ISSN: 1304-7981 Number: **12**, Year: **2016**, Pages: **205-219** **http://jnrs.gop.edu.tr**

Gaziosmanpasa University
Graduate School of Natural and Applied Sciences
Journal of New Results in Science

Received: 15.05.2016 *Invited Editor*: Mahmut Hekim
Accepted: 30.09.2016 *Area Editor*: Cem Emeksiz

Feature Extraction Based on Pan Tompkins Algorithm from ECG Signals and Diagnosis of Arrhythmia Using Multilayer Perceptron Neural Network

Ersin ERSOY[a] (ersin.ersoy.1987@gmail.com)
Mahmut HEKIM[b] (mahmut.hekim@gop.edu.tr)

[a] *Republic of Turkey Ministry of Foreign Affairs*
[b] *Gaziosmanpasa University, Department of Electrical and Electronics Engineering, 60150 Tokat*

In this study, Electrocardiogram (ECG) signals giving information about the state and functioning of the heart are divided into segments, waves and intervals by resting upon temporal limitations and feature vector of each section is obtained by means of arithmetic mean which is one of basic statistical parameters. Arrhythmia (rhythm disorders) occurring in the heart are diagnosed by the obtained feature vectors used as the inputs into multilayer perceptron neural network (MLPNN) model. For this purpose, ECG signals are divided into sections that are 10-minute-equal-length. These sections are divided into subsections (segment and intervals) which are admitted for each segment and wave interval and give information on arrhythmia by temporal limitations and arithmetic average of each interval is used as the inputs into the model of MLPNN for the diagnosis of arrhythmia. As a conclusion, it is proved that the proposed approach has reached to high accuracy rates of classification for the diagnosis of arrhythmia through ECG signals.

Keywords -
ECG signals, multilayer perceptron neural network, signal processing, diagnosis of disease, arrhythmia diagnosis

1. Introduction

Arrhythmia is a kind of disease with abnormal heartbeats. These types of abnormal heartbeats may lead to a rise or decrease in blood pressure and cause paralysis, stroke or even death. Cardiac arrhythmias are abnormalities or disorders in electrical behaviours of the heart. These disorders result in arrhythmia named as abnormalities in the rate and rhythm of heartbeat. Those intervals do not occur periodically or the facts that its initial and final durations are longer or shorter than some specific values are symptoms of arrhythmia.

This arrhythmia on Electrocardiogram (ECG) measurements spontaneously appear as observed deformations and irregularities in the form of waves. Arrhythmia emerges due to three reasons in general: psychiatric reasons, primarily due to physical and emotional stress and are cardiac causes [1, 2].

ECG signals are the ones which arise while the parts of the heart are being strained and relaxed. Thanks to these signals, one may get information regarding the functioning of the heart. There exist P, Q, R, S, T waves and PR, QRS, ST, QT, RR intervals on the ECG signal with 12 derivations. Arrhythmia can be diagnosed by the locations and durations of these waves and intervals. Six different types of QRS complex are observed in the signals and all the parts of QRS complex may not be seen at every deviation. Thus, QRS complex is a common classification. [1]. Figure1.1 shows ECG measurement parameters.

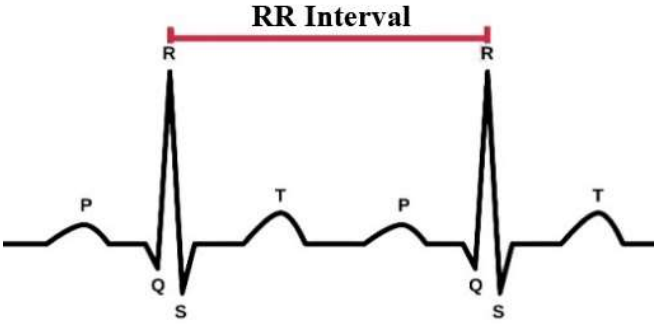


Figure1.1: ECG measurement parameters.

Artificial neural network (ANN) models are extensively used for the classification of ECG signals and disease diagnosis. In this study, ECG signals which give information related to the state and functioning of the heart have been discovered by means of arithmetic average that is divided into segments and waves and feature vector of each segment is one of basic statistical parameters by depending on limitations on time. Arrhythmia occurring in the heart is detected by the obtained feature vectors being used as the introduction into the model of MLPNN. For this purpose, ECG signals are divided into 10-minute-equal-length sections. These sections are divided into subsections (wave, segment and intervals) which are admitted for each segment and wave interval and give information on arrhythmia by the limitations on time and arithmetic average of each interval is used as the access into the model of MLPNN for the diagnosis of arrhythmia. As a conclusion, it is proved that the proposed approach has reached up to high-classification accuracy rates for the detection of arrhythmia through ECG signals.

2. Literature Review and Material

ANN is composed by plenty of cells, which are named as neurone, being connected to each other. ANN model is an artificial intelligence approach which can identify the relations between the given inputs and outputs of the samples and learn about these relations, and generalize over the obtained outcomes. Multilayer Perceptron Neural Network (MLPNN) is an ANN model that is commonly used because of its ease of use and high success rates at the classification of the signals. A number of ANN-based approaches among ECG signals have been used in order to detect arrhythmia: studies such as the development of a system based on machine learning with an aim of supporting the treatment processes of the patients with heart rhythm [1], the detection of cardiac arrhythmia on ECG signs via the

use of Ann [2], the comparison of standard ECG and Lewis ECG for the detection of A-V dissociation on the patients with wide QRS tachycardia[3], a fuzzy support vector machine approach for ECG analysis [4], the classification of ECG contractions with the help of GAL (Grow and Learn) network [5], the classification of ECG contractions by ANN and genetic algorithms being used [6] are carried out.

Definition 2.1. Rhythms formed during the functioning of the heart come into existence through electrical impulses appearing in the sinoatrial node. Firstly, atriums are strained with the help of electrical impulses and blood is transferred into ventricles and later on ventricles are strained and the blood in atriums is pumped into aorta and pulmonary artery through this process. Electrocardiogram, ECG in other words is acquired as a result of the fact that electrical variations, which occur in the heart in this way, are put on paper [1, 5].

Definition 2.2. Diseases resulting from the problems arising during the outlet or transmission of the electrical impulse are named as arrhythmia in general. Besides, formation of normal or abnormal impulse, abnormalities in impulse transmission or the variation of the normal rhythm depending on the fact that both originate together may also be defined as arrhythmia [7].

Definition 2.3. The first part of P wave is constituted by the depolarization of right atrium and the second part of it is constituted by the depolarization of left atrium. Normally, the width of P wave is less than 0.11 secs whichever derivation is discussed [7].

Definition 2.4. PR interval is gained through the measurement of duration between the beginning of P wave and the one of QRS complex. With adults, normal value for PR interval is between 0.12-0.20 seconds [7].

Definition 2.2. ST segment is the interval which combines junction point where QRS complex is over with the beginning of T wave. Its duration varies inversely proportional to heart rate and is between 0 and 0.15 secs. [7].

Definition 2.6. T wave reflects the re-polarization of the ventricles. With adults, the duration of normal T wave is between 0.10-0.25 secs.

Definition 2.7. RR interval equals distance between two R points.

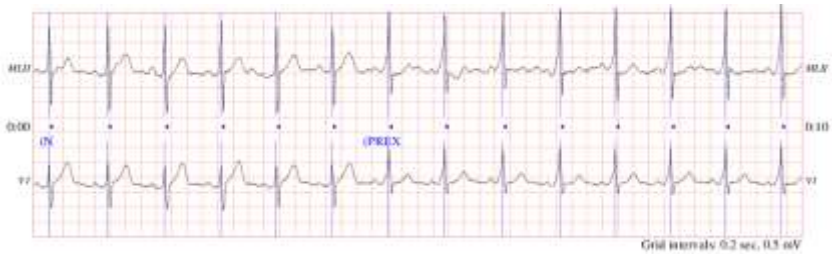
Definition 2.2. QT interval calculated through the measurement of the duration from the beginning of QRS complex until the end of T wave. QT interval amended in accordance with the heart rate is expressed as QTc. QTc is calculated through the division of QT duration into square root of RR duration (Bazett Index). According to Bazett Index, the upper bound of the amended QT interval (QTc_B) that is calculated is 0.44 secs and equation is calculated through 1:

$$QTB = \frac{QT}{\sqrt{RR}} \tag{1}$$

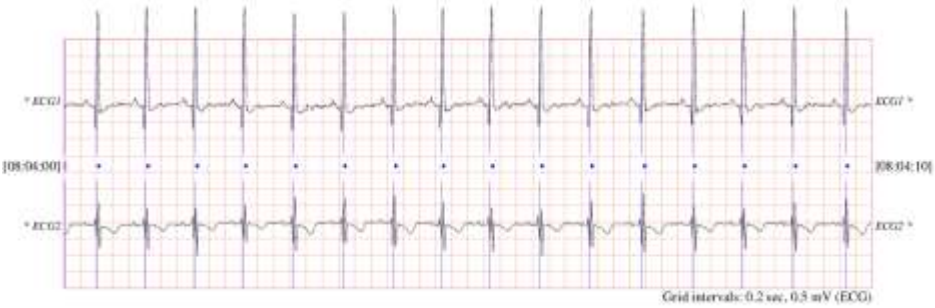
Hereby, it shows amended QT interval that is calculated with the use of QTc_B Bazett index.

2.1. ECG Signal

As ECG signal, "Physionet ECG databases" have been used. "MIT-BIH Normal Sinus Rhythm Database" [8] for normal ECG signal and "MIT-BIH Arrhythmia Database" for arrhythmia signal have been used [9]. Figure 2.1 shows 10-second- segments taken as an example from normal and arrhythmia ECG signals.



a) Normal (sine) ECG signal sample.



b) Arrhythmia ECG signal sample.

Figure 2.1: ECG signal (10 seconds).

Figure 2.2 shows P, Q, R, S, T heights and PR, QRS, ST, QT intervals of the waves observed at ECG signals of a healthy human being.

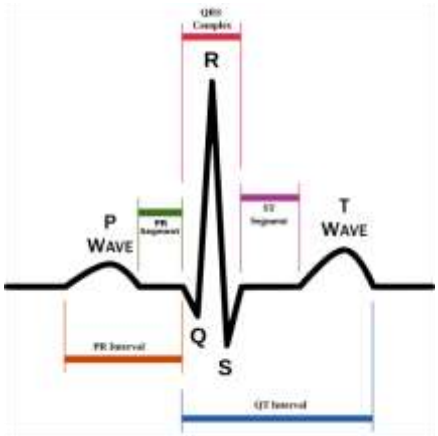


Figure 2.2: criteria for ECG signals.

As seen in Figure 2.2, ECG signal is characterized by repetitive wave sequence of P, QRS and T waves that are related with every heartbeat. QRS complex is the most distinctive one which is constituted through ventricular depolarization and atrial repolarization. As soon as

positions of QRS complexes are discovered, locations of ECG's other waves such as P,T waves and QT, ST segments are determined by the position of QRS complexes in order to analyse cardiac period in its entirety.

In this study, average-mean-based key feature vectors of P, PR, QRS, QT, ST, T and RR intervals of ECG signals have been calculated with the use of their temporal distance to R point. Pan-Tompkins algorithm is used for the identification of R point at ECG signals. Pan-Tompkins algorithm consists of five stages: band pass filter, differentiator, square receptor, sliding window integration, threshold adjustment [10]. Rough 10-second normal ECG signal is seen in Figure 2.3.

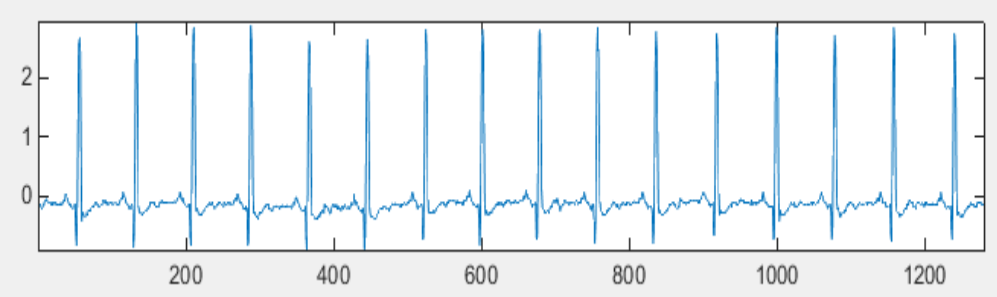


Figure 2.3: Normal ECG signal.

2.1. Feature extraction based on the temporal intervals using Pan Tompkins algorithm

The first stage of Pan Tompkins algorithm applying band pass filter to determine the noise within ECG signals. Figure 2.4 shows the 10-second part of the signal whose noise has been reduced by means of band pass filter.

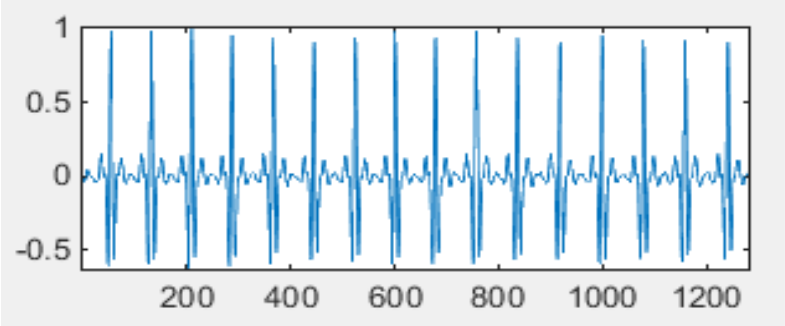


Figure 2.4: ECG signal screened out through band pass filter.

Band pass filter used in Pan Tompkins algorithm has been attained through low pass filter and high pass filter. Sampling frequency for high pass filter equals 200 Hz and critical frequency is 11 Hz and sliding quantity equals 5 samples, namely 25ms. Critical frequency of high pass filter is 200 Hz and sampling frequency is 5Hz and sliding quantity equals 16 samples, 80 ms in other words [11-12].

ECG signal that is filtrated has been applied to differentiator in order that QRS can become evident and low-frequency components can be supressed under derivation and ECG signal that is low-frequency-component free, which is shown below, has been obtained:

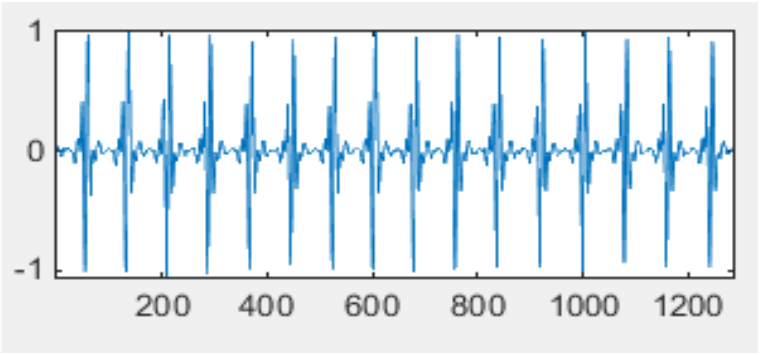


Figure 2.5: ECG signal to which derivation filter has been applied.

Finally, as is seen in Figure 2.6, smoothing procedure is carried out by means of square receptor and sliding window integration [2].

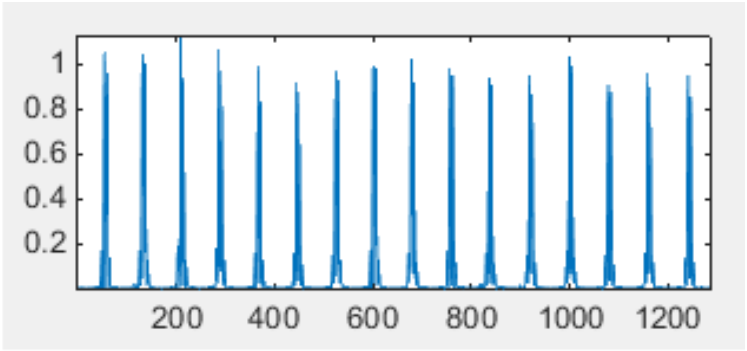
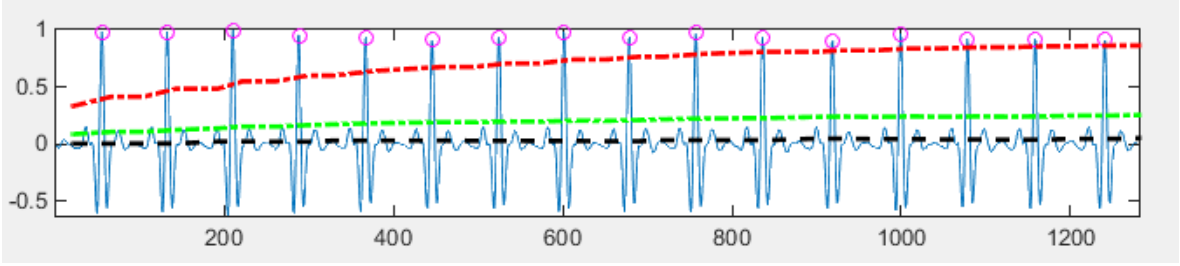
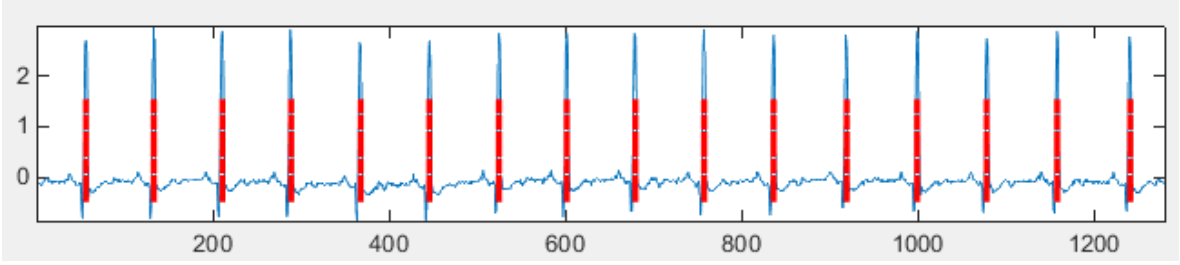


Figure 2.6: Squared ECG signal.

Figure 2.7 shows QRS points that are identified through the use of Pan-Tompkins algorithm.



a) QRS points detected on ECG signal filtered through Pan-Tompkins algorithm.



b) QRS points shown on Standard ECG signal.

Figure 2.7: Identification of QRS points.

In this study, signal groups of the waves at ECG signal have been formed in accordance with their temporal intervals after R points at QRS segment were determined as is seen in

Figure 2.7 and the average of temporal distances up to R points, which are identified, has been calculated as below:

Step 1. Deviance at RR intervals: The average of all RR points (RR_{avg}) on the signal has been worked out and the difference between every RR block distance and the calculated average has been found out. The resulting difference reveals that R points periodically continue. If the resulting difference is high, it means that R points are not periodically formed. The averages of RR internal deviations are calculated through Equation 2 shown below.

$$RR_{deviation} = \sum_{i=1}^{RR_{final}} \frac{|RR_i - RR_{avg}|}{RR_{number}} \quad (2)$$

Hereby, RR_{avg} indicates the average of all RR intervals and RR_{number} demonstrates the number of all RR intervals and $RR_{deviation}$ points out the average of the difference between all RR intervals and the calculated RR_{avg} value.

Step 2. QRS complex: Q interval cannot exceed 25 % of total QRS and total duration of QRS cannot go beyond 0.11 secs. Besides, it should be Q<0.04 secs. Suppose that R point equals the middle of QRS block,

QRS_{half} : It is temporally half of QRS wave.
R_{half} : It is temporally half of R wave.
QRS_{half} = 0.11/2 = 0.055
Q= 0.11/4 = 0.0275 (It is also in conformity with Q<0,04 requirement).
If $R_{half} = QRS_{half} - Q$, R_{half} = 0.055-0.0275 = 0.0275
R=2 R_{half} =0.055 and
If S = QRS – S – R, S=0.0275.

In Figure 2.8, the intervals used for the calculation of QRS complex have been demonstrated.

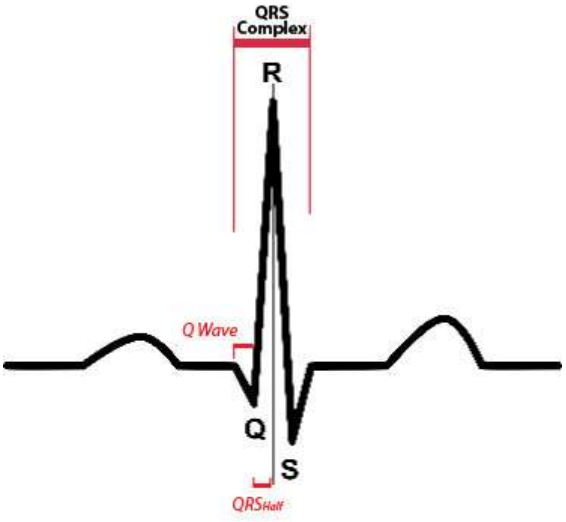


Figure 2.8 : Intervals used for the calculation of QRS complex.

Step 3. Temporal distances from intervals to R point:

$P_{initial}$: The temporal distance from P wave initial point to R point.
P_{final} : The distance from P wave final point to R point.
$PR_{initial}$: The temporal distance from PR interval initial point to R point.
PR_{final} : The temporal distance from PR interval final point to R point.
$Q_{initial}$: The temporal distance from Q interval initial point to R point.
S_{final} : The temporal distance from S interval final point to R point.
$ST_{initial}$: The temporal distance from ST segment initial point to R point.
ST_{final} : The temporal distance from ST segment final point to R point.
$T_{initial}$: The temporal distance from T wave initial point to R point.
T_{final} : The temporal distance from T wave final point to R point.
$QTc_{initial}$: The temporal distance from the modified QT interval initial point to R point.
QTc_{final} : The temporal distance from the modified QT interval final point to R point.
R_{half} : R wave is temporally half of the width.

Step 4. P interval: In Figure 2.9, it is demonstrated how to work out the distance from P interval to R point.

If P_{final} = PR + QR, 0.2 + 0.55= 0.255 secs.
$P_{final} = P_{initial} - 0.10 = 0.155$ secs.

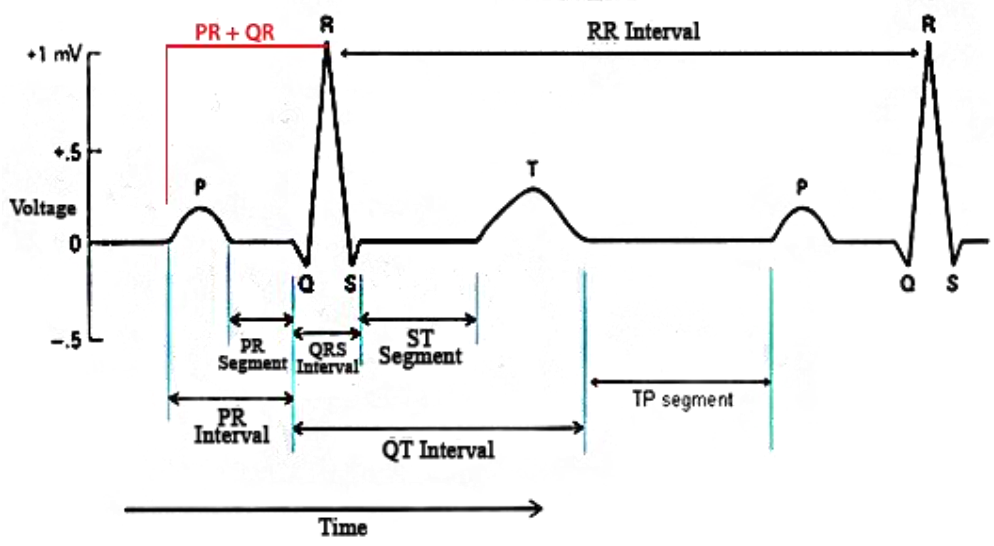


Figure 2.9: Calculation of P interval

Step 5. PR interval: In Figure 2.10, the ranges within the calculation of PR interval are indicated.

$PR_{initial} = P_{initial}$
$PR_{final} = PR_{initial} - 0.2 = 0.055$ secs.

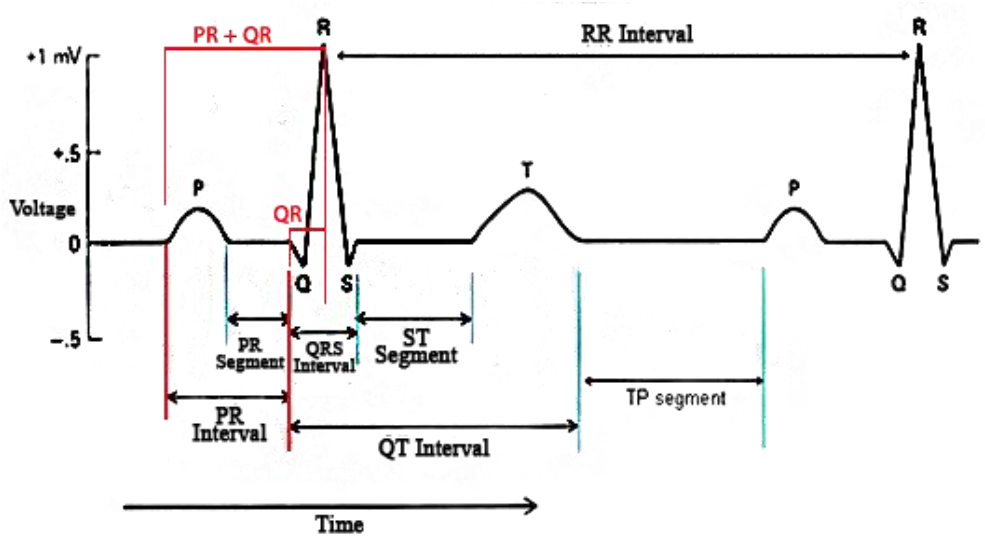


Figure 10: Calculation of PR interval

Step 6. QRS interval: In figure 2.11, QRS interval is shown.

$Q_{initial}$ = 0.55 secs
S_{final} = 0.55 secs

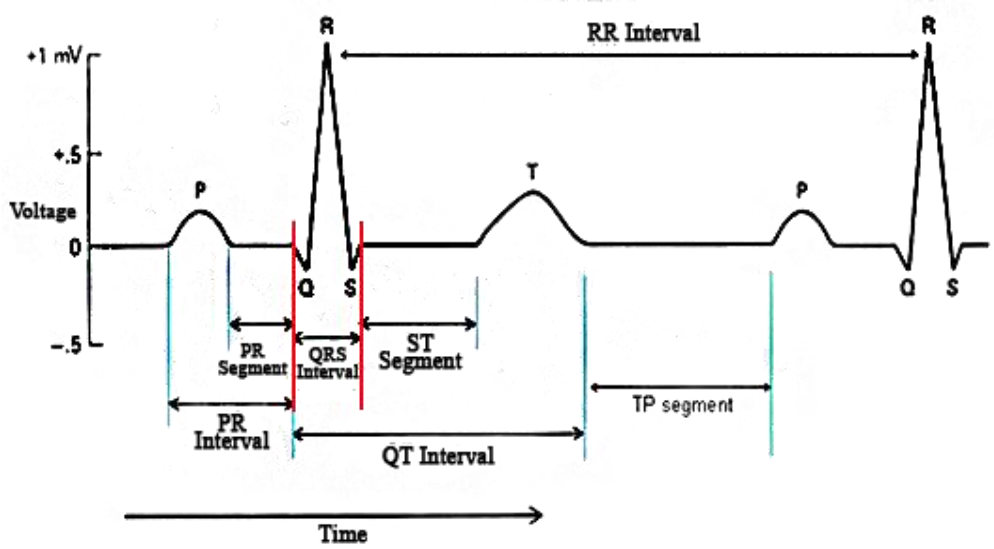


Figure 2.11: Calculation of QRS interval

Step 7. ST segment length: In Figure 2.12, intervals for the calculation of ST segment length are demonstrated.

$ST_{initial} = S_{final}$
If $ST_{final} = ST_{initial} + 0.15$,
ST_{final} = 0.205 secs.

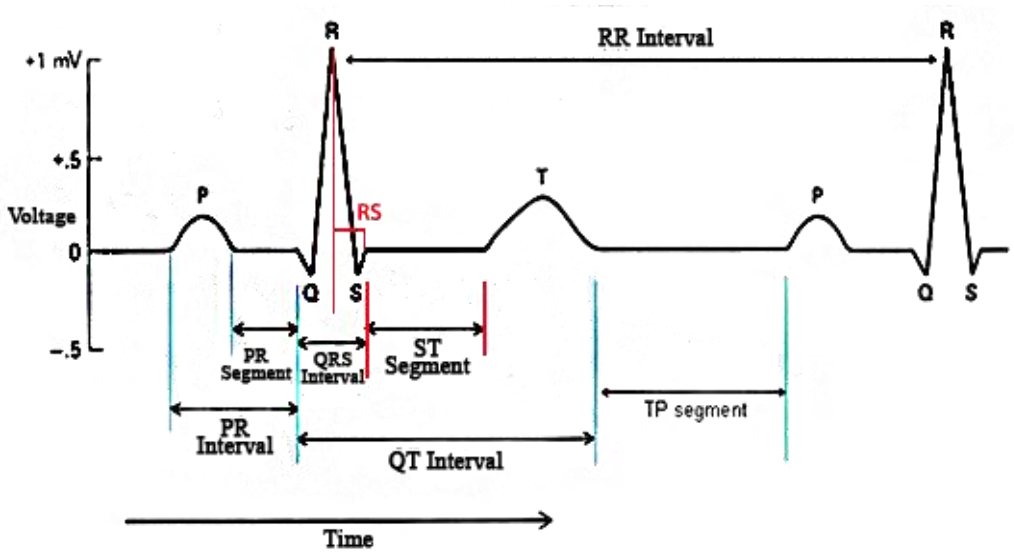


Figure 2.12: Calculation of the length of ST segment.

.

Step 8. T interval: In Figure 2.13, intervals for the calculation of T wave are exhibited.

$T_{initial} = ST_{final}$
If $T_{final} = T_{initial} + 0.25$, $T_{final} = 0.455$ secs.

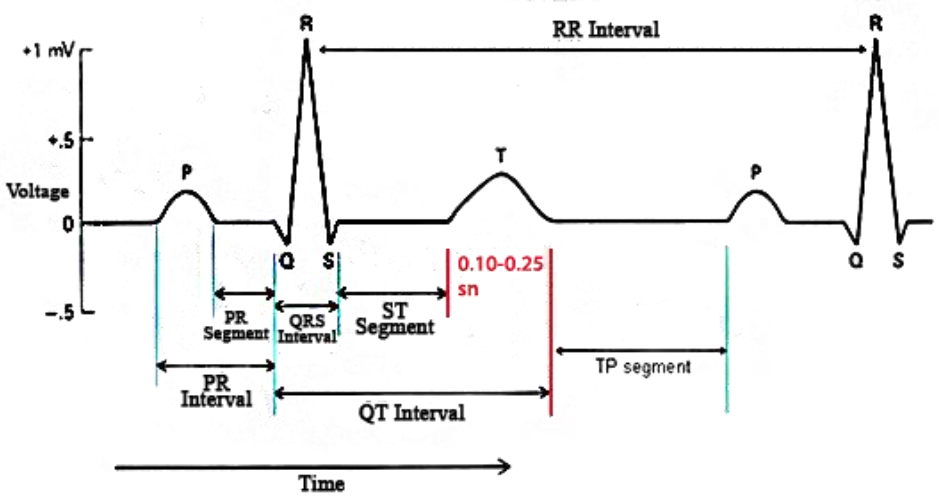


Figure 2.13: Calculation of T wave.

Step 9. QTc interval: QTc interval is counted according to Bazett index. In Figure 2.14, intervals for the calculation of QT interval are indicated.

$QTc_{initial} = Q_{final}$
$QTc_{final} = QTc - R_{half}$

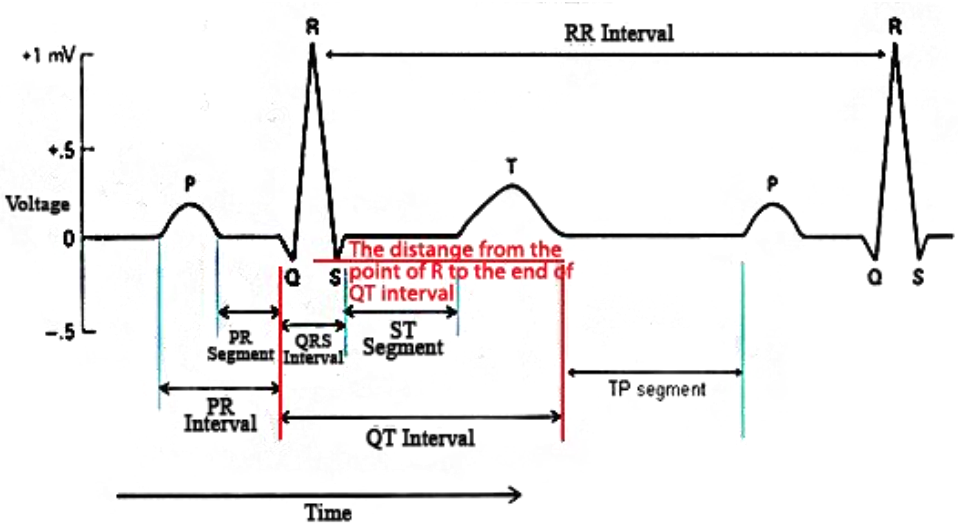


Figure 2.14: Calculation of QT interval.

According to the operation steps cited above, averages of all waves and intervals have been counted. As a result of the calculation, 90 arrhythmia have been observed and 90 usual sine rhythms have been recorded totally and a total of 180 signal segments have been used.

2.2. Formation and training of MLPNN model

A MLPNN model is composed of three sequential layers: input layer, hidden layer, and output layer. The number of neurons on input layer equals the number of the features used. The number of the neurons on output layer, on the other hand, reflects the features required. Hidden layers used may optionally be increased or decreased in order to enhance the capability of the network. Hidden layer functions as classifier and transmits the values on input layer to output layer as a result. On one hand, the increase of the number of the neurons on hidden layer can improve the network's capability but it may adversely affect the operation duration [13] on the other hand. Figure 2.15 reveals the block diagram of a MLPNN model.

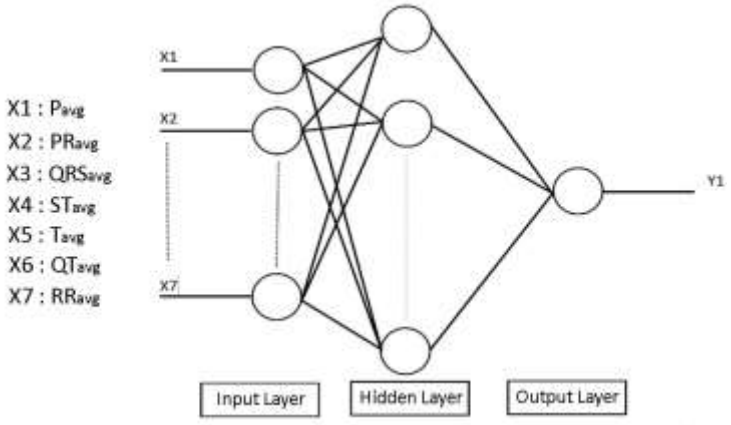


Figure 2.15: MLPNN based classifier model.

As seen in Figure 2.15, 7 feature vectors including P_{avg}, PR_{avg}, QRS_{avg}, ST_{avg}, T_{avg}, QT_{avg} and RR_{and} have been used as an access into MLPNN model whose hidden layer includes 10 neurons for the determination of arrhythmia through ECG signals. Classifier model Levenberg-Marquardt (LM) whose activation function has been set as tangent-hyperbolic

activation function has been trained with backpropagation algorithm. Classifier model has been operated hundred times and the final result has been worked out by the average of acquired classification success measures being calculated. For training, 90 feature vectors pertaining to normal signal and 90 feature vectors belonging to arrhythmia signal have been implemented on the classifier model. Examples of the feature values used as the inputs to MLPNN model is viewed in Figures 2.16-17.

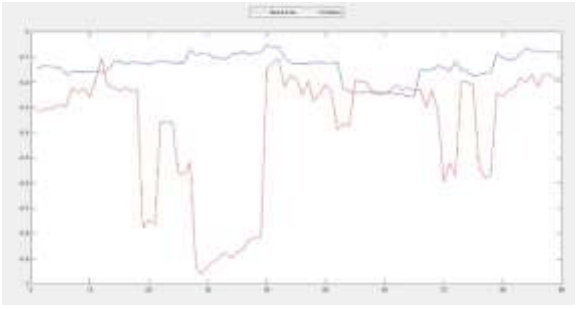


Figure 2.16: P_{avg} values obtained from normal signals and signals with arrhythmia.

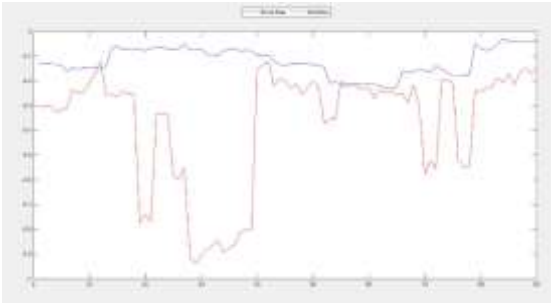


Figure 2.17: PR_{avg} values obtained from normal signals and signals with arrhythmia.

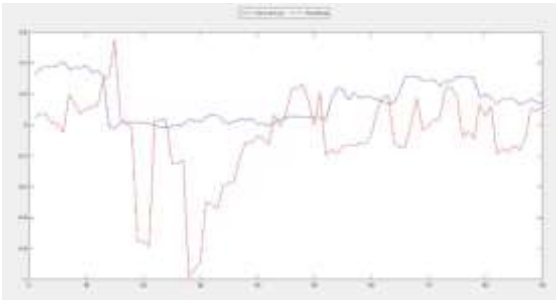


Figure 2.18: QRS_{avg} values obtained from normal signals and signals with arrhythmia.

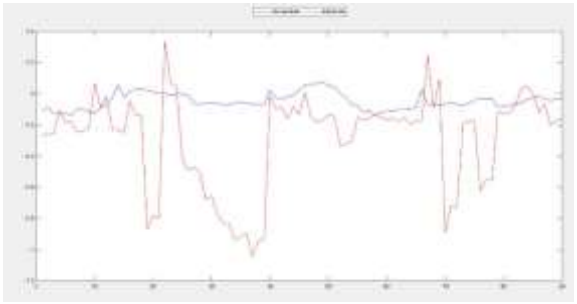


Figure 2.19: QT_{avg} values obtained from normal signals and signals with arrhythmia.

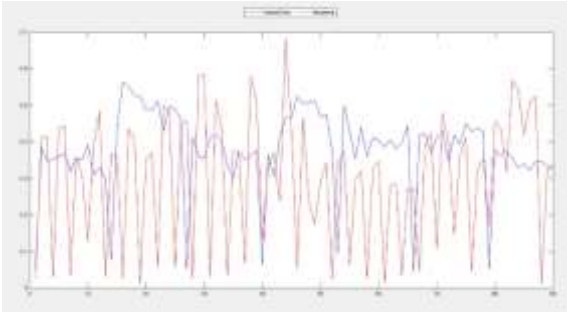


Figure 2.20: RR_{avg} values obtained from normal signals and signals with arrhythmia.

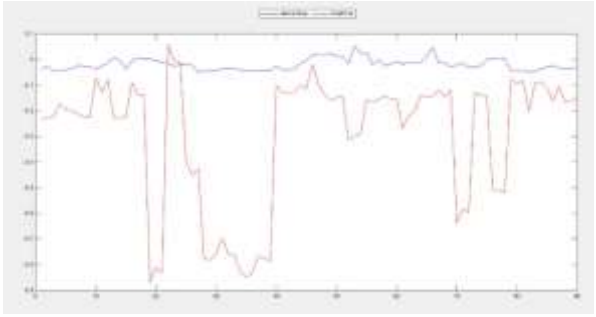


Figure 2.21: ST_{avg} values obtained from normal signals and signals with arrhythmia.

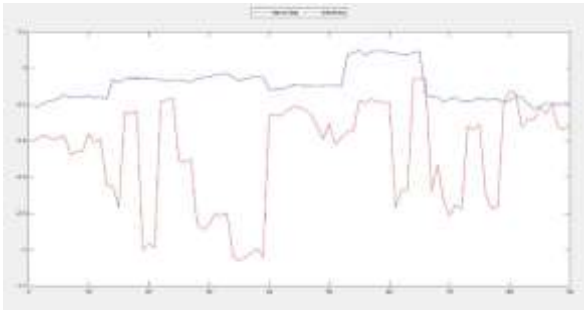


Figure 2.22: T_{avg} values obtained from normal signals and signals with arrhythmia.

10-fold cross validation criterion that is based on random sample selection has been used so as to measure the generalised success of the classifier. In this method, the feature vectors, which are acquired, are randomly distributed into three groups consisting of training, validity and test data. Training data have been selected in a way that it will include 70 % of all data (126 samples) and validity and test data, on the other hand, have been selected in a way that it will contain 15 % of it (54 samples). Once the model's success in the validity data has reached the highest level, the training has been finalised. Ultimately, model's final accurate classification success has been evaluated by means of statistical criteria on test data. The major criteria for this evaluation are certainty, sensitivity, total classification accuracy criteria. Certainty criterion is the ratio of the number of individuals classified as AP to the number of individuals that is actually classified as positive (AP+FN) and sensitivity criterion, on the other hand, is the ratio of the number of individuals classified as AN to the number of individuals that is actually classified as negative (AN+FP) and lastly, criterion of total accurate classification (TAC) is the ratio of the number of individuals classified accurately (AP+AN) to the total number of individuals (AP+AN+FP+FN).

$$Certainty = \frac{AP}{AP+FN} \tag{3}$$

$$Sensitivity = \frac{AN}{AN+FP} \tag{4}$$

$$TAC = \frac{AP+AN}{AP+AN+FP+FN} \tag{5}$$

Hereby, AP (Accurate Positive) indicates the number of individuals who have been classified accurately and diagnosed with arrhythmia and AN (Accurate Negative) points the number of individuals who have been classified accurately and do not have arrhythmia and FP (False Positive) reflects the number of individuals who have been inaccurately diagnosed with arrhythmia but do not have arrhythmia and FN (False Negative), on the other hand, refers to the number of individuals who have been inaccurately diagnosed with no arrhythmia but do have arrhythmia.

3. Evaluation of the findings

In case the class-based distributions of the numbers obtained from sample data are highly diverse and the success rate is high, confusion matrix and ROC curve analysis are utilised to be able to carry out success evaluation [13-15]. Confusion matrix that has been obtained

as a result of the classification of ECG signals with the model recommended is demonstrated in Table 1.

Table 1: Confusion matrix for the diagnosis of arrhythmia

	Negative (Healthy)	Positive (Arrhythmia)
Negative (Healthy)	14 (AP)	0 (FN)
Positive (Arrhythmia)	1 (FP)	12 (AN)

While inaccurate classification for the diagnosis of arrhythmia is not carried out with the proposed approach, in the very unlikely event that an inaccurate classification for a healthy individual with no arrhythmia may be performed as is seen in Table 1. A certainty rate of 93,3, a sensitivity rate of 100 % and a TAC rate of 96,3 % have been calculated with the use of equation 3,4 and 5.Thispoints out that the classifier has high rates of success. Figure 3.1 demonstrates ROC curve of the classification test performed for the diagnosis of arrhythmia through ECG signals.

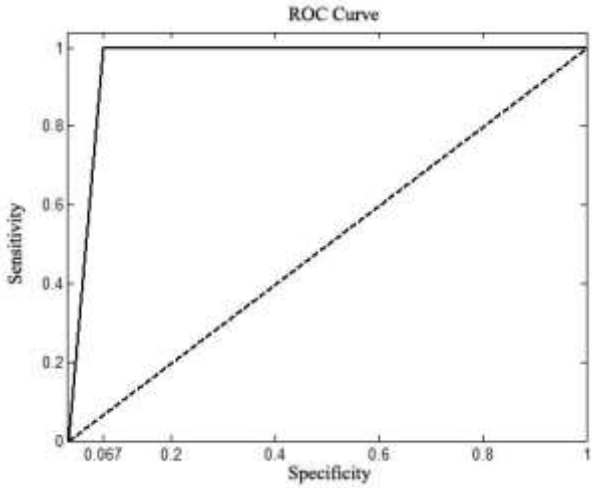


Figure 3.1: ROC curve of the classification.

As seen in Figure 3.1, the proposed approach has an acceptable classification capability for the diagnosis of arrhythmia according to ROC curve analysis. Accordingly, the fact that it displays broad spaces under ROC curves reveals that it is a classifier model with high certainty and sensitivity.

4. Results

High success rates of classification have been achieved when feature vectors obtained from the signals, which are divided into temporal segments and waves, thanks to arithmetic average have been used as an access into a MLPNN model for the diagnosis of arrhythmia through ECG signals. In addition to the intervals used in the studies of the diagnosis of arrhythmia on the literature, all the waves, segments and intervals occurring when a heart is strained and relaxed have been used as access into MLPNN. It is revealed that ECG signals' segment and wave intervals, which pertain to the parts that are divided into 10-

minute equal-length sections, are indicative of a significant feature vector for arrhythmia diagnosis. In conclusion, it is a crucial finding because of the fact that arrhythmia can be diagnosed with the help of high accuracy rates of classification after the statistical features of the wave intervals of ECG signals that are resolved into segments are implemented on a ANN-based classifier model.

References

[1] F. Aydin, *Developing a machine learning based system to assist treatment processes of arrhythmia patients*, Trakya University Institute of Science and Computer Engineering, 2011.

[2] B. Gumus, S. Yazgi, *Automatic detection of cardiac arrhythmia ECG signals using artificial neural network*, In: Proceeding of Electrical-Electronics-Computer and Biomedical Engineering 13th National Congress, 23-26 2009 (*in Turkish*).

[3] U. Aksu, *Comparison of standard and lewis ECG in detection of A-V dissociation in patients with wide QRS tachycardia*, Atatürk University Faculty of Medicine, Department of Cardiology, 2014.

[4] N. Ozcan, *A fuzzy support vector machine approach for ECG analysis*, Bogazici University in Systems and Control Engineering, 2010.

[5] M. Korueki, S. Metin, *Classification of electrocardiogram beats using GAL network*, Istanbul Technical University, Institute of Science and Technology, 2002.

[6] Z. Dokur, *Classification of ECG beats by using artificial neural networks and genetic algorithms*, Istanbul Technical University Institute of Science Electrical and Electronics Engineering Department, 1999.

[7] B. Ilerigelen, H. Mutlu, *Continuing Medical Education Faculty of Cerrahpaşa Medical Techniques*, ECG course booklet, http://www.ctf.edu.tr/stek/EKG_Kurs_Kitap.pdf (*in Turkish*).

[8] *Physionet, MIT-BIH Normal sinus rhythm Database* http://www.physionet.org/physiobank/database/nsrdb/ (May 10, 2016).

[9] *Physionet, MIT-BIH Arrhythmia Database*, http://www.physionet.org/physiobank/database/mitdb/ (May 10, 2016).

[10] S.L. Pingale, N. Daimiwal, *Detection of various diseases using ECG signal in Matlab*, International Journal of Recent Technology and Engineering. ISSN: 2277-3878, 2014.

[11] J.H. Indik, E.C. Pearson, K. Fried, R.L. Woosley, *B.* Fridericia, *QT correction formulas interfere with measurement of drug-induced changes in QT interval*, Sarver Heart Center of University of Arizona College of Medicine, 2006.

[12] J. Pan, W.J. Tompkins, *A real-time QRS detection algorithm*, IEEE Transactions on Biomedical Engineering, BME-32 (3) 230-236, 1985.

[13] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press Inc. New York, NY, USA 1995.

[14] M.W. Gardner, S.R. Dorling, *Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences*, School of Environmental Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, 1998.

[15] M. Hekim, U. Orhan, M. Ozer, *Epileptic seizure detection using artificial neural network and a new feature extraction approach based on equal width discretization*, Gazi University Journal of the Faculty of Engineering and Architecture, 26(3) 575-580, 2011.