# Disease Classification from ECG Signal Using R-Peak Analysis with Artificial Intelligence

Pooja Sharma[1] and D.V Gupta[2]

*Assistant Professor, Seth Jai Parkash Mukand Lal college of Engg. & Technology(JMIT)*
*Professor, Department of Mathematics, College of Engg. Roorkee(COER)*
*[1]Pooja_100@jmit.ac.in, [2]Dvgupta.rke@gmail.com*

## *Abstract*

*Electrocardiogram (ECG) is a non-invasive technique which is used as a main diagnostic device for cardiovascular diseases. A cleaned ECG signal offers essential information for electrophysiology for the heart diseases and ischemic changes that might happen. It gives significant information for the functional aspects of the heart and cardiovascular system. In this research, the detection of cardiac arrhythmias in the ECG signal consists of: pre-processing using DWT, detection of QRS complex in ECG signal; feature extraction from detected QRS complexes; classification of beats using extracted feature set from QRS complexes. To see the presence of cardiovascular disease in the R-peaks, GA (Genetic algorithm) and NN (Neural network) are considered. Initially, the extraction of the R peaks are done precisely through feature extraction and then they are optimized to have the small value of actual R peaks through which the user can view the presence of disease with the variation in the beat. To check the performance of the proposed work, varied parameters, viz. Precision, recall, F-measure, Accuracy, Error and time are analyzed and calculated.*

## 1. Introduction

In the recent development in the modern technologies, sensors, wireless communication technologies and efficient signal processing tools allows storing and transmitting of the biomedical signals and waves for diagnosis of patient diseases [1]. ECG defines heart's electrical action which shows regular reduction and easing of the muscles. The tool which is taken as support for diagnosis of condition of heart is known as electrocardiography. Various heart abnormalities are detected by the ECG waveform [2]. The processing technique of signal contains (a) de-noising (b) parameter extraction (c) baseline correction and (d) detection of arrhythmia. An ECG electrical wave (waveform) is composed of mainly 5 signal strokes M, N, P, Q and T waves and sometimes U waveform [3]. Every action is defined in the form of waves which are normally different every time. The waveforms are the statistical description of the disturbance generated by arties and ventricles. The tool is mainly in the hospitals for diagnosis and prediction of the heart abnormities. The signals of the ECG consist of many notches; hence the important ECG signal is extracted [4]. The signal processing is divided into two parts: (a) pre-processing (b) feature extraction. In the pre-processing phase, contaminations in the signal are minimized like power line interference, noises, baseline wandering, and artifacts of the patients. In the next phase of feature extraction, amplitude and spacing functions are used for automatic analysis.

The peak signals uses five different forms of waves and three pulses which are P wave, QRS complex and T waves. P peak describes the atrial depolarization. QRS defines the ventricular depolarization; T wave is referred to ventricle re-polarization [5]. The non-functioning of heart alters the waveform of the QRS complex. During deep injuries, the deep QRS waves occur and in smaller injuries, Q-waves are less deep. The main component of the ECG waveform is PRS complex. The pacemaker cells in the sinoatrial (SA) node are used to produce and control the heart rhythm that is situated at the right atrium top. Atrial depolarization is continuously monitored by ventricular depolarization. In ST (Sinus Tachycardia), the making rate is quicker where as in the SB (Sinus Bradycardia) the creation rate is in smaller amount, but the wave will be directed by the usual route [6].
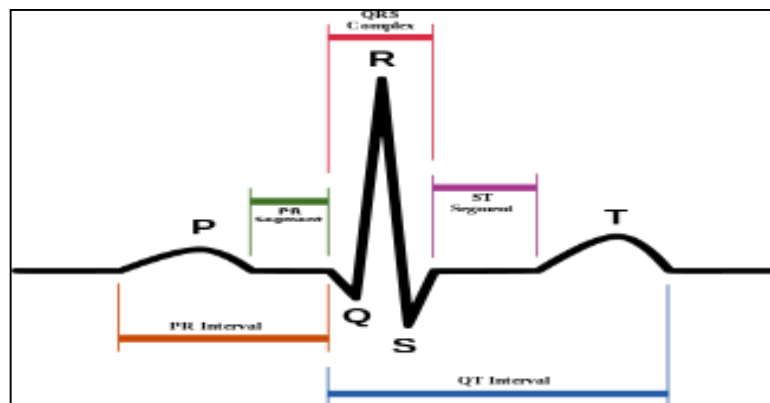


**Figure 1. ECG Waveform Representation**

The variability in the heart rate is a natural occurrence of change in the time period amid the heartbeat which is mainly computed in the beat to beat interval. In this, R is the point of the QRS complex of the electrocardiograph wave [7]. The RR interval is generally between the succeeding R's. The methods which are normally used to measure the heart beat are ECG, ballistocardiograms, blood pressure and pulse wave signal which is taken from the PPF (photo plethysmograph). ECG is deliberated as higher as it is offers more clear waveform, and excludes heart-beats which does-not originate in the sinoatrial node. As the processed beats are basically normal beats so NN is taken in place of RR. There are time intervals in ECG waveform as defined below [8]:

**Table 1. ECG Time Intervals**

| Time interval | Description |
|---|---|
| PR interval | It depicts the interval from initialization of artial depolarization initializing ventricular depolarization. |
| PR segment | It begins from the P end wave and finishes at the beginning of QRS complex. |
| | It shows the time occupied by the impulse from sinus node towards ventricle. |
| | The value ranges from 0.12 to 0.20 seconds. |
| | If the value of PR interval is less as compared to 0.12 seconds than it is taken as PR interval. |
| | The PR interval lessens in the duration by means of distributed beats as the impulse doesn't have much to travel. |
| | If the range of PR interval is larger than 0.20 seconds than the blockage of atrioventricular |

| QR interval | It defines the time from depolarization towards the re-polarization ventricles. |
| --- | --- |
| | It begins at the QRS complex onset and finishes at T wave end point. The variation of QT peak differs with the heart rate variation. |
| Short QT interval | This wave interval occurs in hyperthyroidism, hypercalcaemia, hyperkalaemia, and digital is toxicity and class IB drugs. |
| Long QT interval | It occur when the heart rate is less, class IA and III drugs and coronary heart failurs. |

RR interval values are accessed to see the level of different diseases as shown in Figure 2. In the previous studies, RR intervals are taken for middle-aged person when the myocardial infarction occurred to see the presence of coronary heart disease which is due to the inconsistency in the R-R interval [9] The values of ultralow-frequency power, low frequency, very low frequency, HF, LF and the ratio of (LF / HF) power are calculated. The main disease due to the substantially lower RR variability is chronic coronary heart disease.
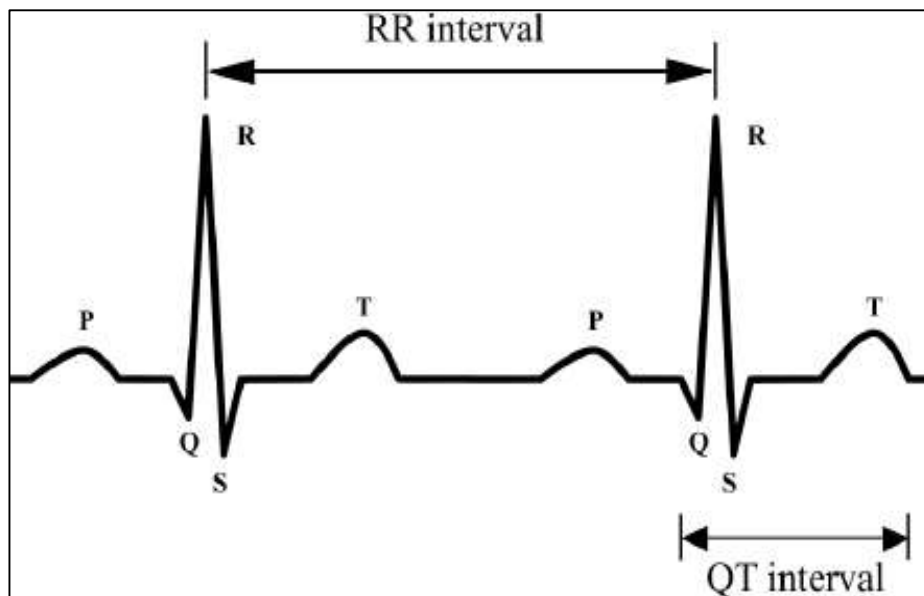


**Figure 2. R to R Interval**

When any discomfort in the shoulder, jaw, arm or chest pain may cause myocardial infarction which is usually known as heart attack [10]. The pain often occurs in the left side of chest which can last for over few minutes. The discomfort can relate to hurt burn. Other indications may consist of shortness of breath, feeling faint, nausea, feeling tired and old sweat. Sometimes the MI (myocardial infarction) occurs but has no history symptoms. The diseases which are associated with usually lower heart rate variability (HRV) are "Congestive heart failure", depression, "diabetic neuropathy", susceptibility to SIDS, "post-cardiac transplant" and reduced persistence in untimely infants [11].

## 2. Related Work

Number of authors has done their work in the field of ECG signal analysis. This section defines the work done till date and is defined in tabular form covering the objectives, methods with the outcome achieved.

**Table 2. Comparative Analysis of Existing Work**

| Author | Disease | Objective | Method | Result |
|---|---|---|---|---|
| Patrick E. McSharry | | Generate realistic synthetic ECG signals | Dynamical model which is based on 3 coupled ordinary differential equations | - Assess the methods of biomedical signal processing which are used to calculatethe figure from the ECG.<br>- Generate multi ECG signals |
| M.M. Al Rahhal | - | | Deep learning approach for classification of ECG signals | - More accuracy<br>- Fast online retraining |
| Sarang L. Joshi | - | "Equiripple notch filter"was the finest choice to eliminatethe PLI "power line interference" for removing motion artifact and EMG noise | - Equiripple notch filter<br>- discrete Meyer wavelet | - Not good for decomposition approach. |
| Philip de Chazal | - | Heartbeats classification | - The method assignedphysically notice heartbeat to 1 of 5 beat groups suggested by "ANSI/AAMI EC57:1998 standard".<br>- On the basis of different beats like ventricular ectopic beat (VEB); supraventricular ectopic beat (SVEB) and fusion of a normal and a VEBthe sensitivity, FPR and | - Sensitivity =75.9%, positive predictivity = 38.5%.<br>- For "SVEB class"; FPR = 4.7%<br>- For "VEB"; sensitivity = 77.7%, positive predictivity |

| | | | | =81.9%, and FPR = 1.2%. |
|---|---|---|---|---|
| Philip de Chazal | | Classification of heartbeats | Discussed the growth of a patient adjustingheartbeatclassifier mode | - Performance of a patient adjustable classifier upsurges with the quantity of training of system on local record.<br>- performance of system can be boosted with a small volume of adaptation even when all strokes used for adaptation are from same class. |
| Julian F. Thayer | Cardiovascular | The risk factors of CVD is notices through research | Risk factors in the CVD such as hypertension, obesity, family history ad work stress are reviewed | - the previous study shows that the risk factors have high impact on CVD |
| Ram bilas pichori | Diabetes | Differentiate between the normal R-R interval and the diabetic interval | Non-linear method based on the empirical mode decomposition | Features used $A_{asr}$, $A_{sodp}$ gave 95% tendency to differentiate between the two intervals. |
| M. sabarimalai manikandan | | R peal detection | R-peak detector based on automatic peak finding logic | - average detection accuracy = 99.80%<br>- sensitivity = 99.93% |
| M.G. | | Arrhythmic | "RR-interval" sliding | - arrhythmic |

| Tsipouras | | beat classification | window is taken in the classification techniqueof arrhythmic beat | beat classification; accuracy= 98% <br> - "arrhythmic episode detection" = 94% |
|-----------|--|---------------------|-------------------------------------------------------------------|--------------------------------|

From above table, it has been observed that there are some research gaps in the existing work and the current work, therefore, to fulfil this gap, we have proposed an analysis of ECG signal to see the presence of cardiovascular diseases in R-R interval using AI (Artificial intelligence) with GA (Genetic algorithm) as a optimization technique. Here, GA is used as an optimization algorithm for analysis of R-peak for different category of ECG signal using their peaks.

## 3. Proposed Framework

This research deals with exploring ECG signals for human heart to classify diseases. A novel method has been proposed for the classification on the basis of GA (Genetic Algorithm) for feature optimization and ANN (Artificial neural network) for the classification of disease. The performance of the system has been checked by utilizing Precision, Recall, F-measure and accuracy.

### 3.1. GA (Genetic Algorithm)

GA is generally used in the applications with more search space. The benefit of GA is that the process is fully automatic and evades the local minima [10]. GA consisted of number of components, namely, Crossover, Fitness function and mutation. The crossover operation is used for generating a new chromosome from parents set but the mutation operators add variation. The fitness function is used to execute a chromosome which is dependent on the survival chance. Population is the collection of chromosomes. New population is executed by using standard GA operations, like Single point Crossover, Selection operator and mutation. The procedure for the execution of genetic algorithm is defined below and shown in Figure 3:

Step 1: Initialization of arbitrary population with chromosomes.

Step 2: Calculation of fitness function in the population.

Step 3: Development of new population having individuals

Step 4: Selection of parent chromosomes for better fitness function

Step 5: Calculate crossover for having copy of parents

Step 6: Calculation of mutation for mutating the new offspring

Step 7: Placement of new offspring in the population

Step 8: Iterate the steps for the better solutions

Step 9: Stop
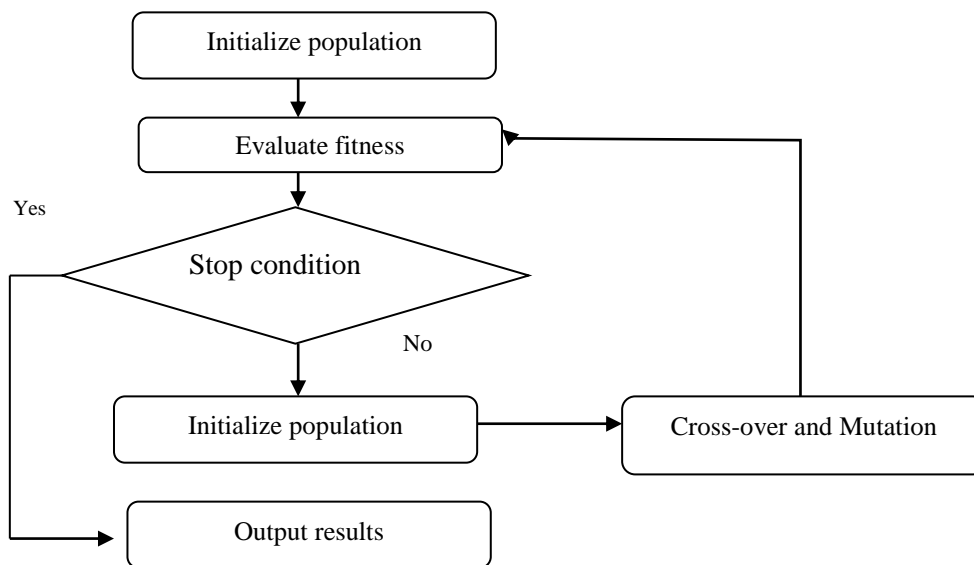
**Figure 3. GA Working Flowchart**

### 3.2. NN (Neural Network)

It is a mathematical model or the computational model defined on the basis of biological neural networks. The classifier of NN is dependent on neural network having interconnected neurons [11]. The neurons has positive as well as negative stimuli that is a numeral value taken from another neurons and when the weighted total of stimuli is more than the provided threshold value and it gets activated. The outcome of the neurons is generally a non-linear change of stimuli sum.

Neural networks are the neurons in the graphs. NN is modelled as a graph as a compilation of neurons that are integrated in acyclic graph. It can also be said that the outcome of few neurons can resulted in the input to the another neuron. NN is generally connected into different neuron layers. For usual NN, the frequent utilized layer type is fully connected layer having neurons among two distinct layers which are completely pair wise integrated, but the neurons having single layer with no connections.
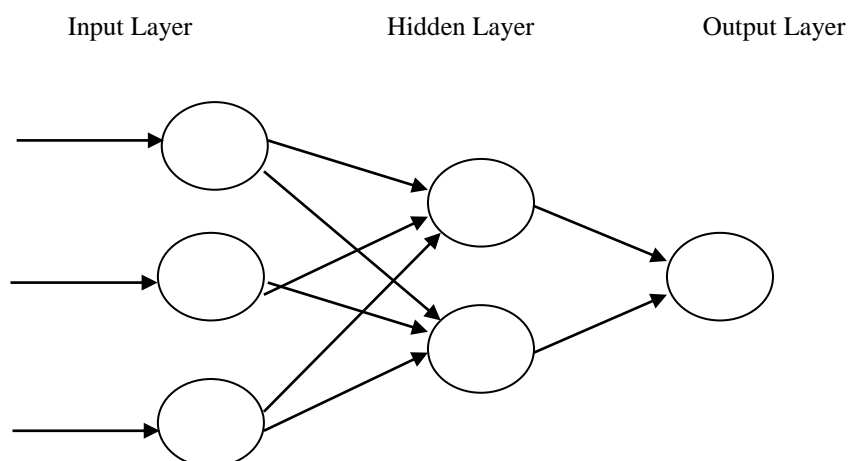


**Figure 4. Neural Network**

To implement the proposed work, below flow diagram has been shown in Figure 5. According to the flow diagram, simulation model has been designed that helps to find the

disease by R peak analysis. The implementation of proposed work is described in below section with experimental results analysis.
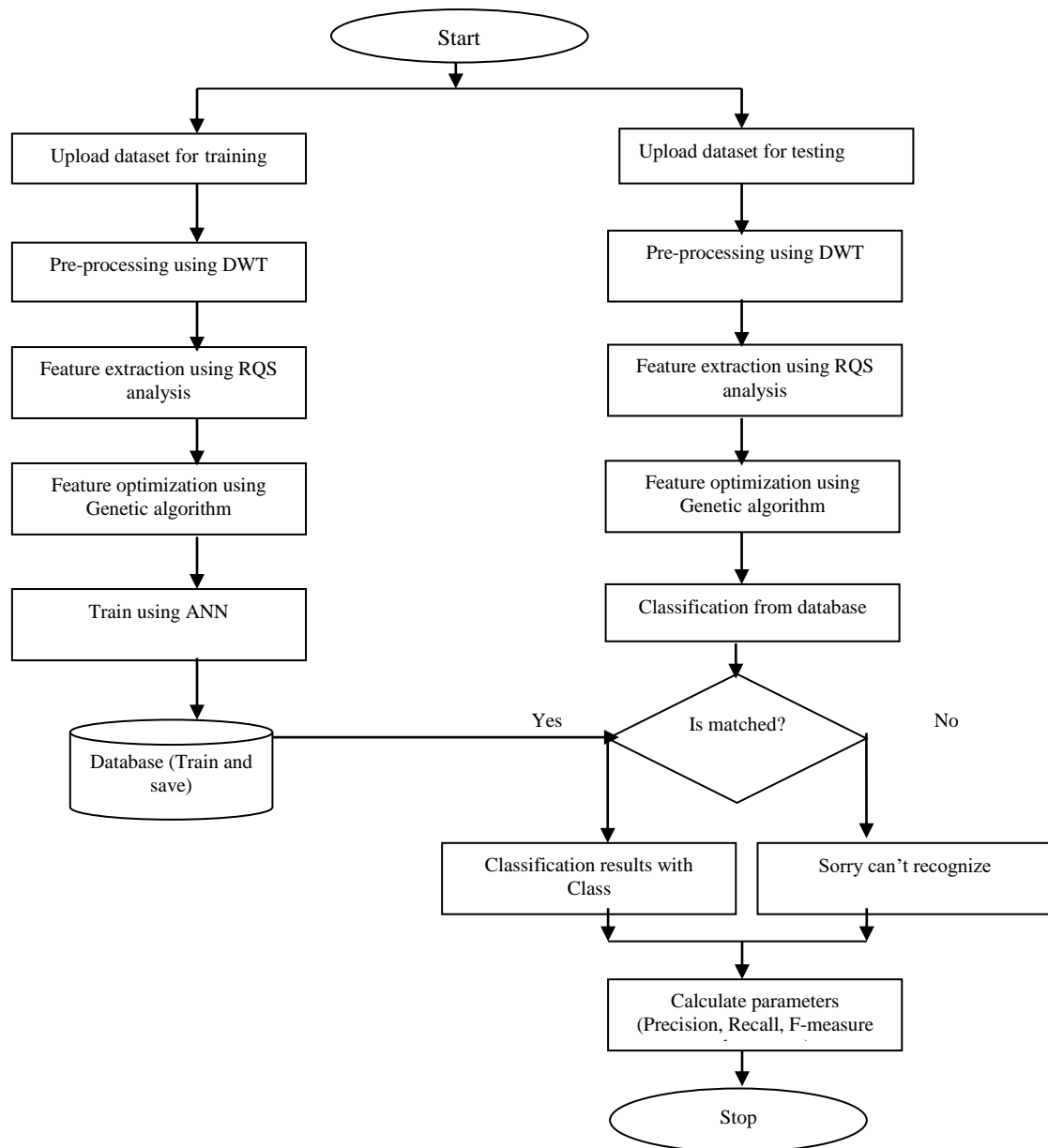


**Figure 5. Proposed Workflow**

## 4. Result and Analysis

This section defines the results obtained after the simulation of the proposed work. For the experimentation, parameters, viz. Precision, Recall, F-measure and accuracy has been calculated. The results have been shown in the form of tables and graphs.

**Table 2. Performance Parameters**

| S. No. | Precision | Recall | F-measure | Accuracy | Error | Time |
|--------|-----------|---------|-----------|----------|-------|-------|
| 1 | 0.9971 | 0.99150 | 0.99510 | 92.9555 | 7.05 | 12.45 |
| 2 | 0.9962 | 0.99090 | 0.98540 | 93.1885 | 6.82 | 13.93 |
| 3 | 0.9860 | 0.98920 | 0.99815 | 92.9880 | 7.02 | 12.65 |
| 4 | 0.9890 | 0.99171 | 0.99260 | 93.1880 | 6.82 | 15.63 |
| 5 | 0.9980 | 0.99170 | 0.99170 | 93.1840 | 6.84 | 13.56 |
| 6 | 0.9888 | 0.98920 | 0.99520 | 93.1960 | 6.81 | 11.87 |
| 7 | 0.9970 | 0.98560 | 0.99480 | 93.1880 | 6.82 | 13.53 |
| 8 | 0.9858 | 0.99170 | 0.99490 | 92.1890 | 7.82 | 14.67 |
| 9 | 0.9875 | 0.98230 | 0.99390 | 92.1920 | 7.81 | 19.56 |
| 10 | 0.9980 | 0.99165 | 0.99450 | 93.1953 | 6.81 | 11.69 |
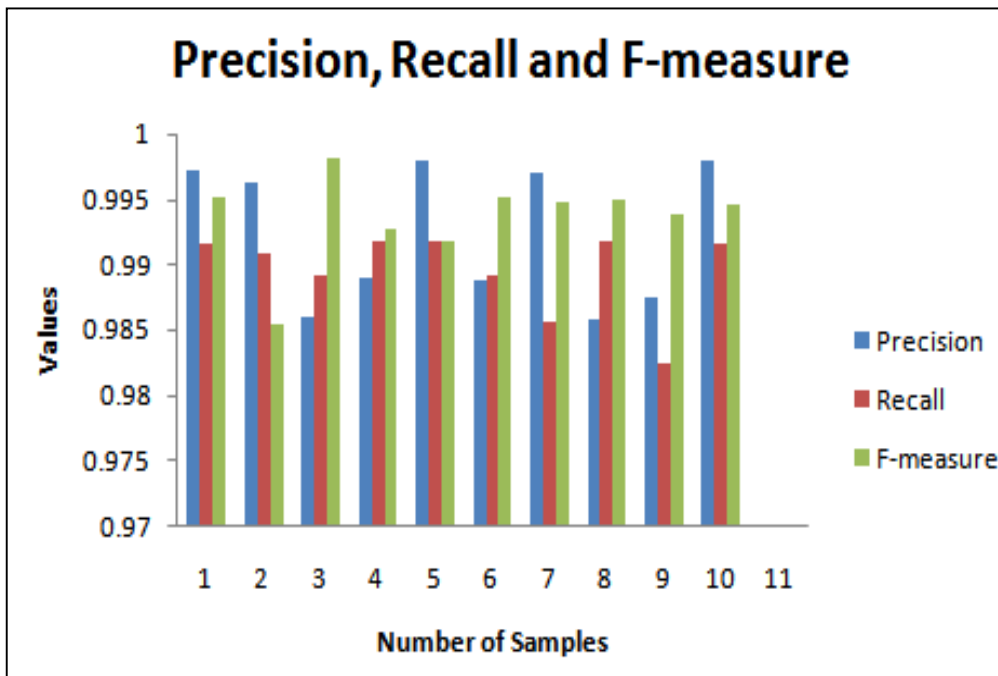


**Figure 6. Precision, Recall and d F-measure**

Above figure defines the results of precision, recall and F-measure. As depicted, X-axis defines the number of samples and Y-axis defines the obtained values. Blue bar is for precision, red bar is for recall and green bar defines the values of F-measure. The average precision rate is 0.9924 and it is good for the proposed work. The maximum precision rate is 1 and we have obtained 0.9924 during the classification of ECG signal using artificial neural network. The average recall rate is 0.9896 and it is also good for the proposed work. The recall rate is always less than the precision rate because it is the rate with respect to the all training data but the precision rate is the rate that can be defined with respect to the all testing data. F-measure is the harmonic mean of precision and recall and it is also better for the proposed work and the average value of F-measure is 0.9936.
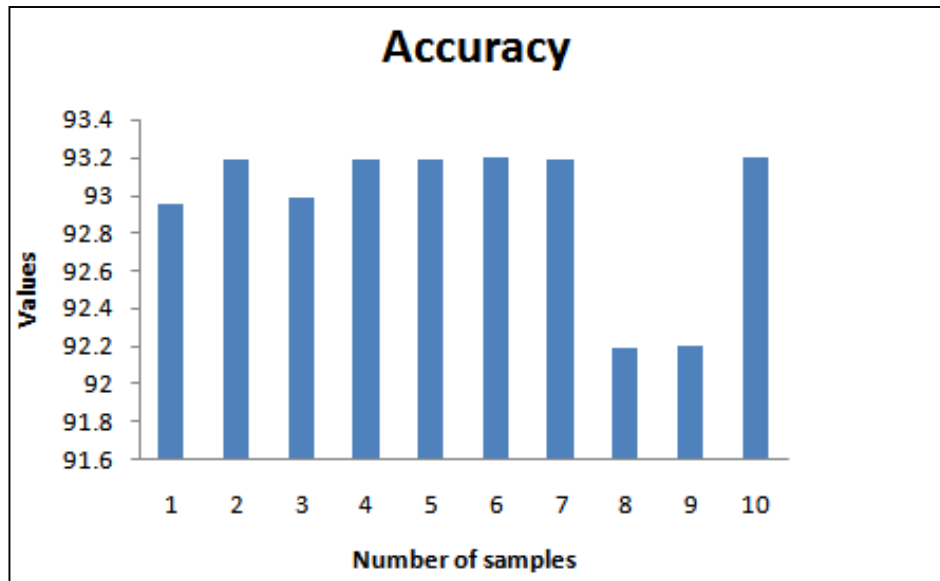
**Figure 7. Accuracy**

Above figure defines the result of accuracy parameter. As depicted, X-axis defines the number of samples and Y-axis defines the value obtained after the simulation of the proposed work. The average value of obtained accuracy is 92.94. Accuracy defines how accurately a system executes. It has been seen that the proposed work is 92.94 accurate.
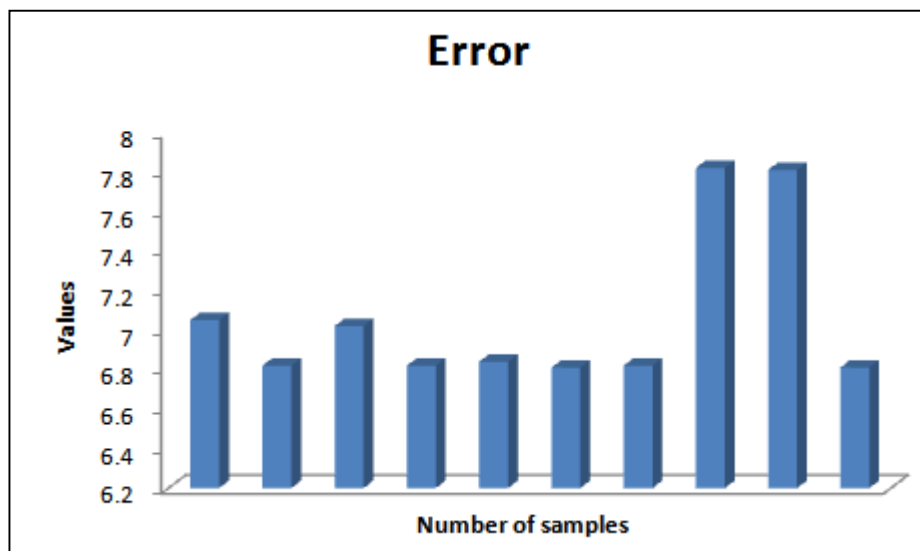


**Figure 8. Error**

Figure 8 defines the error obtained in the proposed work. In the above figure, X-axis defines the number of samples, Y-axis defines the obtained values. The presence of errors in the simulation work is 7.06 % approximately and it is better for proposed work.
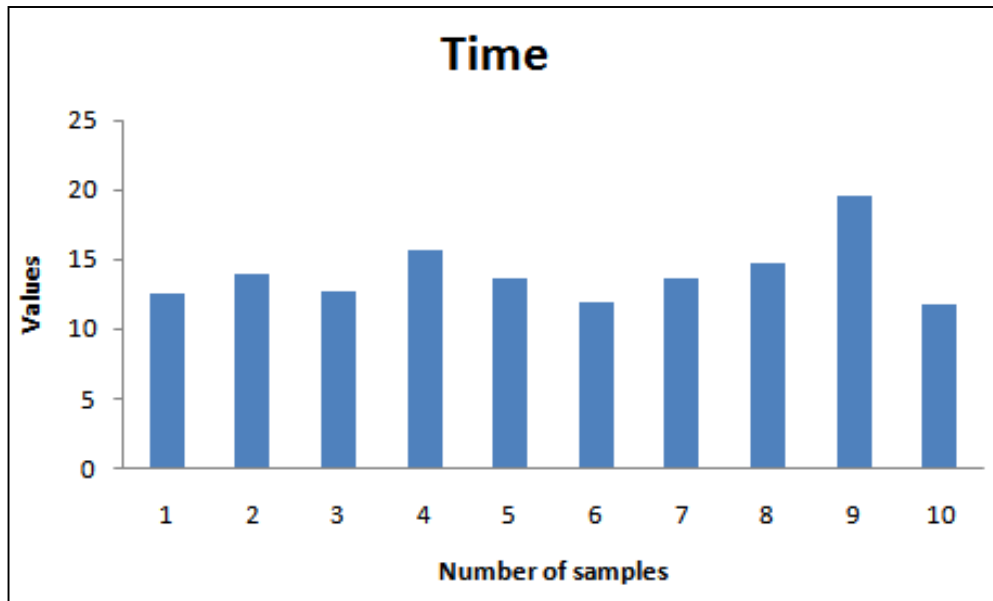
**Figure 9. Time**

Figure 8 defines the execution time of proposed work during the testing of ECG signal samples. In the above figure, X-axis defines the number of samples; Y-axis defines the obtained values for time. The average value obtained for time is 13.954sec to execute a sample of ECG signal.

## 5. Conclusion

Electrography is mainly used for detection of HPV in the heart rate or the beats of the QRS complex. The variability in the R-R intervals are checked and are compared with the previously taken heart signals. Accurate diagnosis is done to predict the cardiovascular disease. For this, firstly pure R-R intervals should be extracted to find the relative disease. Optimization algorithms like genetic and ant bee colony is put on the ECG signal to get the pure R-R intervals. In this research, ECG analysis has been performed in cardiovascular diseases in R-R interval using AI (Artificial intelligence) with GA (Genetic algorithm) as an optimization technique. To check the performance of the work, parameters, such as, Precision, recall, f-measure, accuracy, time and error has been measured. Average precision rate is 0.9924 and it is good for the proposed work. The maximum precision rate is 1 and we have obtained 0.9924 during the classification. The average recall rate is 0.9896 which is also fine for the proposed work. The obtained F-measure is 0.9936. It has been seen that the proposed work is 92.94 accurate. The presence of errors in the simulation work is 7.06 approximately and the average value obtained for time is 13.954.

Future scope lies in the use of former classifiers like SVM with the aim of having multidimensional data and making use of feature reduction algorithms, so that accuracy rate can be enhanced. SVMs bring a unique solution, since the optimality problem is rounded. This is an advantage to artificial neural network (ANN) which has several solutions related with local minima and for this reason may not be tough over different samples. For optimization algorithms like artificial bee colony (ABC) and Particle swarm optimization (PSO) would be used.

## References

[1]    M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals", Information Sciences, vol. 345, **(2016)**, pp. 340-354.

[2]    P. E. McSharry, G. D. Clifford, L. Tarassenko and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals", IEEE transactions on biomedical engineering, vol. 50, no. 3, **(2003)**, pp. 289-294.

[3]    S. L. Joshi, R. A. Vatti and R. V. Tornekar, "A Survey on ECG Signal De-noisingTechniques", International Conference on Communication Systems and Network Technologies (CSNT), **(2013)**, pp. 60-64.

[4]    C. Saritha, V. Sukanya, and Y. Narasimha Murthy, "ECG Signal Analysis Using Wavelet Transforms", Bulgarian Journal of Physics, vol. 35, **(2008)**, pp. 68-77.

[5]    P. De Chazal, M. O'Dwyer and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features", IEEE Transactions on Biomedical Engineering, vol. 51, no. 7, **(2004)**, pp. 1196-1206.

[6]    P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features", IEEE Transactions on Biomedical Engineering, vol. 53, no. 12, **(2006)**, pp. 2535-2543.

[7]    J. F. Thayer, S. S. Yamamoto and J. F. Brosschot, "The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors", International Journal of cardiology, vol. 141, no. 2, **(2010)**, pp. 122-131.

[8]    Y. C. Yeh and W. J. Wang, "QRS complexes detection for ECG signal: The Difference Operation Method", Computer methods and programs in biomedicine, vol. 91, no. 3, **(2008)**, pp. 245-254.

[9]    R. B. Pachori, P. Avinash, K. Shashank, R. Sharma and U. R. Acharya, "Application of empirical mode decomposition for analysis of normal and diabetic RR-interval signals", Expert Systems with Applications, vol. 42, no. 9, **(2015)**, pp. 4567-4581.

[10]  M. S. Manikandan and K. P. Soman, "A novel method for detecting R-peaks in electrocardiogram (ECG) signal", Biomedical Signal Processing and Control, vol. 7, no. 2, **(2012)**, pp. 118-128.

[11]  M. G. Tsipouras, D. I. Fotiadis and D. Sideris, "An arrhythmia classification system based on the RR-interval signal", Artificial intelligence in medicine, vol. 33, no. 3, **(2005)**, pp. 237-250.

[12]  C. Tu, Y. Zeng and X. Yang, "A new approach to detect QRS complexes based on a histogram and genetic algorithm", Journal of medical engineering & technology, vol. 29, no. 4, **(2005)**, pp. 176-180.

[13]  S. S. Omran, S. M. R. Taha and N. A. Awadh, "ECG Rhythm Analysis by Using Neuro-Genetic Algorithms".

[14]  S. Jain, V. Bajaj and A. Kumar, "Efficient algorithm for classification of electrocardiogram beats based on artificial bee colony-based least-squares support vector machines classifier", Electronics Letters, vol. 52, no. 14, **(2016)**, pp. 1198-1200.