IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

ECG Signal Classification using Support Vector Machine and Linear Discriminant Analysis

S. Grover[1*], Shailja[2]

[1,2]Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Punjab, India

[*]*Corresponding Author: shifagrover@yahoo.com, Tel.: 9056191919*

DOI: https://doi.org/10.26438/ijcse/v7i5.17201725 | Available online at: www.ijcseonline.org

Accepted: 20/May/2019, Published: 31/May/2019

***Abstract*—**With the increase in the number of the patients of heart diseases, it is important to analyse the heart activity so that we can easily classify and diagnose the disease. Since, the Electrocardiogram (ECG) signals are used for detecting the cardiac diseases so, in this study analyses and classification of ECG signal are done using Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) data mining techniques. A cleaned ECG signal provides vital information about the heart diseases and ischemic changes that may occur. It provides necessary information about the functional characteristics of the heart. In this paper, R-peaks of the ECG signal is analysed and its optimization is done using the Genetic Algorithm (GA) as the optimization algorithm. The optimized features are selected using this algorithm. Classification of heart disease is done using SVM and LDA data mining techniques. The two cardiac disorder named bradycardia and tachycardia is classified using SVM and LDA techniques. The comparison of these two techniques is performed on the basis of precision value. In this study, SVM showed better results.

***Keywords*—**ECG Signal, Genetic Algorithm, SVM, LDA

I. INTRODUCTION

Data mining is defined as a process of extracting useful information from existing huge amount of data. In other words we can say that data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis [1]. Thus data mining cab be used to analyse medical data. In this work, data mining techniques are used to analyse the ECG signals. ECG is known as a method to measure and record different potentials of the heart. It represents the relaxation and contraction activities of the heart. It is divided into the phases of depolarization and repolarization of muscle fibres making up the heart [2]. In this study R- peak of ECG signal is used for detecting and classification of heart diseases. The two cardiac disorder named bradycardia and tachycardia is classified through the analysis of R-peak of the ECG signal.

The main objective of this work is to optimize the ECG signal using genetic algorithm as an optimization algorithm and then to categorize these optimized disease dataset using SVM and LDA algorithm and to train the system on the source of the features extracted and to test the ECG signal. The comparison of these two techniques is performed on the basis of precision value.

Rest of the paper is organized as follows, Section II contains the introduction to ECG Signal, Section III contains introduction to Genetic Algorithm, Section IV contains introduction to Linear Discriminant Analysis, Section V contains introduction to Support Vector Machine, Section VI contains the related work for classification of ECG Signal, section VII explains the methodology of proposed work with flow chart, Section VIII describes results and discussions of proposed work, and Section IX concludes the entire work.

II. ECG SIGNAL

ECG also known as an electrocardiogram, which are biomedical signals representing cardiac electrical activity. ECG signal consists of different waves such as P wave, QRS complex and T wave as shown in figure 1. These waves having different frequencies and amplitudes are also separated from each other without repeating a stationary marking itself. These waves; P wave, the QRS complex and the T wave. P-wave that occurs during the compression of the blood to the right ventricle from the atrium during atrial depolarization occurs. Q, R and S wave of the QRS complex can be separated into sections, when the right ventricle, the aorta, the right ventricular blood jam occurs when depolarization. QRS complex has a period of between 80 ms and 120 ms. However, each of the Q wave of the QRS complex, It does not include an R-wave and an S wave; It has many formal form of a QRS complex. T-wave, the

ventricular heart rate is generated in preparation for the next time period repolarized [2].

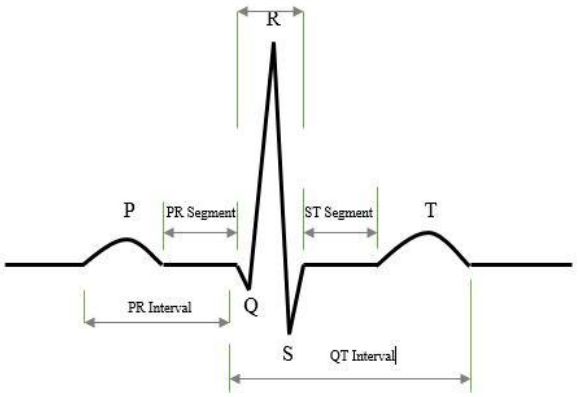

Figure 1. ECG signal

It offers useful information about the heart's rhythm and work. The analysis of the ECG signal helps in determining the different types of heart disease and to diagnose them [3]. Thus in this work, R-peak of the signal has been used for the classification two different kind of heart abnormality.

III. GENETIC ALGORITHM

Genetic algorithm (GA) is one of the optimization algorithm. It was introduced by John Holland in the early seventies. It is considered as a powerful stochastic algorithm, which is inspired by the some of the natural evolution processes. This algorithm is based on the mechanism of natural selection and natural genetics, which has been quite successfully, applied in machine learning and optimization problems. The main mechanisms of Genetic Algorithm are named as the crossover, mutation, and a selection function. The GA differing from standard search techniques, begins with the initial set of random solution called population. Each individual with in population referred to as chromosomes. The chromosomes evolve through number of ordered iterations, known as generations. During every generation, the chromosomes are evaluated using the fitness function. The next generation new chromosomes, known as offspring are formed either by fusion of two chromosomes from the current generation using a crossover operation or by modifying the chromosomes through the mutation operation. A new generation is generated by selecting according to the fitness value and rejecting those which does not satisfy the fitness function condition. Thus, in this way the population size remains constant. After number of generation the algorithm covers the best chromosomes that may represents the optimal solution to the problem [4]. The flow chart for genetic algorithm is shown in figure 2.

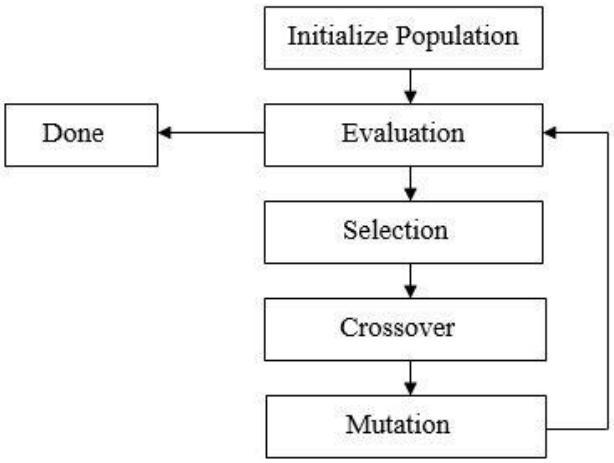

Figure 2. Flow Chart of Genetic Algorithm

A. Steps of Genetic Algorithm

Step1: To reset random population having chromosomes.
Step2: To compute suitability function in the population.
Step3: To develop the novel population of persons.
Step4: To choice parent chromosomes for greatest fitness function.
Step5: To do crossover to have the duplicate of parents.
Step6: To do the alteration to mutate novel off springs.
Step7: To place novel offspring in the people.
Step8: To repeat the steps to get a fulfilled solution.

IV. LINEAR DISCRIMINANT ANALYSIS

Linear Discriminant Analysis (LDA) technique was originally established in 1936 by R. A. Fisher. The term Discriminant Analysis is usually well-defined as a statistical approach which is employed to classify the entities into mutually exclusive and exhaustive groups on the basis of a set of measurable features of the entities. It is established upon the conception of searching for a linear combination of variables (predictors) that proficiently separates two classes (targets). It is mainly a technique that can be used in statistics, pattern recognition and machine learning in order to find a linear combination of features that characterizes or separates two or more classes of objects or events. It is mainly a technique that can be used in statistics, pattern recognition and machine learning in order to find a linear combination of features that characterizes or separates two or more classes of objects or events. LDA easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. It maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal reparability. It is mostly used in the machine learning problems like pattern recognition, face

recognition, feature extraction and data dimensionality reduction [5].

V. SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a supervised learning technique which passes a linearly differ hyper plane via dataset for the classification of data in two groups. Hyper plane is a linear differentiator for some dimension. It can be 2D line and 3D plane. It is a decision boundary that separates the data into two different types of classes. The hyperplane is that which enhances the margin. Margin is known as a distance among the hyperplane with some of the close points. The close points are known as support vectors as those points control the hyperplane [5]. The SVM technique consist of two module i.e. training and testing module. In training module, the designed model is trained using the class labels and feature values. Testing of the designed module is done using this trained model. Margin between the two classes decides the hyper plane. The optimal hyper plane is considered that line which separates two classes in well manner and has maximum vector margin [6].

VI. RELATED WORK

R. Varatharajan *et al.* [5] introduced a SVM model with a weighted kernel function method to classify more features from the input ECG signal. In this heart diseases are classified into Left Bundle Branch Block (LBBB)**,** Right Bundle Branch Block (RBBB), Premature Ventricular Contraction (PVC) and Premature Atrial Contractions (PACs). The performance of the proposed LDA with enhanced kernel based (SVM) method is comparatively analysed with other machine approaches such as Linear Discriminant Analysis (LDA) with multilayer learning perceptron (MLP), Linear Discriminant Analysis (LDA) with Support Vector Machine (SVM), and Principal Component Analysis (PCA) with Support Vector Machine (SVM).

Rahime Ceylan [7] presented two well-known dictionary learning algorithms to extract features of ECG signals. The features of ECG signals have been extracted by using Method of Optimal Direction (MOD) and K-Singular Value Decomposition (K-SVD) and the extracted features have been classified by Artificial Neural Network (ANN). Twelve different ECG signal classes which taken from MIT-BIH ECG Arrhythmia Database have been used. From the experiment it is seen that performance of classifier increases in usage of K-SVD for feature extraction. The highest classification accuracy upto 98.74 % has been obtained.

C. Venkatesan *et al.* [8] introduced KNN based algorithm to classify normal and abnormal ECG signal. LMS based adaptive filters are used in ECG signal pre-processing, but this consume more time for processing due to long critical path. To overcome this problem, a novel adaptive filter with delayed error normalized LMS algorithm has been utilized to attain high speed and low latency design. Low power design has been achieved in this design by applying pipelining concept in the error feedback path. Arrhythmic beat classification has been carried out by KNN classifier on HRV feature extracted signal. Classification performance reveals that the proposed DWT with KNN classifier provides the accuracy of 97.5% which is better than other machine leaning techniques.

Sandeep Raj *et al.* [9] describes that signal processing techniques are an obvious choice for real-time analysis of electrocardiography (ECG) signals. This paper presents a new approach, i.e., discrete orthogonal stockwell transform using discrete cosine transform for efficient representation of the ECG signal in time–frequency space. In addition, the dynamic features (i.e., RR-interval information) are computed and concatenated to the morphological features to constitute the final feature set, which is utilized to classify the ECG signals using support vector machine (SVM). In order to improve the classification performance, particle swarm optimization technique is employed for gradually tuning the learning parameters of the SVM classifier.

Pengfei Guo *et al.* [4] focused on three different kinds of the novel enhanced genetic algorithm procedures including the hybrid genetic algorithm, interval genetic algorithm and hybrid interval genetic algorithm are respectively presented. As the results of the proven systems show, the hybrid genetic algorithm can determines the better optimum design than the traditional optimization algorithms and genetic algorithm. The interval genetic algorithm and hybrid interval genetic algorithm can avoid calculating system slope in traditional interval analysis and determines the optimum interval range of the parameters under allowable corresponding objective error boundary. It is the first time that genetic algorithm has been applied to interval optimization process.

Tanoy Debnath *et al.* [10] proposed an efficient method of analysing ECG signal and predicting heart abnormalities. In the proposed scheme, at first the QRS components have been extracted from the noisy ECG signal by rejecting the background noise. This is done by using the Pan Tompkins algorithm. The second task involves calculation of heart rate and detection of tachycardia, bradycardia, asystole and second degree AV block from detected QRS peaks using MATLAB. The results show that from detected QRS peaks, arrhythmias which are based on increase or decrease in the number of QRS peak, absence of QRS peak can be diagnosed. The final task is to classify the heart abnormalities according to previous extracted features. The Back Propagation (BP) trained feed-forward neural network

has been selected for this research. Here, data used for the analysis of ECG signal are from MIT database.

Sayanti Chattopadhyay *et al.* [11] described that ECG peak detection is important for different disease purposes or cardiac dysfunction. For different diseases, this article gives a table of special ECG signals. This paper described simulation method in which ECG data analysis and peak detection were completed.

VII. METHODOLOGY

Getting an ECG signal can be a simple job, other than to obtain a dependable ECG signal towards only if a clinic analysis by a specialist is a more difficult job, that's why manipulating and filtering a signal is an intricate task. ECG signal is a graphical illustration of cardiac activity and it used to compute number of cardiac diseases and abnormalities there in heart. ECG signals have P wave, QRS complex with T wave with any divergence in parametric shows abnormalities in heart. In this work main problem of optimization of features set. If we cannot find the exact feature of Bradycardia and Tachycardia attack then we cannot train our module properly. If training is not well according to all the diseases information then at the point of classification we cannot classify accurate categories of test data. The problem of this research work is to classify the Heart disease dataset using optimization algorithm i.e. GA algorithm and train the system on the basis of the features extracted and as well to check the image on the basis of the features at the database and the features extracted of the signal to be tested.

In this work, classification system has been designed for ECG signal on the basis optimized R-peaks of signal using genetic algorithm as an optimization algorithm with SVM and LDA as a classifier. It has been designed using MATLAB simulator. For this work, MIT-BIH Arrhythmia database has been used. Firstly, a GUI is created in MATLAB that comprises of two panels namely testing panel and training panel. In training panel, the designed model is trained using SVM and LDA technique. In testing panel, ECG signal has been tested for two different ECG disease classes. The features of the signal have been extracted based on the threshold signal as per the R-peak of the test signal. Then the features of the extracted R-peaks have been optimized using genetic algorithm. After that the disease is classified using SVM and LDA classifier and efficiency has been determined on the basis of performance parameter such as precision value. The flow chart for the proposed work is shown in figure 3.

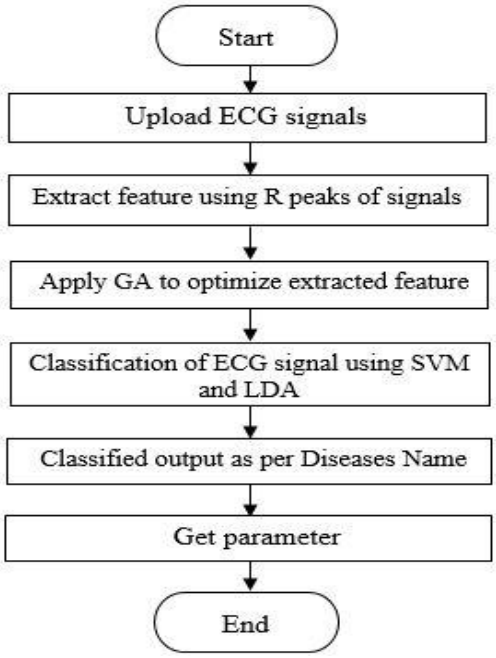

Figure 3. Flowchart of Proposed Work

VIII. RESULTS AND DISCUSSION

Figure 4 shows the main window of the proposed ECG Disease Detection System. The system is mainly categorized into three steps: Feature Extraction & optimization, training and testing. In below figure there are two sections (i) Bradycardia (ii) Tachycardia. Every section has three sub-sections namely; upload sample, R-peak and genetic. Upload sample is used to select the folder and upload the information of ECG signal. R-peak is use for feature extraction. The extracted feature is improved by using optimization technique known as GA optimization.

Figure 4. Main GUI

Figure 5 shows the R-peak analysis of uploaded signal. In this part, the R-peaks of the uploaded signal are obtained

using the threshold value. Threshold value is considered as an average value. After the R-peak analysis the signal is ready for the optimization.

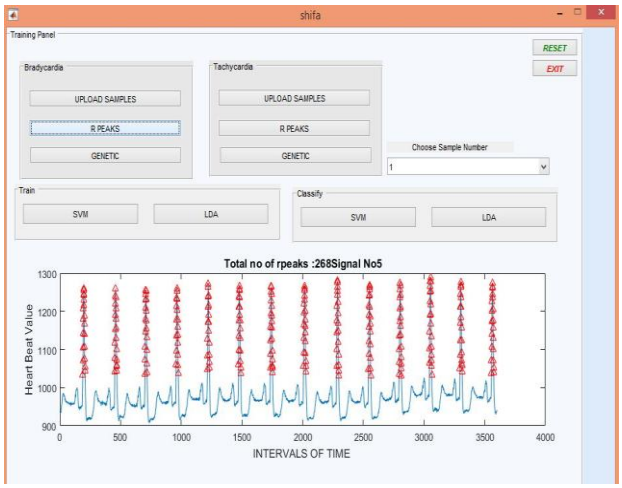

Figure 5. R-Peak Analysis

Figure 6 shows the optimized signal. After applying genetic algorithm the optimized R-peaks are obtained. Genetic algorithm helps to highlight the optimized R peak of the ECG signal. This increases the quality of the ECG signal. On the basis of these optimized R peak SVM and LDA algorithm train the system. With the optimization the training is done in better way and optimized results are obtained.

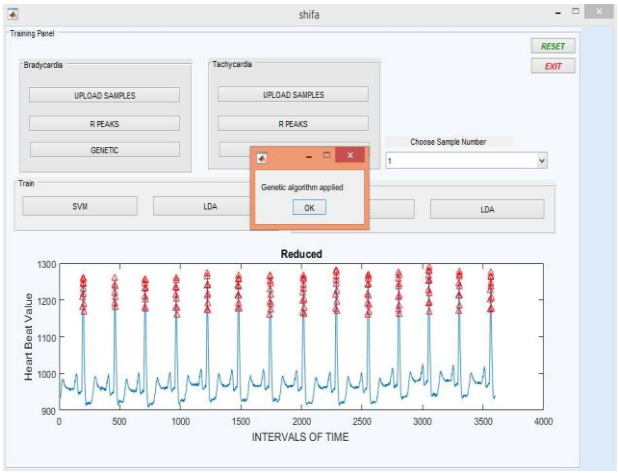

Figure 6. Optimized Signal

Figure 7 represents the SVM plot for the bradycardia and tachycardia signal classification. The 1 in the figure denotes the bradycardia signal and the 2 denotes the tachycardia signal. The kernel or decision boundary used in SVM for signal classification is a polynomial kernel. The training of the work using SVM classification algorithm is shown in this figure in the graphical form.

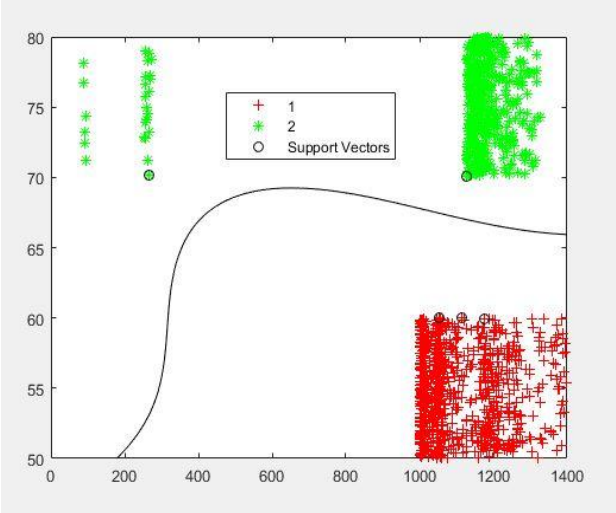

Figure 7. Classification Using SVM

Figure 8 represents the precision value measured after applying LDA and SVM in the proposed work. Blue and orange colour signifies the value of precision for SVM and LDA. From the figure it is clear that SVM perform well as compared to LDA classifier.

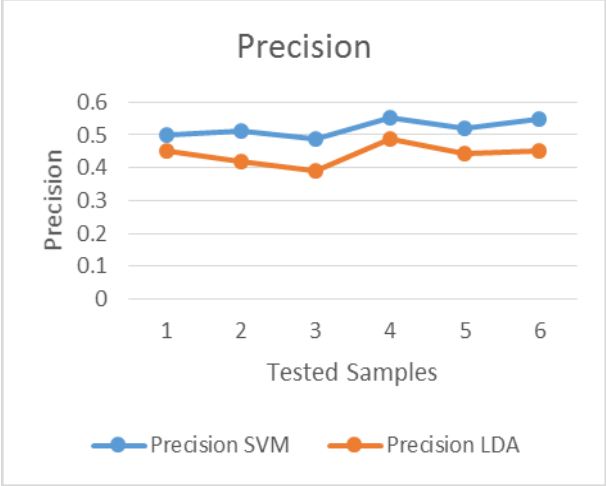

Figure 8. Precision Value

The formula for Precision can be written as below:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Table 1. Comparison of SVM and LDA

Tested Sample Count	Precision SVM	Precision SVM
1	0.5	0.45
2	0.51	0.42
3	0.489	0.39
4	0.551	0.489
5	0.521	0.445
6	0.55	0.4521

Table 1 shows the comparison of precision value for both the techniques for some of the samples. The average value of precision measured for the SVM and LDA technique are 0.52 and 0.44 respectively. It has been concluded that there is an increment of 18.18% when SVM is used.

IX. CONCLUSION

The objective of this research work is to categorize the disease dataset using SVM and LDA algorithm and to train the system on the source of the features extracted and moreover to test the ECG signal. This research is based on studying the implemented approaches in the ECG diseases and then to propose a novel technique or algorithm for classification of two cardiac disorders named as Bradycardia, and Tachycardia dependent on SVM and LDA. Heart Disease is the disease that mainly affects the heart. Most of the losses are due to heart diseases. This project presents the ECG Disease Detection System based on GA, and SVM and LDA, in which detection is based on precision value. The precision of the proposed work with SVM is 18.18% more than that of LDA technique.

REFERENCES

[1] V.K. Gujare , P. Malviya, "*Big Data Clustering Using Data Mining Technique*", International Journal of Scientific Research in Computer Science and Engineering, Vol. **5**, Issue.**2**. pp.**9-13**, **2017.**

[2] A. Diker, E. Aci, Z. Comert, D. Avci, E. Kacar, I. Serhatlioglu, "Classification of ECG Signal by using Machine Learning Methods", In the Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, **Turkey**, pp**. 1-4**, **2018.**

[3] A. Gaikwad1 , M.S. Panse, "*Extraction of FECG from Non-Invasive AECG signal for Fetal Heart Rate Calculation*", Int. J. Sc. Res. in Network Security and Communication, , Vol. **5**, Issue.**3**. pp.**109-112**, **2017.**

[4] P. Guo, X. Wang, Y. Han, "The Enhanced Genetic Algorithms for the Optimization Design", In the Proceedings of the 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, **China**, pp. **2990-2994**, **2010**.

[5] R. Varatharajan, G. Manogaran, M.K. Priyan, "*A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing*", Multimedia Tools and Applications, Vol.**77**, Issue.**8**, pp.**10195-10215**, **2018**.

[6] K. Kaur, L.S. Mann.,"*Heart Disease Prediction System Using ANOVA, PCA and SVM Classification*", International Journal of Advanced Research, Ideas and Innovations in Technology, Vol. **2**, Issue.**3**, pp**.1-6, 2016.**

[7] R. Ceylan, "*The Effect of Feature Extraction Based on Dictionary Learning on ECG Signal Classification*", International Journal of Intelligent Systems and Applications in Engineering, Vol.**6**, Issue.**1**, pp.**40-46**, **2018.**

[8] C. Venkatesan, P. Karthigaikumar, R. Varatharajan, "*A novel LMS algorithm for ECG signal preprocessing and KNN classifier based abnormality detection*", Multimedia Tools and Applications, Vol.**77**, Issue.**8**, pp.**10365-10374**, **2018**.

[9] S. Raj, K.C. Ray, "*ECG signal analysis using DCT-based DOST and PSO optimized SVM*", IEEE Transactions on Instrumentation and Measurement, Vol.**66**, Issue.**3** pp.**470-478**, **2017**.

[10] T. Debnath, M. Hasan, T. Biswas, "Analysis of ECG signal and classification of heart abnormalities using Artificial Neural Network", In the Proceedings of the International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh, pp.**353-356**, **2018**.

[11] S. Chattopadhyay, S. Das, A. Nag , J.K. Ray, S. Bhattacharjee, B. Neogi, "*Design and Simulation Approach Introduced to ECG Peak Detection with Study on Different Cardiovascular Diseases*", International Journal of Scientific and Research Publications, Vol.**2**, Issue.**12**, pp.**348-353**, **2012**.

Authors Profile

Ms. S Grover studied Bachelor of Technology in Information Technology from Guru Nanak Dev engineering College Ludhiana (Punjab) in 2015. She is currently pursuing Master of Technology in Computer Science & engineering from Guru Nanak Dev Engineering College, Ludhiana (Punjab). Her main research focuses on development of disease detection system for ECG signal using genetic algorithm, SVM, LDA.

Ms. Shailja received a Bachelor degree in Computer Science Engineering in 2005 and a Master degree in Computer Science & Engineering in 2011. She is with Guru Nanak Dev Engineering College, Ludhiana (Punjab) India as Assistant Professor, in CSE department since 2013- 2019. Her research area is image processing.

