

A Technical Review on Statistical Feature Extraction of ECG signal

Asutosh Kar¹, Leena Das²

¹ Dept. of Electronics and Telecommunication Engineering, IIIT, Bhubaneswar, India, ² Dept. of Computer Science Engineering, Sambalpur University, Sambalpur, India

ABSTRACT

ECG Feature Extraction plays a significant role in diagnosing most of the cardiac diseases. In this paper a comprehensive review has been made for statistical feature extraction of ECG signal analyzing classifying method which have been proposed during the last decade and under evaluation that includes digital signal analysis, Fuzzy Logic methods, Artificial Neural Network, Hidden Markov Model, Genetic Algorithm, Support Vector Machines, Self-Organizing Map, Bayesian and other method with each approach exhibiting its own advantages and disadvantages. To diagnose the condition of the heart Electrocardiography is an important tool but it is a time consuming process to analyze a long duration ECG signal as it may contain thousands of heart beats. Hence it is desired to automate the entire process of heart beat classification and preferably diagnose it accurately. For subsequent analysis of ECG signals its fundamental features like amplitudes and intervals are required which determine the functioning of heart.

Index Terms -Artificial Neural Network, Discrete Wavelet Transform, ECG Signal, Fuzzy Logic, Lyapunov Exponent, Support Vector Machine.

1. INTRODUCTION

Electrocardiogram (ECG) is a nearly periodic signal that reflects the activity of the heart. A lot of information on the normal and pathological physiology of heart can be obtained from ECG. However, the ECG signals being non-stationary in nature, it is very difficult to visually analyze them. Thus the need is there for computer based methods for ECG signal Analysis. Clinical observation of ECG can take long hours and can be very tedious ECG being a non-stationary signal, the irregularities may not be periodic and may show up at different intervals. Moreover, visual analysis cannot be relied upon. This calls for computer-based techniques for ECG analysis. At every beat, the heart is depolarized to trigger its contraction. This electrical activity is transmitted throughout the body and can be picked up on the skin which is the principle behind the ECG. An ECG machine records this activity via electrodes on the skin and displays it graphically. An ECG involves attaching 10 electrical cables to the body: one to each limb and six across the chest. ECG is a wave that represents an electrical event in the heart such as atria depolarization, ventricular depolarization, atria repolarization, ventricular repolarization. The signal consists of a series of repetitive complex waveforms with a frequency of approximately 1 Hz. One cardiac cycle in an ECG signal consists of the P-QRS-T waves. The majority of the clinically useful information in the ECG is originated in the intervals and amplitudes defined by its features.

Cardiovascular disease remains the number one cause of mortality in the western world, responsible for more than 16 million deaths annually worldwide [1]. Changes in life-style, such as reducing cholesterol intake and exercising regularly can reduce the chances of a fatal event associated with CVD. Therefore, early detection is a critical step in the prevention of death associated with CVD. A regular doctor visit, which includes an ECG, is a vital step towards early detection, results in large volumes of patient data that must be carefully scrutinized. Conventional methods of monitoring and diagnosing electrocardiographic changes rely on detecting the presence of particular signal features by a human observer [2]. QRS complex is the most prominent feature in electrocardiogram because of its shape; therefore it is taken as a reference in ECG feature extraction. Computer based medical diagnostic systems have been

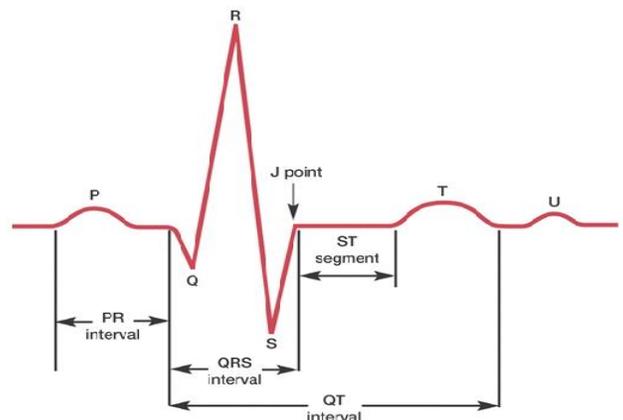


Figure.1 ECG signal showing P-QRS-T wave

developed in order to assist medical professionals in the analysis of large volumes of patient data. Various signal processing techniques have been utilized in extracting features from the biomedical signals and analyzes these features which have their own merits and demerits. Such techniques work by transforming the mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem [2]. The techniques have been used to address this problem such as the analysis of ECG signals for detection of electrocardiographic changes using the autocorrelation function, frequency domain features, time-frequency analysis, and wavelet transform. Some methods consist of series of band pass filters having frequency range of QRS complexes but these methods have limited accuracy in analyzing ECG features in presence of high frequency noise as well as the ECG signal affected by severe base line drift. In recent years, there has been an increasing interest in applying techniques from the domains of nonlinear analysis and chaos theory in

studying the behavior of a dynamical system from an experimental time series such as ECG signal. The purpose of these studies is to determine whether dynamical measures can serve as clinically useful parameters [3]. Various techniques proposed earlier in literature for extracting the features from ECG is analyzed this paper discusses and a review has been made to find out the best among them with less computational complexity and more accuracy in prediction and feature extraction.

2. FEATURE EXTRACTION OF ECG

Feature extraction method using wavelet transform and classification using support vector machines was first proposed in [1]. A new approach to the feature extraction was presented for reliable heart recognition. Three main steps were performed-

- [1] Data preprocessing,
- [2] Feature extraction and
- [3] Classification of ECG signals.

Two methods were applied together to extract the features of ECG signal which gives the feature vector of ECG data set. To extract the coefficients of the transform as the features of each ECG segment wavelet transform is used. Concurrently, autoregressive modeling (AR) is also applied to get hold of the temporal structures of ECG waveforms. Then finally the support vector machine (SVM) with Gaussian kernel is used to classify different ECG heart rhythm. The results of computer simulations reached the overall accuracy of 99.68%. The cardiac depolarization route has been shown in fig.2. The wave of depolarization then proceeds rapidly to the bundle of His where it splits into two pathways and travels along the right and left bundle branches [3]. The impulse travels the length of the bundles along the interventricular septum to the base of the heart, where the bundles divide into the Purkinje system [3]. The wave of depolarization is then distributed to the ventricular walls and initiates ventricular contraction. The first step in extracting ECG features starts from the accurate detection of R peaks in the QRS complex.

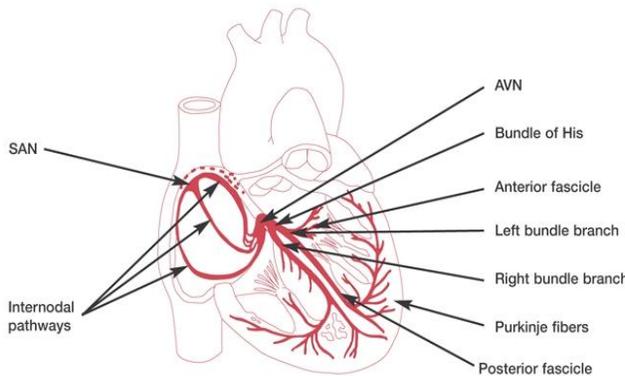


Figure.2 Cardiac depolarization route [3]

Table 1: Electrophysiology [4]

Action	Effect
Depolarization	Shifting of electrolytes across the cell membrane causes change in electric charge.
Repolarization	Internal negative charge is restored and the cells return to their resting state.

Table 2: Conduction System and Functions [4]

Structure	Function and Location
Sinoatrial (SA) Node	Dominant pacemaker of the heart, locate in upper portion of right atrium. Intrinsic rate 60-100 bpm.
Internodal Pathways	Direct electrical impulses between SA and AV nodes.
Atrioventricular (AV) node	Part of AV junctional tissue. Slows conduction, creating a slight delay before impulses reach ventricles. Intrinsic rate 40-60 bpm.
Bundle of His	Transmits impulses to bundle branches. Located below AV node.
Left bundle Branch	Conducts impulses that lead to left ventricle.
Right bundle Branch	Conducts impulses that lead to right ventricle.

A robust R wave detector using the wavelets was developed [2] by Awadesh and Manabendra. The wavelets used for detection are Daubechies and Symmetric. The database has been collected from MIT-BIH arrhythmia database and the signals from Lead-II have been analyzed. The selection of detail coefficient d4 has been done based on the following important parameters i.e.

- [1] Energy
- [2] Frequency and
- [3] Cross-correlation analysis

of decomposition structure of ECG signal. Forty two records were tested for R peaks. The overall of detection using db6 and sym11 are 96.65% and 84.37% respectively. The importance of using wavelet transform has been highlighted in which the noise is filtered at each level of decomposition thus eliminating the requirement of any preprocessing. This ensures the robustness of the method. Further confirmation is done using different records of

the database with noise present in it. The results with db6 have been found to be more stable by varying threshold than sym11 which picks up false peaks [4]. The effect of zero padding has also been eliminated during energy analysis making the algorithm simpler and less time consuming.

Table 3: Normal heart beat rates determined by QRS complex [3]

Number of large squares between QRS complexes	Heart rate(bpm)
5	60
4	75
3	100
2	150

A novel approach for feature extraction of ECG signal was proposed in [3]. The proposed letter present an algorithm, based on the wavelet transform for feature extraction from an (ECG) signal and recognition of abnormal heartbeats. Since wavelet transforms provides time frequency localization they developed a method for choosing an optimal mother wavelet from a set of orthogonal and bi-orthogonal wavelet filter bank. Choosing a wavelet function depends on the ability to reconstruct the signal from the wavelet decomposition and to preserve the energy under the transformation. The next step of the approach is to remove noise from the ECG signal by a soft or hard threshold with limitation of 99.99 reconstructs ability and then each PQRST cycle is decomposed into coefficients vector by the optimal wavelet function [3]. The coefficients, approximations of the last scale level and the details of all levels, are used for the ECG analyzed. Coefficients of each cycle was divided into three segments that are related to (A) P-wave, (B) QRS complex, and (C) T-wave. The summation of the values from these segments provided the feature vectors of single cycles. This algorithm was tested on two ECG signals, the first was taken from the MIT biomedical database was decomposed into four levels and denoised by the optimal wavelet "sym4" with global threshold value 1.3073. This ECG signal was with local abnormal heartbeat activity [5]. The second ECG signal was recorded from a patient during an epileptic seizure. The optimal wavelet function was "coif5" with global threshold 23.217. The three waves of the QRS complex represent ventricular depolarization [3].

- [1] Small Q waves correspond to depolarization of the interventricular septum. Q waves can also relate to breathing and are generally small and thin. They can also signal an old myocardial infarction (in which case they are big and wide)
- [2] The R wave reflects depolarization of the main mass of the ventricles –hence it is the largest wave
- [3] The S wave signifies the final depolarization of the ventricles, at the base of the heart

Another technique using the Daubechies wavelet transform for feature extraction of ECG signal was proposed [4]. A electrocardiogram (ECG) feature extraction system based on the multi-resolution wavelet transform had been developed and

evaluated. The ECG signals from Modified Lead II (MLII) were chosen for processing. For better detection the wavelet filter with scaling function similar to the shape of the ECG signal was chosen. The foremost step of their approach was to denoise the ECG signal by removing the equivalent wavelet coefficients at higher scales. Then QRS complexes are detected and each one complex is used to trace the peaks of the individual waves, including onsets and offsets of the P and T waves which are present in one cardiac cycle. The experimental results revealed that the proposed approach for ECG feature extraction achieved sensitivity of 99.18% and a positive prediction of 98%.

For the detection of QRS complexities, an algorithm was presented [5]. The recognition of QRS complexes forms the origin for more or less all automated ECG analysis algorithms. The feature considered here in this algorithm for the detection of QRS complex is the slope of ECG signal. A succession of transformations of the filtered and baseline drift corrected ECG signal is used for mining of a new modified slope-feature. In the presented algorithm, filtering procedure based on moving averages provides smooth spike-free ECG signal, which is appropriate for slope feature extraction. The foremost step is to extort slope feature from the filtered and drift corrected ECG signal, by processing and transforming it, in such a way that the extracted feature signal is significantly enhanced in QRS region and suppressed in non-QRS region. The proposed method has detection rate and positive prediction of 98.56% and 99.18% respectively.

Tayel and Bouridy [6] together put forth a technique for ECG image classification by extracting their feature using wavelet transformation and neural networks. Features are extracted from wavelet decomposition of the ECG images intensity. The obtained ECG features are then further processed using artificial neural networks. The features are:

- I. Mean & median
- II. Maximum & minimum value
- III. Standard deviation, variance, & mean absolute deviation.

The introduced ANN was trained by the main features of the 63 ECG images of different diseases. The test results showed that the classification accuracy of the introduced classifier was up to 92%. The extracted features of the ECG signal using wavelet decomposition was effectively utilized by ANN in producing the classification accuracy of 92%.

Another algorithm for feature extraction of ECG signals was proposed by in [7]. The basic focus of the work was to evaluate the classification performance of an automatic classifier of the electrocardiogram (ECG) for the detection of abnormal beats. The concept of feature extraction was completely new. The obtained feature sets were based on ECG morphology and RR-intervals. Configuration adopted Kohonen self-organizing maps (SOM) for examination of signal features and clustering. A classifier was developed with SOM and learning vector quantization (LVQ) algorithms using the data from the records recommended by ANSI/AAMI EC57 standard. Moreover the proposed work compares two strategies for classification of annotated QRS complexes:

- Based on original ECG morphology features
&

- Proposed new approach - based on preprocessed ECG morphology features.

The mathematical morphology filtering was used for the preprocessing of ECG signal.

Another technique for feature extraction was proposed in [8]. Three artificial neural network models (a. MLP, b. RBF neural networks and c. SOFM) were considered for ECG classification. Four types of ECG beats were chosen to be recognized, including

- I. Normal sinus rhythm,
- II. Premature ventricular contraction,
- III. Atrial premature beat &
- IV. Left bundle branch block beat.

Six different features (Mean R-peak value, mean power spectral density, Area under QRS complex, energy of the signal, Q-S distance and autocorrelation value) were extracted for characterizing the four classes of heart beats. MLP gives the best performance as far as overall accuracy is considered, five types of learning rules namely

- Step,
- Momentum,
- Conjugate gradient,
- Quick prop and
- Delta bar-delta

are evaluated in MLP and momentum is chosen as the learning algorithm because it gives out the best results when the speed performance is calculated. SOFM also exhibits good performance RBF neural network also performs very well, its main strength being, faster training, as compared to MLP. The three classifiers are subjected to noise analysis, and the performances of the classifiers were evaluated by contaminating the original ECG signal with white Gaussian noise of varying strength. Performance of all three classifiers in presence of noise is fairly robust. This is an indication of the effectiveness of the selected features.

In [9] various machine learning algorithms are discussed for the classification of ECG. It summarizes some of the principle machine learning algorithms to ECG classification, evaluating them in terms of the features they employ, the type(s) of cardiovascular diseases to which they are applied, and their classification accuracy. The various classification techniques discussed in the letter. The data indicate that a variety of features have been utilized, yielding a range of classification accuracies which are within the range of 70-100% for the most part, depending on the exact metric used, here no comparison is made between the various machine learning algorithms i.e. no systematic study has been performed which has examined the accuracy of a variety of classifiers on a single dataset. It has been discussed that these algorithms are highly accurate and are useful within a clinical setting but it depends on the role of the classifier. If the results are to be used to determine a course of treatment, then specificity and sensitivity issues must be taken into account.

A method for automatic extraction of both time interval and morphological features, from the Electrocardiogram (ECG) to classify ECGs into normal and arrhythmic was described in [10]. For feature extraction of ECG signals Linear Discriminant Analysis (LDA) technique was used and for the classification purpose Artificial Neural Network was used. Five ECG features namely a. RR, b. RT, c. T wave amplitude, d. T wave skewness, and e. T wave kurtosis were used in their method. These features

are obtained with the assistance of automatic algorithms. The onset and end of the T wave were detected using the tangent method. The three feature combinations used had very analogous performance when considering the average performance metrics.

In [11] the classification accuracies of ME (Mixture Experts) trained on composite features and MME (Modified Mixture Experts) trained on diverse features were compared. Basically here it is tried to find out whether the automated diagnostic systems with diverse features (MME) or composite features (ME) improve the capability of classification of the ECG signals. These features are calculated using the various Eigen Vector Methods- Pisarenko method, MUSIC method, Minimum Norm method. The present model consists of three main modules:

- A. A feature extractor that generates a feature vector from the ECG signals,
- B. Feature selection that composes diverse and composite features (power levels of the PSDs obtained by the eigenvector methods) &
- C. Feature classifiers that output the class based on the diverse and composite features (mixture of experts – ME, modified mixture of experts – MME).

The ECG signals (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat, atrial fibrillation beat) from the Physiobank database were used for training and testing of classifiers. MME classifier trained on the three diverse feature vectors produce better performance than that of the ME trained on the composite features. The results of the present study demonstrated that the MME can be used in classification of the ECG signals by taking into consideration the misclassification rates.

Sufi et al [12] formulated a new ECG obfuscation method for feature extraction and corruption detection. A new ECG obfuscation method was presented, which uses cross correlation based template matching approach to distinguish all ECG features followed by corruption of those features with added noises. Reconstruction of the obfuscated features was difficult without the prior knowledge of the templates used for feature matching and the noise. Therefore, three templates and three noises were considered for P wave, QRS Complex and T wave comprise the key, which is only 0.4%-0.9% of the original ECG file size. With this obfuscation model, the corrupted ECG appears as regular ECG without encryption, noise can be represented in enormous number of combinations establishing unmatched security and the key distribution is efficient due to its small size.

In [13] an approach for effective feature extraction from ECG signals was described. This research work deals with a composite method which has been developed for

- A. Data Compression
- B. Signal retrieval &
- C. Feature extraction of ECG signals.

It has been found that signal retrieval from the compressed data not only compresses the data but also improves the quality of the retrieved ECG signal with respect to elimination of high-frequency interference present in the original signal. The best topology with two hidden layers and four elements in each hidden layer has been finalized for ECG data compression using a Military Hospital (MH) data base. It has been observed that a higher compression ratio can be achieved using ANN, as compared with other methods of data compression, because the compression ratio in this method depends on the number of cycles taken for

compression. Moreover the features extracted by amplitude, slope and duration criteria from the retrieved signal match with the features of the original signal. Their experimental results at every stage are steady and consistent and prove beyond doubt that the composite method can be used for efficient data management and feature extraction of ECG signals in many real-time applications.

A feature extraction method using Discrete Wavelet Transform (DWT) was proposed in [14]. It used a discrete wavelet transform (DWT) to extract the relevant information from the ECG input data in order to perform the classification task. The proposed work includes the following modules

- A. Data acquisition,
- B. Pre-processing beat detection,
- C. Feature extraction and
- D. Classification.

In the feature extraction module the Wavelet Transform (DWT) is designed to address the problem of non-stationary ECG signals. It was derived from a single generating function called the mother wavelet by translation and dilation operations. Using DWT in feature extraction may lead to an optimal frequency resolution in all frequency ranges as it has a varying window size, broad at lower frequencies, and narrow at higher frequencies. The DWT characterization will deliver the stable features to the morphology variations of the ECG waveforms. The nonlinearity of ECG signal was considered in this letter and chaos theory was put forward to study the behavior of dynamical system from an experimental time series (ECG Signal) [15]. This consideration was tested successfully using the nonlinear dynamics tools, like the computation of Lyapunov exponents. The four ECG beats (normal beat, congestive heart failure beat, ventricular tachyarrhythmia beat, atrial fibrillation beat) obtained from the Physiobank database were classified using the Multilayer Perceptron Neural Network architectures. The computed Lyapunov exponents of the ECG signals were used as inputs of the MLPNNs trained with back propagation, delta-bar-delta, extended delta-bar-delta, quick propagation, and Levenberg–Marquardt algorithms. To reduce the dimensionality of the extracted features statistical features were used. The results confirmed that the MLPNN trained with the Levenberg–Marquardt algorithm has potential in detecting the variabilities of the ECG signals. Total classification accuracy is 95%.

In [16] an algorithm was proposed based on chaos theory for ECG feature extraction. Numerous chaos methods, including phase space and attractors, correlation dimension, spatial filling index, central tendency measure and approximate entropy were discussed. A new feature extraction environment was created called ECG chaos extractor to apply the above mentioned chaos methods. A new semi-automatic program for ECG feature extraction has been implemented and is presented in this article. Graphical interface is used to specify ECG files employed in the extraction procedure as well as for method selection and results saving. The program extracts features from ECG files.

Another algorithm called the Slope Vector Waveform for the detection of QRS complex and RR interval evaluation was discussed [17]. In this proposed method variable stage differentiation is used to achieve the desired slope vectors for feature extraction, and the non-linear amplification is used to get better of the signal-to-noise ratio. The method allows for a fast and accurate search of the R location, QRS complex duration, and RR interval and yields excellent ECG feature extraction results. In order to get QRS durations, the feature extraction rules are needed.

3. FUTURE ENHANCEMENT

ECG signal plays a vital role in diagnosing various cardiac disorders so it is useful to extract the features of ECG signal. Many time domain and frequency domain methods [18] are used for feature extraction which has its own advantages and limitations. The future work focuses on the fact that the different techniques used for extracting the features must provide high accuracy and should be fast and easy to implement.

4. CONCLUSION

This paper provides an overview of the various techniques and algorithms for feature extraction of ECG signal proposed earlier in literature. Advantages and limitations of many methods [18] have been discussed. Need of frequency domain methods [19],[20] are described as it is not always necessary to analyze all the features. The feature extraction technique or algorithm developed for ECG must be highly accurate and should ensure fast extraction of features from the ECG signal.

5. REFERENCES

- [1] Qibin Zhao, and Liqing Zhan, "ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines," International Conference on Neural Networks and Brain, ICNN&B, vol. 2, pp. 1089-1092, 2005.
- [2] Awadhesh Pachauri, and Manabendra Bhuyan, "Robust Detection of R-Wave Using Wavelet Technique," World Academy of Science, Engineering and Technology 56, 2009.
- [3] Ashley EA, Niebauer J., "Conquering the ECG," London: Remedica; 2004.
- [4] F.A Davis, "ECG notes," 2005.
- [5] V. S. Chouhan, and S. S. Mehta, "Detection of QRS Complexes in 12-lead ECG using Adaptive Quantized Threshold," IJCSNS International Journal of Computer Science and Network Security, vol. 8, no. 1, 2008.
- [6] M.B. Tayel, and Mohamed E. El-Bouridy, "ECG Images Classification Using Feature Extraction Based On Wavelet Transformation And Neural Network," ICGST, International Conference on AIML, June 2006.
- [7] P. Tadejko, and W. Rakowski, "Mathematical Morphology Based ECG Feature Extraction for the Purpose of Heartbeat Classification," 6th International Conference on Computer Information Systems and Industrial Management Applications, CISIM '07, pp. 322-327
- [8] R.Ghongade and A. Ghatol, "An Effective Feature Set for ECG Pattern Classification," Lecture Notes in Computer Science, Volume 4901, Medical Biometrics, pp. 25-32, 2007
- [9] Abdael-Badeeh M. Salem, Kenneth Revett, El-Sayed A. El-Dahshan, "Machine Learning in Electrocardiogram Diagnosis," Proceedings of the International Multiconference on Computer Science and Information Technology, pp. 429 – 433, 2009.
- [10] C. Alexakis, H. O. Nyongesa, R. Saatchi, N. D. Harris, C. Davies, C. Emery, R. H. Ireland, and S. R. Heller, "Feature Extraction and Classification of Electrocardiogram (ECG)

- Signals Related to Hypoglycemia," Conference on computers in Cardiology, IEEE, pp. 537-540, 2003.
- [11] E.D Übeyli, "Feature extraction for analysis of ECG signals," 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008
- [12] F. Sufi, S. Mahmoud, I. Khalil, "A new ECG obfuscation method: A joint feature extraction & corruption approach," International Conference on Information Technology and Applications in Biomedicine, pp. 334-337, May 2008.
- [13] S. C. Saxena, A. Sharma, and S. C. Chaudhary, "Data compression and feature extraction of ECG signals," International Journal of Systems Science, vol. 28, no. 5, pp. 483-498, 1997.
- [14] Emran M. Tamil, Nor Hafeezah Kamarudin, RosliSalleh, M. Yamani IdnaIdris, Noorzaily M.Noor, and Azmi Mohd Tamil, "Heartbeat Electrocardiogram (ECG) Signal Feature Extraction Using Discrete Wavelet Transforms (DWT)."
- [15] E.D.Übeyli, "Detecting variabilities of ECG Signals by Lyapunov Exponents," Neural computing and applications, vol.18, no. 7, pp. 653-662, 2009.
- [16] Alan Jovic, and Nikola Bogunovic,"Feature Extraction for ECG Time- Series Mining based on Chaos Theory," Proceedings of 29thInternational Conference on Information Technology Interfaces, 2007.
- [17] Xiaomin Xu, and Ying Liu,"ECG QRS Complex Detection Using Slope Vector Waveform (SVW) Algorithm," Proceedings of the 26th Annual International Conference of the IEEE EMBS, pp. 3597-3600, 2004.
- [18] Y. H. Hu, S. Palreddy, and W. Tompkins, "A Patient Adaptable ECG Beat Classifier Using A Mixture Of Experts Approach", IEEE Transactions on Biomedical Engineering, vol. 44, pp. 891-900, 1997.
- [19] Costas Papaloukas, Dimitrios I. Fotiadis, Aristidis Likas, Lampros K. Michalis, "Automated Methods for Ischemia Detection in Long Duration ECGs", 2003.
- [20] S.Karpagachelvi, M.Arthanari, M.Sivakumar, "ECG Feature extraction techniques- A survey approach," International Journal of Computer Science and Information Security, vol. 8, No. 1, April 2010
- [21] Ashley EA, Niebauer J., "Conquering the ECG," London: Remedica; 2004.