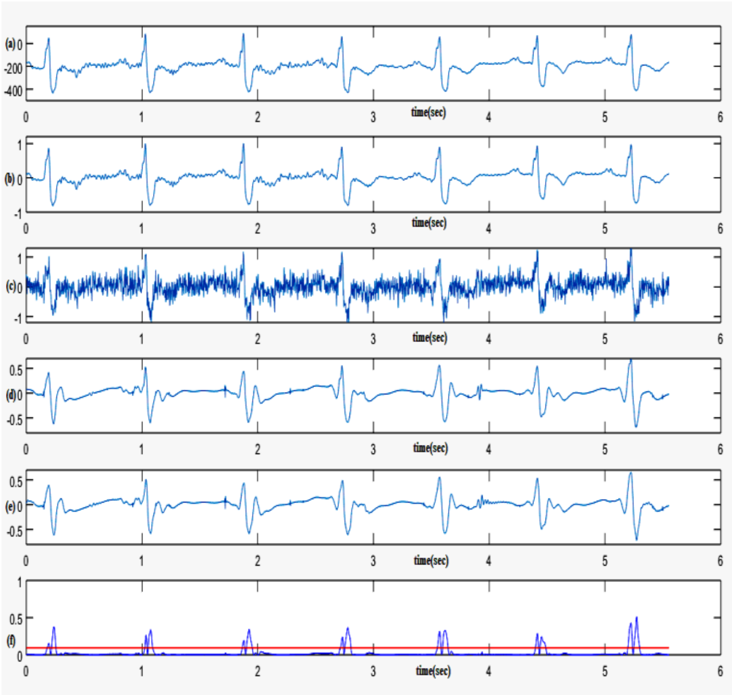

Fig - 5 : Right bundle branch block (RBBB) (6sec) (a) Original signal (b)Normalization with dc gain removed (c) Addition of white Gaussian noise (d) DWT based baseline wonder and noise removal (e) Removal of power frequency carrier (45-55hz) using Butterworth filter (f) R-peak detection using thresholding (th=0.18*max(abs(x))^2) for segmentation (100+ R-peak + 200 samples).

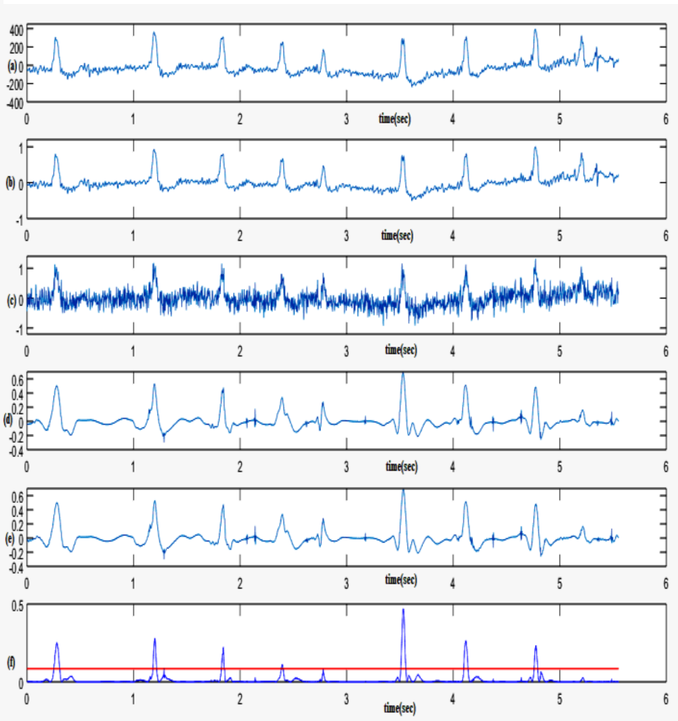

Fig – 6 : Premature ventricular contractions (PVC) (6sec) (a) Original signal (b)Normalization with dc gain removed (c) Addition of white Gaussian noise (d) DWT based baseline wonder and noise removal (e) Removal of power frequency carrier (45-55hz) using Butterworth filter (f) R-peak detection using thresholding (th=0.18*max(abs(x))^2) for segmentation (100+ R-peak + 200 samples).

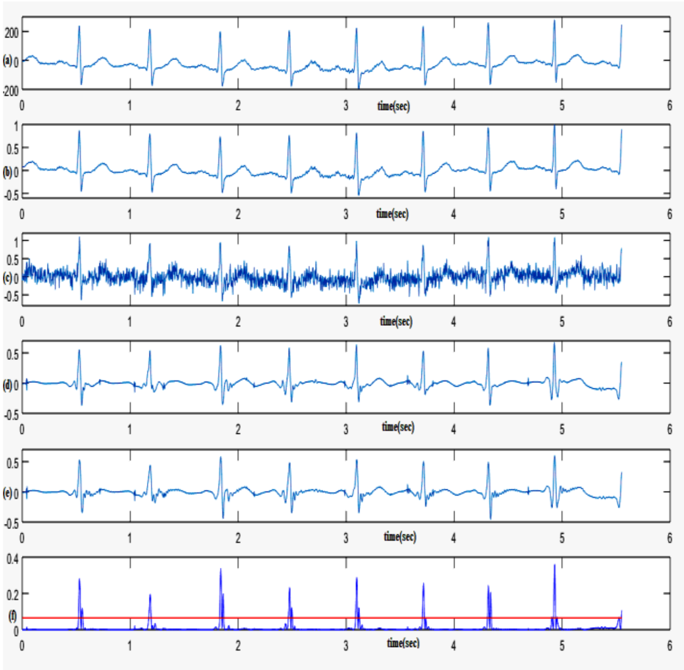

Fig – 7 : Atrial premature contractions (APC) (6sec) (a) Original signal (b)Normalization with dc gain removed (c) Addition of white Gaussian noise (d) DWT based baseline wonder and noise removal (e) Removal of power frequency carrier (45-55hz) using Butterworth filter (f) R-peak detection using thresholding (th=0.18*max(abs(x))^2) for segmentation (100+ R-peak + 200 samples).

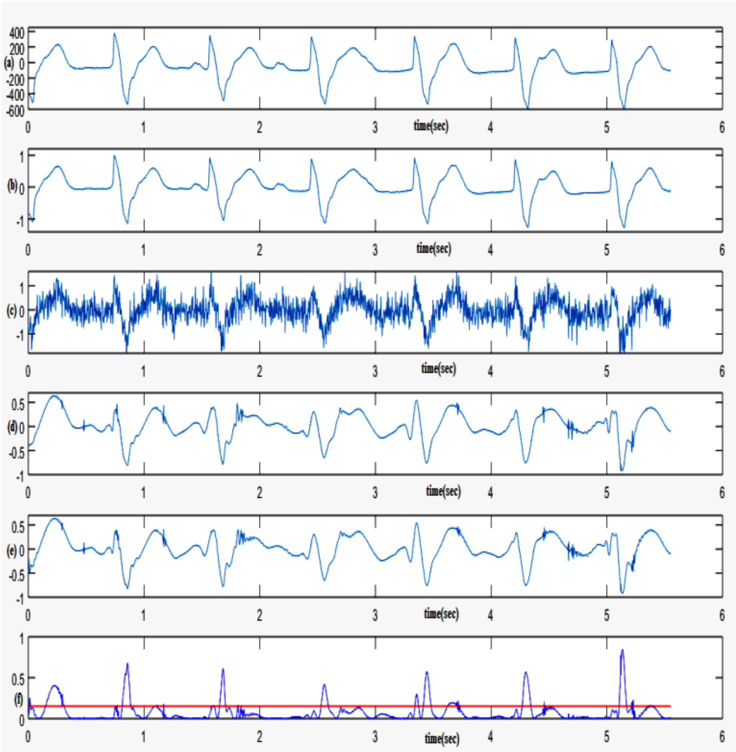

Fig – 8 : Paced rhythm beat (PR) (6sec) (a) Original signal (b)Normalization with dc gain removed (c) Addition of white Gaussian noise (d) DWT based baseline wonder and noise removal (e) Removal of power frequency carrier (45-55hz) using Butterworth filter (f) R-peak detection using thresholding (th=0.18*max(abs(x))^2) for segmentation (100+ R-peak + 200 samples).

4. CONCLUSION

From the result section, it is observed that the suggested process provided an overall accuracy of 97.44% which is above the acceptable limit as provided by Association for the Advancement of Medical Instrumentation (AAMI).

5. Future Scope

The designed automated arrhythmia detector achieves a decent performance to pin point the type of disorder in the activities of the heart. However, the system lacking the scope of acquisition of ECG signals under real circumstances. So ample scope is there in designing of acquisition arrangement and then the suggested method can be applied upon those acquired signals to actually measure its performance. Also, the computational and time complexity can be measured and compared with other existing techniques. Instead of using MATLAB as a platform for implementation, the effect of using computer languages might have provided more flexibility in design and modifications in the program.

REFERENCES

[1] U. Desai, et al., "Discrete Cosine Transform Features in Automated Classification of Cardiac Arrhythmia Beats." Emerging Research in Computing, Information, Communication and Applications, Springer India, 2015.

[2] Qibin Zhao, and Liqing Zhan, 2005 "ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines," International Conference on Neural Networks and Brain, ICNN&B '05, vol. 2.

[3] He, H.; Tan, Y. Automatic pattern recognition of ECG signals using entropy-based adaptive dimensionality reduction and clustering. Appl. Soft Comput. 2017.

[4] Rahul Kher (2019) Signal Processing Techniques for Removing Noise from ECG Signals. J Biomed Eng 1: 1-9.

[5] R.J. Martis, C. Chakraborty, and A.K. Ray, "An integrated ECG feature extraction scheme using PCA and wavelet transform," in: IEEE India Annual Conference (INDICON), 2009, IEEE, 2009.

[6] Herve Abdi and Lynne J. Williams "Principal Component Analysis", Wiley Interdisciplinary Reviews: Computational Statistics, 2011.

[7] Qin, Q.; Li, J.; Zhang, L.; Yue, Y.; Liu, C. Combining Low-dimensional Wavelet Features and Support Vector Machine for Arrhythmia Beat Classification. Sci. Rep. 2017.

[8] Padmavathi Kora and K.Sri Rama Krishna, Hybrid firefly and Particle Swarm Optimization algorithm for the detection of Bundle Branch Block, International Journal of the Cardiovascular Academy, Volume 2, Issue 1, March 2016,doi.: doi.org/10.1016/j.ijcac.2015.12.001.

[9] H. Prasad, R.J. Martis, U.R. Acharya, C.M. Lim, J.S. Suri, "Application of higher order spectra for accurate delineation of atrial arrhythmia, "Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE,2013.

[10] Cao, X.C.; Yao, B.; Chen, B.Q. Atrial fibrillation detection using an improved multi-Scale decomposition Enhanced residual convolutional neural network. IEEE Access 2019.

[11] Qiao Li, Cadathur Rajagopalan, Gari D.Clifford, A machine learning approach to multi-level ECG signal quality classification, Computer Methods and Programs in Biomedicine, Volume 117, Issue 3, December 2014.

[12] https://en.wikipedia.org/wiki/Support-vector_machine