IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Review of Methodologies for Classification of ECG by Artificial Intelligence

Ashwini Ashokrao Jagnade[1] , Ms. Sugandha Nandedkar[2]
[1]PG Student, Department of Computer Science and Engineering, Marathwada Shikshan Prasark Mandal's Deogiri Institute of Engineering & Management Studies, Aurangabad MH, India
[2]Professor, Department of Computer Science and Engineering, Marathwada Shikshan Prasark Mandal's Deogiri Institute of Engineering & Management Studies, Aurangabad MH, India

Abstract: The concept of pattern recognition relates to the categorization of data patterns and the differentiation of these patterns into a set of categories that have been predetermined. This is the analysis. Pattern recognition is put to use in the interpretation of ECG signals. The waveform that is formed by the ECG signal provides practically all of the information that is needed on the activity of the heart. It is a bioelectrical signal that is used in the process of documenting the electrical activity of the heart in relation to the passage of time. In order to properly diagnose cardiac disorders and choose the most effective course of therapy for a patient, early and accurate diagnosis is essential. ECG signals are the parameter that is utilised for the identification of cardiac disorders, and the majority of the data originates from the databases maintained by PhysioDataNet and MIT-BIH. In this research, the spectral entropy, Poincare plot, and Lyapunov exponent are some of the ECG signal feature extraction characteristics that are put under the microscope. Additionally, this study utilises an artificial neural network as a classifier for the purpose of determining the anomalies associated with cardiac disease.

Index Terms - **Artificial Neural Network, Electrocardiograph (ECG), Arrhythmia, feature extraction, Classification.**

I. Introduction

Electrocardiography is the study of the electrical impulses that are produced by the heart. Due to the nature of biosignals as non-stationary signals, the reflection may take place at any given moment in the time scale. As a result, in order to provide an accurate diagnosis, the ECG signal pattern and the variability of the heart rate would need to be watched for many hours. As a result of the vast amount of data being analysed, the research is both laborious and time consuming. Because of this, the analysis and categorization of heart disorders performed by a computer may be of great use in diagnostic work [1]. The method of recognising patterns on an ECG includes the following steps: pre-processing of the ECG signal; detection of the QRS; feature extraction; and usage of a neural network for signal classification. The early identification provides information about cardiac problems and increases the life expectancy of the human patient. ECG is used to measure the rate and regularity of heartbeats, as well as the size and position of the chambers, the presence of any damage to the heart, and the effects of drugs or devices used to regulate the heart [2]. In order to acquire the signal, ECG devices with varying numbers of electrodes (three to twelve) can be used [2]. The electrocardiogram may, to a certain extent, be broken down into phases that correspond to the depolarization and repolarization of the muscle fibres that make up the heart. The P wave, which represents atrial depolarization, and the QRS wave are the two waves that correlate to the depolarization stages (ventricles depolarization). T-wave and U-wave (ventricular repolarization) are the phases that correlate to the repolarization of the ventricles [3. An irregular heartbeat, also known as arrhythmia or dysrhythmia, is a symptom of a cardiac condition known as arrhythmia or dysrhythmia. This condition is caused by a problem with the heart's electrical system cells. It makes the heart pump blood less efficiently and produces disturbances in the mechanism through which the heart conducts electrical impulses [4]. The discovery of cardiac disease at an early stage is extremely beneficial, both for increasing the length of one's life expectancy and for increasing the development of our technology for detecting arrhythmias. Generally speaking, typical ECG signals are capable of being broken down into three distinct sets of fundamental components, as seen in figure 1.

Waves – deviations from the isoelectric line(baseline voltage) .they are named successively : P,Q,R,S,T,U;
Segments-isoelectric lines periods between waves.
Intervals-periods which include segments and waves.

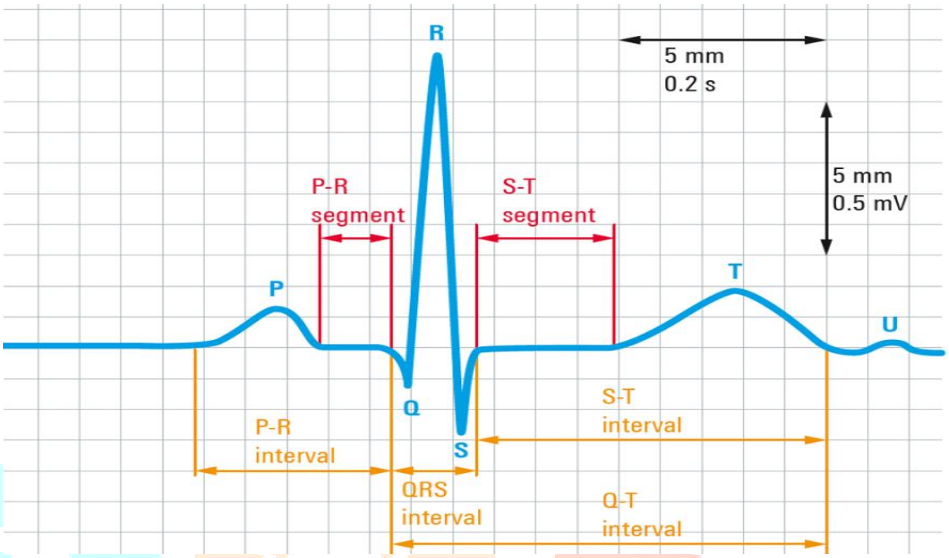

Figure 1: ECG signal

The analysis of an electrocardiogram (ECG) signal relies on the precise and consistent detection of the QRS complex, in addition to the T and P waves. When doing an automated analysis of an ECG signal, the identification of the QRS complex is the most critical job. After the QRS have been located, a more in-depth analysis of the ECG signal is feasible by incorporating the ST segment and the heart rate [5-6]. In the medical literature, a variety of different algorithms for the identification and categorization of electrocardiogram signals have been presented and explored.

Structure Details:

Page size: A4 size only
Text Column: Single **texts align: justify**
Title: 24pt Times New Roman align: centre
Page Margins: Left – 0.51”, Right – 0.51”, Top – 0.75”, Bottom – 0.75”
Font: Use Only Times New Roman for whole paper
Figure caption: Font size- 10”, lower case and Write below the figure, position-center
Table Caption: Font- 10”, lower case and Top of the table, position-center
Paragraph: Paragraph Indentation by- 0.2”
Line Spacing: single
Before: 0” **After:** 0”
Header 0.3” **footer** 0”

For this study secondary data has been collected. From the website of KSE the monthly stock prices for the sample firms are obtained from Jan 2010 to Dec 2014. And from the website of SBP the data for the macroeconomic variables are collected for the period of five years. The time series monthly data is collected on stock prices for sample firmsand relative macroeconomic variables for the period of 5 years. The data collection period is ranging from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance.

II. Literature review

In the body of academic research, many methods for the detection, feature extraction, and categorization of electrocardiogram (ECG) data have been created. In order to identify the key characteristics from a 12-lead electrocardiogram (ECG), Ramli et al. [7] made use of a method called cross correlation analysis. The detected values may be utilised to make predictions on the types of arrhythmias by using approaches that use cross correlation. An automatic classifier was developed by Tadejko and Rakowski [8] using the Kohonen self-organizing maps (SOM) and learning vector quantization (LVQ) techniques. In this study, a comparison is made between the categorization of QRS complexes using the original ECG morphology characteristics and the suggested novel technique, which uses the preprocessed ECG morphology features. The effectiveness of algorithms is evaluated based on their ability to identify beats as either normal or abnormal. In [9], Xu and colleagues presented the slope vector waveform (SVW) technique for evaluating RR intervals and detecting QRS complexes in electrocardiograms. In order to extract features from the ECG wave, they made use of the slope vector, and non-linear amplification was used in order to increase the signal-to-noise ratio. In this research, we offer a QRS detection method that has both high accuracy and a quick reaction. For the purpose of analysing ECG data, Manpreet Kaur and Arora [10] suggested using the K-means clustering algorithm in conjunction with the squared Euclidean distance. Wave form, duration, and amplitude were determined to be the most important characteristics for feature extraction. The approach of K clustering involves adding up all of the clustered Ks, which helps to reduce the total distance from each location to the centroid. Using the K clustering method, the first step will reveal information on the points that have been assigned to the closest cluster around the cetroid. In the second phase, information is provided on lines where the values have been self-resigned. Analyses are done to determine the success rate of categorization for a variety of datasets from MIT-BIH.

The research done by F. de Chazal et al. [23] demonstrates that premature ventricular contraction may result from both normal heartbeats and various types of cardiac disorders.

In order to extract characteristics from an electrocardiogram, a combination of features that are based on morphology and features that are based on time intervals has been suggested. For the purpose of ECG signal classification, the multilayer perceptron (MLP) with a variable number of hidden layers and an algorithm based on the radial basis function and probabilistic neural network are used. According to the findings of the simulation, an accuracy rate of around 97.14 percent may be attained when classifying ECG beats. The MIT-BIH arrhythmia database is used for doing simulations. In this particular piece of research, S. Mitra and colleagues [11] proposed a rough-set theory for the processing of ECG signals. In order to create an inference engine for illness detection, time-domain characteristics are used in the development of a rule-based rough-set decision system. By making use of rough set theory, one may optimise criteria for heart illness detection, so avoiding the complexity of NN and saving oneself some time and effort. At the moment, the system is being evaluated using three distinct variations of electrocardiogram (ECG) data: normal, myocardial ischemia, and myocardial infarction. In addition, the accuracy of all of these different kinds of datasets is determined by comparing them to an untrained dataset.

A wavelet transform approach was presented by Castro et al. [12] for the purpose of extracting features from an electrocardiogram (ECG) signal and identifying irregular heartbeats. This approach is useful for determining which association with the ECG signal is the strongest. After the electrocardiogram (ECG) signal has been denoised with either a soft or a hard threshold, the optimum wavelet function is used to divide each PQRST cycle into a coefficients vector. In order to produce a features vector of the signal cycles, the studied ECG signal coefficients are first subdivided into the P-wave, the QRS complex, and the T-wave, and then they are added together. The adaptive neuro-fuzzy inference system (ANFIS) algorithm was introduced by Nazmy et al. [13] for the purpose of ECG wave categorization. The RR interval of the electrocardiogram serves as the input for the feature extraction, which is carried out with the assistance of Independent Component Analysis (ICA) and Power spectrum. The ECG signals that are discussed in this article are categorised as follows: normal sinus rhythm (NSR), premature ventricular contraction (PVC), atrial premature contraction (APC), ventricular tachycardia (VT), ventricular fibrillation (VF), and supraventricular tachycardia (SVT) (SVT). The classification accuracy is also reached by the use of the ANFIS technique. Alan and Nikola introduced the chaos theory in their paper [14], which was used for ECG feature extraction. An explanation is given for a number of chaotic approaches, such as correlation dimension, phase space and attractors, central tendency measure, spatial filling index, and approximation entropy.

Yuksel and Bekir [17] have used ANN to categorise the various arrhythmias that might be seen on an electrocardiogram. Normal sinus rhythm, sinus bradycardia, ventricular tachycardia, sinus arrhythmia, atrial premature contraction, paced beat, right bundle branch block, left bundle branch block, atrial fibrillation, and atrial flutter have been considered types of arrhythmias that have been employed. After being filtered, these data had their R peaks located, and their patterns were normalised between 0 and 1. These patterns are used in the training of ANN both alone and in combination with a wide variety of arrhythmias. The authors Zhu et al. [18] explored the usage of artificial neural networks for the identification of ECG abnormalities. In this article, the SOM network, the BP network, and the LVQ network were used in order to examine the performance, and an overall accuracy of these networks was attained. Additionally, [19,20,21] offered a comparative research that looked at how a neural network classifies the patterns from training data and detects whether testing data contains that ECG signal patterns.

ANN model was suggested by El-Khafif et al. [19] to identify ischemic heart disease from normal ECG data. For the categorization of ECG data, they used a Feed forward Multilayer perceptron neural network with error back propagation learning technique. In the work that has been described, the use of slices derived from higher-order statistics demonstrates its utility in analysing and categorising nonlinear ECG patterns. In their study [22], Hosseini and colleagues suggested a two-stage feed forward neural network for the categorization of ECG signals. To identify cardiac irregularities, they have selected two network topologies, one based on one stage of feed forward neural networks and the other based on two stages of feed forward neural networks. In their study [28], Manimegalai et al. constructed a system that was based on the discrete wavelet transform for the purpose of detecting and extracting the P wave, QRS complex, and ST segment. And came to the conclusion that this method requires less computing time and yields superior accuracy when it comes to the categorization, analysis, and characterization of normal and pathological ECG patterns.

The neuro-fuzzy approach has been developed to simulate the experimental data in [25,26,27]. These references may be found here. Golpayegani and Jafari [29] offered a comparative study of the performance of MLP neural networks and ANFF Adaptive Neural Fuzzy Filters (ANFF). They observed that the training time of ANFF was much less than the time needed by MLP. In order to represent the chaotic character of various types of ECG signals, Owis et al. [30] have provided the correlation dimension and greatest lyapunov exponent parameters. These characteristics were computed for a large number of independent ECG signals coming from the MIT-BIH Arrhythmia Database by using the recommended implementations. These features were computed for all five distinct kinds of ECG signals. The findings are analysed in order to identify any statistically significant differences that may exist between the various forms of arrhythmia. Last but not least, statistical classification methods are used in order to evaluate the potential for identifying and categorising arrhythmias based on such data.

III RESEARCH METHODOLOGY

Warren McCulloch, a neurophysiologist, and Walter Pits, a logician, are credited as being the first people to develop a neural network in 1943. The power spectral density (PSD) is a statistical measure that may be derived from artificial neural networks (ANNs), a method that is also used in biological research. The power spectral density, or PSD, is a representation of the power distribution as a function of frequency. The probability density function is obtained by normalising the power spectral density with respect to the total spectral power (PDF). An estimate of the spectrum entropy of the process may be obtained by the use of Shannon's channel entropy. Entropy is provided by inspired networks, which are helpful in application fields like as pattern recognition, classification, and so on. The decision-making process of the ANN is holistic, and it is dependent on the characteristics of the patterns that are input. This makes the ANN an excellent candidate for the categorization of biological data. Typically, the generalised back propagation algorithm (BPA) is what's used to train multilayer feed forward neural networks so that they may function as non-linear classifiers [15]. The BPA is a kind of supervised learning algorithm. Within this type of learning, a mean square error function is constructed, and the objective of the learning process is to minimise the amount of error produced by the whole system. The weights of the connections are first chosen at random, and then they are gradually adjusted so as to lower the total amount of inaccuracy caused by the system. The process of updating the weights begins with the output layer and moves backward from there. In order to optimise the pace at which error is reduced, the weight update is being done in the direction of "negative descent." The step size is determined using a heuristic, and in this instance, a learning constant with the value $q = 0.9$ was selected. It is important for the training data set to have a homogeneous distribution throughout the class domains so that the training may be as successful as possible. It is possible to make use of the data at hand in an iterative manner until the error function has been minimised to its smallest possible value. Figure 1 depicts the ANN that was used for the classification process. The input layer was made up of nodes, while the succeeding hidden layers made use of process neurons that were activated using the sigmoid standard function. There were three neurons in the output layer, which allowed for the output domain to be divided into eight classes (000 to 111).

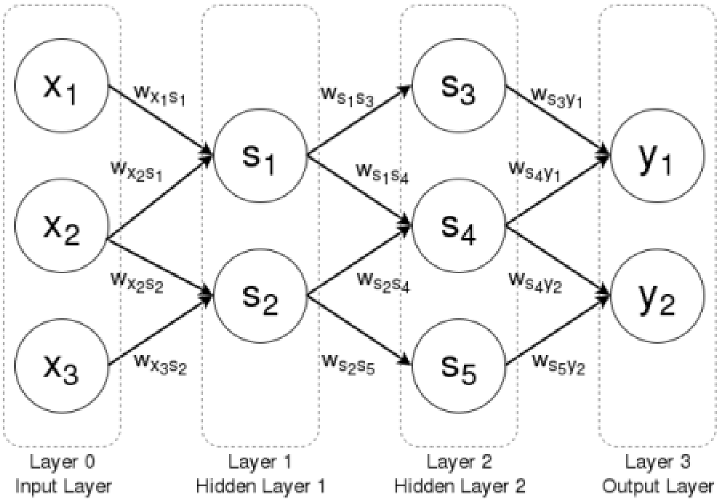

Figure 2: Four-layer feedforword neural network classifier

Disease Classification using ANN

The cardiac disorders were divided into eight categories for the purpose of this research. These categories are as follows: I left bundle branch block (LBBB); (ii) normal sinus rhythm (NSR); (iii) pre-ventricular contraction (PVC); (iv) atrial fibrillation (AF); (v) ventricular fibrillation (VF); (vi) complete heart block (CHB); (vii) ischaemic/dilated cardiomyopathy; and (viii) sick sinus syndrome (SSS). Three factors generated from the heart rate measurements were used to provide information into the ANN classifier. Spectral entropy, Poincare plot geometry, and the greatest Lyapunov exponent were the ones in question (LLE).

1 Spectral Entropy

Quantifying the spectral complexity of the time series is what the spectral entropy does. There are many different kinds of spectral transformations. The power spectral density (PSD) may be derived from a number of different methods; however, the Fourier transformation (FT) is the method that is used most often. The power spectral density, or PSD, is a representation of the power distribution as a function of frequency. The probability density function is obtained by normalising the power spectral density with respect to the total spectral power (PDF). An estimation of the spectrum entropy of the process may be obtained by the use of Shannon's channel entropy, where entropy is provided by

$$H = \sum_f p_f \log\left(\frac{1}{p_f}\right)$$

where pf is the PDF value at frequency f.

In a heuristic sense, the entropy is taken as a measure of the degree to which one is unclear about the event that occurred at f. Therefore, entropy may be used as a measurement for the complexity of a system. The complexity of the HRV signal is characterised by the value of the spectral entropy H(0=H=1), which ranges from 0 to 1. This spectrum entropy, denoted by H, was estimated for each of the distinct cardiac signal types.

2 Poincare Plot Geometry

The nature of R-R interval fluctuations may be explained with the help of Poincare plot geometry, a method that was borrowed from non-linear dynamics. Poincare plot geometry is a graph that shows each R-R interval projected against the following interval [30]. The Poincare plot analysis is a relatively new quantitative-visual method. Using this method, the form of the plot is classified into functional groups that reflect the extent to which a person is suffering from heart failure. We are able to gain both summary and comprehensive information on the behaviour of the heart by using plot. The summary information includes information on how the heart behaves between beats. Calculating the standard deviations of the distances of the R-R(i) to the lines y = x and y = -x 2*R-Rm, where R-Rm is the mean of all R-R, is one way to do a quantitative analysis of the Poincare plot. This may be done by entering the distances of the R-R(i) to the lines in question (i). Standard deviations 1 and 2 are abbreviated as SD1 and SD2, respectively. The variability of the data from beat to beat was defined by SD1, whereas the variability of R-R(i) over a longer period of time was described by SD2. It is also possible to calculate the ratio SD1/SD2 in order to explain the connection between these two components. The Poincar6 plot of a typical person is shown in Figure 2.

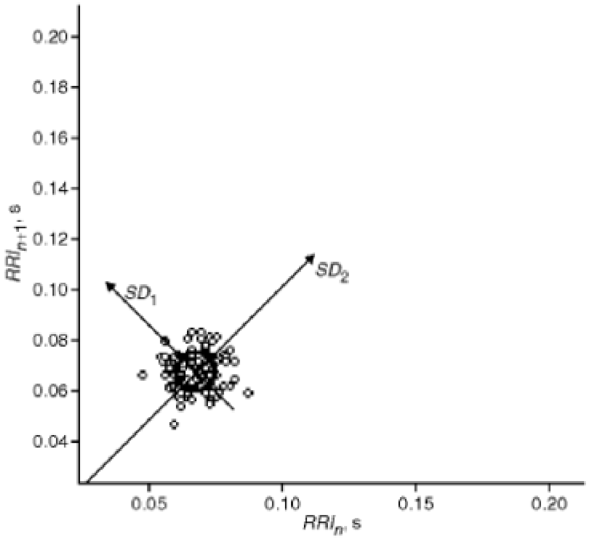

Figure 3: Poincare plot of a normal subject

3. Largest Lyapunov exponent

The rate at which the trajectories diverge from one another is quantified by the Lyapunov exponent, which is denoted by the symbol. A negative exponent indicates that the orbits are getting closer and closer to a central point [5]. When the exponent is 0, it indicates that the orbits are on a stable attractor since they keep their relative locations. Last but not least, the presence of a positive exponent suggests that the orbits are on a chaotic attractor. If there are two locations in a space, X0 and X0 + X0, both of which are functions of time and each of which will create an orbit in that space using some equations or system of equations, then the distance between the two orbits, denoted by the symbol x, will likewise be a function of time. This separation, which takes the form x, is also a function of the position of the beginning value (X0,t). When applied to a data set that is chaotic, the function x (X0,t) will exhibit unpredictable behaviour. Two orbits that were originally close together have a mean exponential rate of divergence that may be described as

$$\lambda = \quad \lim_{n\to\alpha} \quad \frac{1}{t} \quad \ln\left|\frac{\Delta x\ (Xo,t)}{\Delta Xo}\right|$$

The Lyapunov exponent is a valuable tool for differentiating between the different orbits. The Lyapunov exponent (LLE) with the highest value indicates how sensitive the system is to the beginning circumstances and provides a measure of how predictable the system is. When a positive Lyapunov exponent is present, it suggests that chaos is present. Even if a system with m dimensions has m Lyapunov exponents, in the majority of cases it is sufficient to merely calculate the LLE. Therefore, a technique that is reliable regardless of the length of the data is required. This approach searches for the point in phase space that is closest to each one, then measures how far apart those points become over the course of a specified amount of time. Estimating the LLE requires fitting the least squares model to an average line, which is described by

$$y(n) = \frac{1}{\Delta t}\Big(In\big(di(n)\big)\Big)$$

where di(n) represents the distance between the ith phase-space point and the point that is geographically closest to it at the nth time step. This last averaging phase is the primary component that enables an appropriate assessment of the LLE, despite the fact that the data that we have are both brief and noisy

IV. RESULTS AND DISCUSSION

4 The electrocardiogram is most often used in the process of diagnosing cardiac disease. In the research that has been done on this topic, a number of different models of supervised and unsupervised artificial neural networks have been presented for use in ECG signal feature extraction and categorization. The spectral entropy, Poincare plot geometry, and highest Lyapunov exponent (LLE) are the three parameters that were used to feed the ANN classifier. These parameters were produced from the heart rate data that were described. Analysis of the ECG signal, categorization of the ECG signal, and the extraction of an attribute from the ECG signal all play a very essential part in the process of detecting cardiac illness for the future. In addition, the subsequent improvement focuses on making use of a new approach that can classify and extract features with a better degree of precision.

Acknowledgment

I am profoundly grateful to Prof. S. C. Nandedkar for her expert guidance and continuous encouragement since it's commencement to its completion.I would like to express deepest appreciation towards Dr. Ulhas Shiurkar, Director, Deogiri institute of engineering and management Studies, Aurangabad. Prof. S. B. Kalyankar, Head, Department of Computer Science & Engineering. Dr. G. R. Gandhe, Head, PG Department whose invaluable guidance supported us in completing this course of work. At last, I must express my sincere heartfelt gratitude to all the staff members of Computer Science and Engineering Department who helped me directly or indirectly during this course of work.

References

[1] C.T. Lin, C.F. Juang, (2001) "An adaptive neural fuzzy filter and its applications," IEEE Transactions On Systems, MAN, And Cybernetics, VOL. 27, NO. 4, 1103-1110.

[2] Jenisha.J.Hannah and Suja Priyadharsini,2012. Patient Adaptive ECG Beat Classifier using Repetition Detection Approach Enhanced by Neural Networks, International Conference on Computing and Research and Control Engineering (ICCCE 2012), 12 & 13 ISBN:978-1-4675- 2248-9.

[3] K .O. Gupta and Dr. P. N. Chatur, 2012. ECG Signal Analysis and Classification using Data Mining and Artificial Neural Networks, International Journal of Emerging Technology Advanced Engineering, ISSN:2250-2459, 2(1) 56-60.

[4] J. T. Catalano, Guide to ECG analysis, Lippincott, 2nd edition

[5] Bucolo M., Grazia F. D., Sapuppo F., Nikolic D., Vuksanovic B., 2009. Multidimensional Analysis toward the Identification of ECG Nonlinear Dynamics. PHYSCON 2009. Catania, Italy.

[6] Y.C. Yeha, and W. J. Wang, 2008. QRS complexes detection for ECG signals The Difference Operation Method (DOM), Computer methods and programs in biomedicine, vol. 9, pp. 245–254.

[7] A. B. Ramli, and P. A. Ahmad, "Correlation analysis for abnormal ECG signal features extraction", 4th National Conference on Telecommunication Technology, 2003. NCTT 2003 Proceedings, pp. 232- 237

[8] 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",2007, CISIM '07, pp. 322-327.