International Journal of Research in Advent Technology, Vol.6, No.11, November 2018
E-ISSN: 2321-9637
Available online at www.ijrat.org

Classification of ECG Signals using Self Advising Support Vector Machine and Fuzzy C Means Clustering

D. Raj Balaji[1], A.Marimuthu[2]
Ph.D Research Scholar[1], Associate Professor and Head[2]
Manonmaniam Sundaranar University[1], Govt. Arts College, Coimbatore[2]
Email: d.rajbalaji@gmail.com[1], mmuthu2005@gmail.com[2]

Abstract- An electrocardiogram (ECG) is a bioelectrical signal which takes and records the heart's pumping activity as an electrical pulse based on the time. It is an imperative symptomatic tool used to analyze the heart functionality and assessing. The early discovery of abnormalities in the heart is essential for the human beings. ECG signal can be characterized as normal and abnormal based on the heart rate conditions. If the electrical action of the heart is unbalanced and it can make heartbeat be very slow or quick, such stage is called as abnormal. It can occur in the heart beat and be of insignificant result, yet they may lead to show a significant issue that prompts stroke or sudden cardiovascular failure. Such abnormal condition of the heart is found using the electrical activity in the heart, which contains information about the condition of the heart. The Nature of disease affecting the heart is calculated using the P-QRS-T wave. Such ECG signal wave gives the values based on their shape, size and their time intervals between the peak values of the wave signal, which is used to analyze the condition of the heart. In this paper, an automatic classification of ECG is proposed using the combination of clustering and classification techniques. The system was implemented using MATLAB tool, which detect the abnormalities in the ECG as well as it compares the accuracy of the classification algorithms used in this paper. The Dataset used in this paper was downloaded from MIT BIH Arrhythmia Database (https://physionet.org/physiobank/database/mitdb). Based on the result and discussion mentioned in this paper, SA-SVM along with FCM yields good result.

Keywords-Electrocardiogram, SA-SVM, FCM, PSO, Cardiovascular Disease.

1. INTRODUCTION

The analysis of the ECG has been widely utilized for diagnosing numerous heart related diseases. An Electrocardiogram (ECG) is a bio-electrical signal which is utilized to record the heart's electrical movement regarding time. The Heart signals are taken from ECG, or, in other words Electrocardiography and these signs are picked by utilizing anodes in arms, leg, chest of our body. ECG signals are utilized as the parameter for discovery of Cardiac vascular disease. Contingent upon the state of the ECG waveform, one can discover the cardiovascular wellbeing. Early and precise discovery is essential in identifying heart related disease and picking fitting treatment for a patient.

Electrocardiogram (ECG) signals demonstrate electrical action of the heart. The signs are utilized for early conclusion of heart variations from the norm and record changes in the heart [1,2]. Heart arrhythmias happen with deviation from the ordinary physiological conduct of the heart. They are typically connected with anomalous pumping capacity and are caused by any intrusion in the normality, rate, or conduction of the electrical inclination of heart. Arrhythmias diminish the life time for the patient and can even induce sudden passing [1,3]. The programmed division of ECG pulses into sub-classes can be utilized in equipment supported by computer aided design. It diminishes the time spent via cardiologists on the examination of these records. A powerful ECG heart beat grouping by incorporating five imperative modules such modules are Pre-processing, Feature extraction, Feature Selection, Clustering and Classification. Feature extraction and reduction are two essential strides that regularly affect influence the classification execution of any heartbeat classification framework. Thus, the fundamental responsibilities of arrhythmia arrangement issues are the extraction of sufficient highlights and the decrease of their sizes for classifiers with the end goal to achieve ideal grouping results.

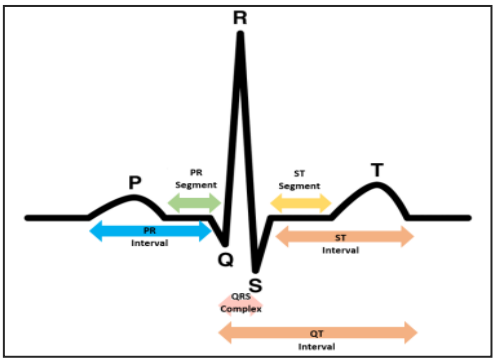

Figure 1: P-QRS-T Wave segments and intervals present in a ECG Signal.

The ECG is a reasonable record of the direction and magnitude of the electrical upheaval that is created by depolarization and repolarization of the atria and ventricles. One cardiovascular cycle in an

ECG flag comprises of the P-QRS-T waves. Figure 1 demonstrates an example of ECG signal. The P-QRS-T wave shape, measure and their time intervals between its different pinnacles contain helpful data about the environment of disease influencing the heart.

The organization of the paper is mentioned as given below. In section 2, the literature review about the ECG Classification, feature extraction and its related study done by the previous researcher was discussed. In section 3, the proposed research methodology and phase divisions were discussed. In section 4 the experimental results and the framework designed for the execution was discussed. In section 5 the conclusion of the study that made in this paper was discussed

2. LITERATURE REVIEW

In [1] Sheng proposed a hybrid Methodology of Fuzzy along with Support Vector Machine, which is based on the Fuzzy C Means Clustering techniques for each class that present in the training set. The fuzzy parameter q, acted as a support vectors, which becomes as cluster centers in the clustering process.

In [2] the ECG Classification is based on the Neural Network and fuzzy Logic classification techniques. The final result is calculated based on the Heart Rate Variability, simply denoted as HRV.

Some specific features extracted from this Heart Rate Variability and given as a input for the classification techniques for the prediction of the abnormalities present in the ECG Signal. Melgani et al. concentrated on a SVM-based way to deal with computer aided system for finding of ECG beats in their investigation. More over the (PSO) Particle Swarm Optimization technique was utilized and the capacity of the SVM calculation to sum up was progressed [3]. Güler et al. utilized a joined Neural Network methodology to group ECG beats. The ECG signal was disintegrated utilizing Wavelet Transform for the time-recurrence portrayal in the examination. They computed the measurable values speaking to the dissemination and grouped four beat types [4].

3. RESEARCH METHODOLOGY

In this paper, the ECG Classification is based on the SA-SVM and KNN classification techniques. The final result is calculated based on the Heart Rate Variability, simply denoted as HRV. Some specific features extracted from this Heart Rate Variability and given as a input for the classification techniques for the prediction of the abnormalities present in the ECG Signal. The final output may be normal or abnormal based on the results of the classification techniques.

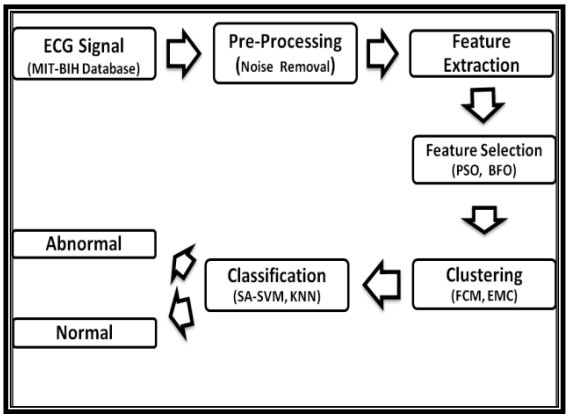

Figure 2: Overall Architecture of Proposed ECG Classification System.

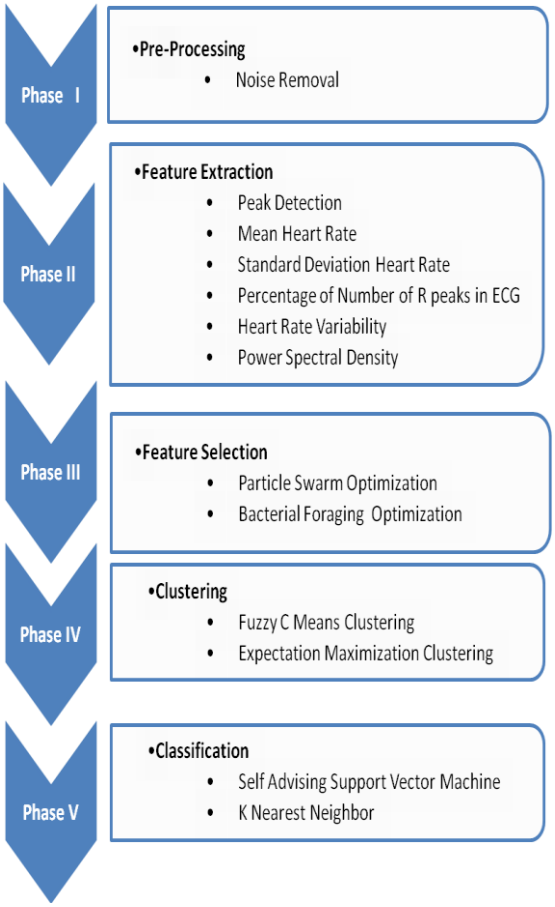

Figure 3: Phases of the proposed ECG Classification system.

A healthy person's average rate of heart beat is 72 beats per minute. For a Normal Person, it will be in between 70–80 beats per minute. In a normal sinus rhythm, the Abnormalities like Bradycardia means a

resting heart rate of below 60 bpm and Tachycardia will have a heart rate above 90 bpm. Heart rate can be measured by using the formulae

Heart rate=60 /RR interval in seconds -------Equ(1)

3.1 PRE-PROCESSING

A. NOISE REMOVAL USING FILTERING METHOD

Electrocardiogram signal inalienably contains of different kind of undesirable noise and a piece of rarity impacts like motor artifacts, baseline drift, polarization noise, the internal amplifier noise, noise due to muscle movement, noise of electrode contact and, such noise stimulate artifacts noise in the ECG signal.

Accordingly with the final goal, to make the ECG signal prepared for feature extraction step, one should expel baseline drift and dispose the noise the present in the ECG by utilizing Band pass channel which is built by a techniques know as high pass filter. This channel takes out baseline dissimilarity. The outcome of this channel is fall with a low pass channel. This channel evacuates high recurrence noise.

3.2 FEATURE EXTRACTION AND FEATURE SELECTION

Feature selection and Extraction process used in any research play an important place in terms of size reduction. Each feature that additionally used in the classification scheme will drastically increase the cost of the system in the means of calculation and the runtime. Therefore, it is important to develop the system and model using fewer features.

3.2.1. FEATURE EXTRACTION

Usually some preferred features were used in ECG Arrhythmia classification, such as subspace domain feature, statistic based feature, time domain feature and morphological features[7,8]. In this paper, a sample ECG beat of 200 points were used to calculate the space domain and time domain features like, Mean Heart Rate, Standard Deviation Heart Rate, Percentage of Number of R peaks in ECG, Heart Rate Variability and Power Spectral Density.

- **Mean Heart Rate**
  This feature indicates the mean value of the heart rate within the range of one minute in all segments
- **Standard Deviation Heart Rate**
  This feature denotes the standard deviation of instantaneous heart rate in all segments
- **Percentage of Number of R peaks in ECG**
  This feature indicates the quantity of consecutive dissimilarity of 64 R-R intervals that fluctuates higher than 50 ms, correspondingly, divided by 64.
- **Heart Rate Variability**
  This feature indicates the integral of the histogram, overall amount of RR intervals to the altitude of the histogram
- **Power Spectral Density**
  This Feature computed the HF and LF bands and the relative quantity of the LF and HF bands power (LF/HF) is taken as the frequency domain feature of the HRV.

3.2.2. FEATURE SELECTION

A. PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization algorithm is a quick, straightforward and proficient population based optimization strategy mostly used evolutionary computation tool in Genetic algorithm calculations. Such Genetic algorithm is also mentioned as Bio Inspired Algorithms [11]. In PSO, population of particles exists in the n-dimensional hunt space. Every molecule has certain measure of learning and will move about the inquiry space based on this information. The molecule has some latency in certified to it and thus will keep on having a segment of movement toward the path it is moving. The molecule knows its area in the search space and will experience with the best arrangement, which helps the further process pick up the best node among the population.

B. BACTERIAL FORAGING OPTIMIZATION

Bacterial Foraging Optimization algorithm is a bio inspired algorithm based on the Chemotaxis process of the bacteria named, E.Coli Bacterium. The two process used by the E.Coli bacteria at the time of Chemotaxis is Tumbling or swimming in the clockwise direction or counter clockwise direction to find the nutrient gradient[12]. It can survive by moving for a longer distance based on the two movement process and it adopts the nutrient gradient and avoid noxious environment. This technique was adopted to find the needed features which gives optimum solution for the classification process.

3.3 CLUSTERING

A. FUZZY C MEANS CLUSTERING

A Clustering technique that includes in limiting and gathering the specified functions based on the given data. At that point, when a process can limit the errors using the functionality then it is mentioned as C-Means, such that, 'c' the quantity of classes or bunches, and if the utilized classes are utilizing the Fuzzy procedure or essentially Fuzzy, at that point it is known to be Fuzzy C Means Clustering technique. The FCM approach utilizes a Fuzzy participation which allots a degree of membership for each class[14]. The significance of degree of membership in fuzzy clustering is like the pixel likelihood in a blend reproduction statement. The advantage of using Fuzzy C Means Clustering is the development of new clusters from the information focuses that have close participation esteems to existing classes. Fundamentally, there are three essential functionalities in FCM clustering technique. i) The Creation of Fuzzy Membership Function, ii) Inference Engine iii) the Crisp Values for the Objective function.

B. EXPECTATION MAXIMIZATION CLUSTERING

The Expectation Maximization, simply called as EM algorithm, which produces Maximum Likelihood (ML) estimation for the parameters, in a critical situation, where there is a mapping structure like many-to-one, from a hidden dissemination to the distribution governing the perception. The EM calculation comprises of two noteworthy phases. The first one is i) An Expectation Step, ii) A Maximization Step. In the expectation step the hidden factors utilizing the current estimate of the parameters and adapted upon the perceptions. The Maximization step at that point which gives an improvised or a new estimation of the parameters.

3.4 CLASSIFICATION

Classification of ECG signals is a demanding issue due to problems associated with grouping the process data's[9]. The most important issues in ECG characterization are absence of consistency in ECG features, changeability among the ECG features, independence of the ECG designs, non presence of ideal characterization rules for ECG grouping, and changeability in ECG waveforms of patients[10]. In this paper, by keeping the above issues in the classification problem, we selected two classification algorithms to classify the ECG signals and predict the abnormalities present in the ECG signal.

A. SA-SVM

Self Advising Support Vector Machine is basically a quadratic programming problem depends on the function of the support vectors that present in the training dataset. Basically, Support Vector Machine shows a successful methodology and obtained a good result among the text classification methods. Self-Advising Support Vector Machine attempts to create ensuing learning from the misclassified information from the training phase of the SVM [13].

This misclassified information can originate from two possible places such as Outliers or as information that have not been directly isolated by utilizing any kind of kernels. Traditional SVM disregards the training data that has not been isolated straightly by kernels in the training stage. Self-Advising Support Vector Machine proposed to manage the disregarding of the information that can be separated from the misclassified information. This should be possible by creating Advice weights dependent on utilizing of misclassified training data, if conceivable, and utilize these weights together with choice estimations of the SVM in the test stage. These weights assist the calculation with eliminating the outlier information.

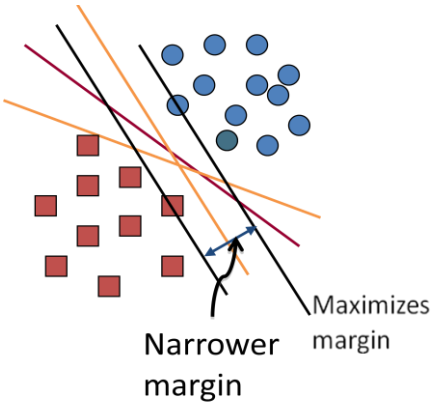

Figure 4: SA-SVM Classification based on the selected features.

B. KNN

K- Nearest Neighbor, simply denoted as KNN, is a non-parametric technique and is generally utilized in Classification and pattern recognition. As indicated by its name, it processed based on the distance measures, the test set is allocated to the class correspond to the larger part of its k closest neighbors from the training set. The k number of nearest neighbors is found by choosing the k neighbors which are insignificantly far off to the test data set. In KNN,

the distance measure of the nodes present in the data is viewed as just in choosing the k closest neighbors. The node assignment depends on larger part of a vote and the vote of every k closest neighbor has the equivalent weightage.

This would imply that the vote of a neighbor which is exceptionally close to the test data set and that of one which is to a greater degree, which is far away convey a similar weight on the off chance that they are among the k closest neighbors choice. Likewise, if the estimation of k number that picked is even, there is the likelihood of a tie happening because of casting a positive vote. This must be settled by arbitrarily appointing a class to the test from among the classes getting a similar number of greatest votes. This technique requires a specific parameter that should be picked dependent on which characterization will be completed. All examples present in training data set have their own esteem. The Major functionalities in the KNN are learning and Classification. Learning means storing all training instances and Classification means assigning target function to a new instance.

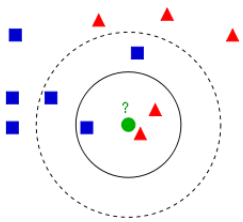

Figure 5: KNN Classification based on the selected features.

4. EXPERIMENTAL RESULTS

A. PERFORMANCE MEASURES

The Performance Metrics like, Sensitivity, specificity, Accuracy, Precision and Recall were used in this paper to compare the proposed ECG classification System based on the executed results.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \quad \text{----Equ(2)}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100 \quad \text{----Equ(3)}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad \text{----Equ(4)}$$

Sensitivity means No. of proportion of patients with a disease who's test results shows positive. In other words, Sensitivity means the test correctly identifies the patient, who has the Cardiac Vascular disorder.

Specificity means No. of proportion of patients without the disease who's test results shows negative. In other words, Specificity means the test correctly identifies the patient, who does not have the cardiac vascular disorder.

Accuracy describes the correctness of the measurement that predicts correct value.

B. EXPERIMENTS

The performance of the ECG signal classification system was depend on numerous important factors including filtering process for noise removal, Classification Methods selected, dataset used for experimental purpose etc.. The ECG Signals used for the Experimentation was downloaded from the MIT-BIH Arrhythmia database. The framework was designed and the entire programming was created using Matlab 2016a. Totally 1062 samples of beats were used in that 802 samples were recorded from Female patients and 260 samples were recorded from Male patients.

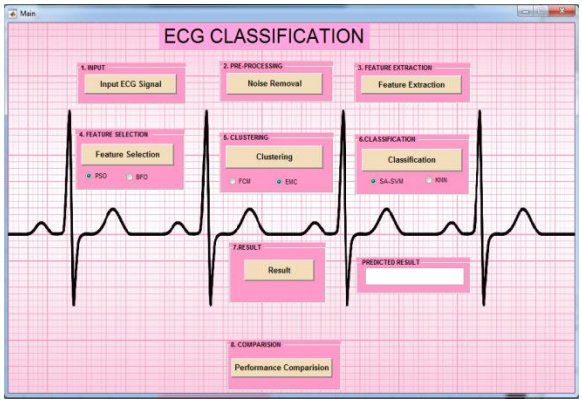

Figure 6:Proposed ECG Classification Framework

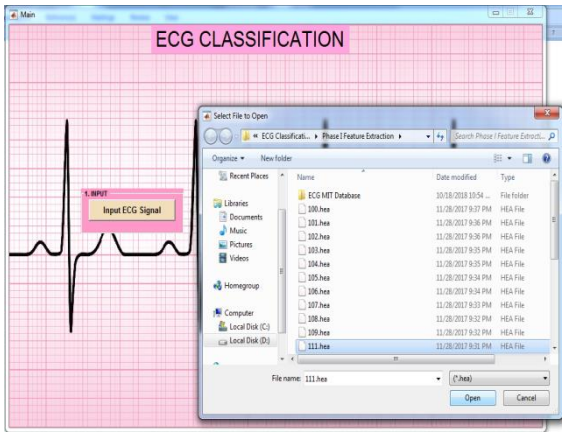

Figure 7 : Input file for ECG signal Classification

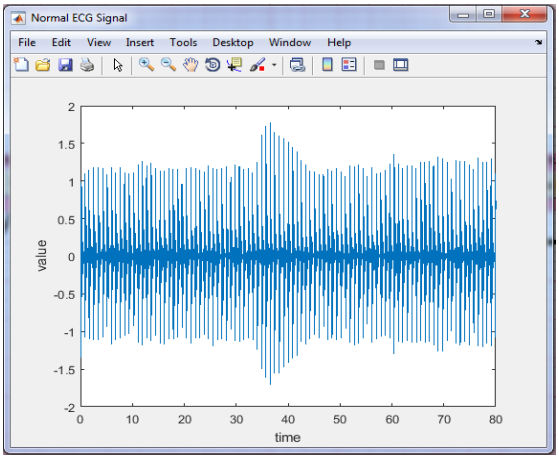

Figure 8: Normal ECG Signal

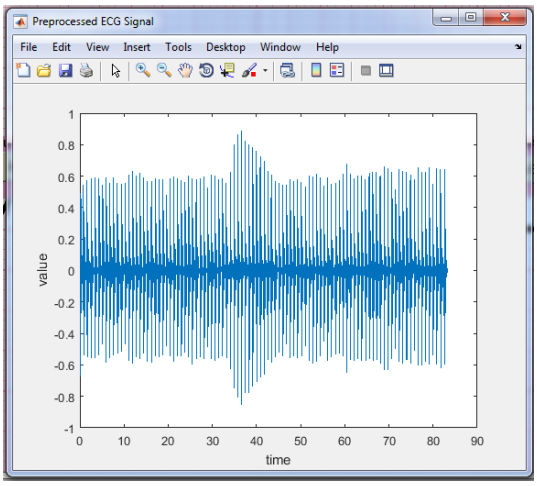

Figure 9 : Pre-processed ECG Signal

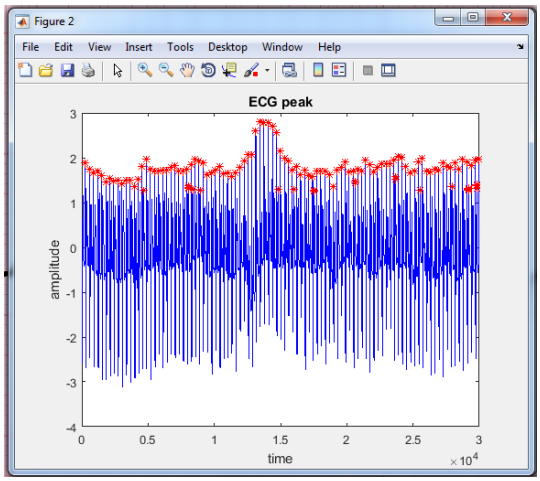

Figure 10: R - Peak Detection for ECG Signal

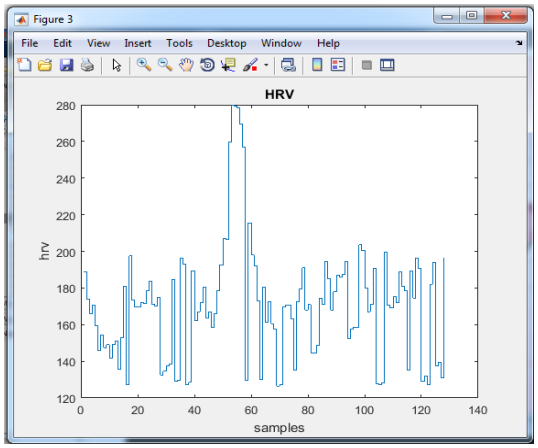

Figure 11 : Heart Rate variability for the given ECG Signal

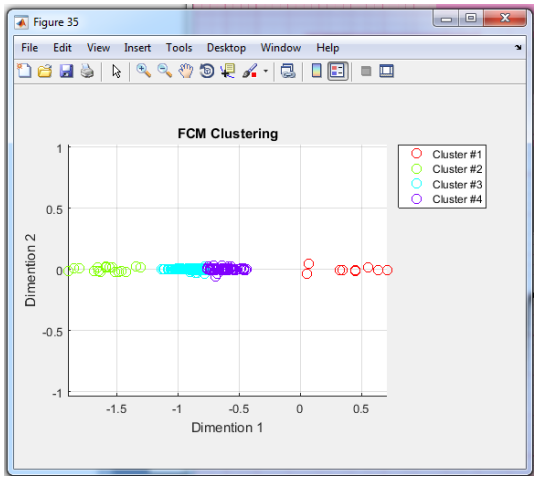

Figure 12: Fuzzy C Means Clustering for the selected features

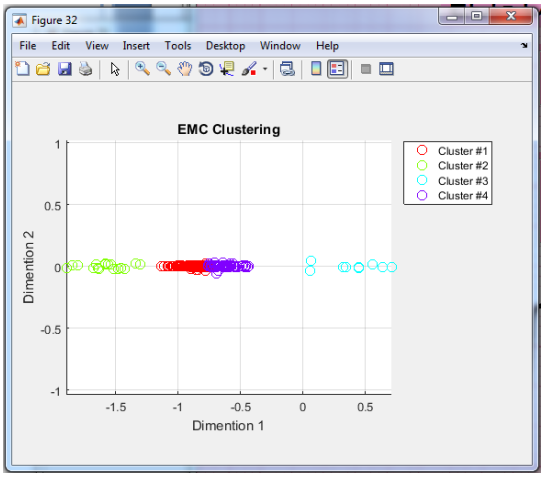

Figure 13 : Expecation Maximization Clustering for the selected features

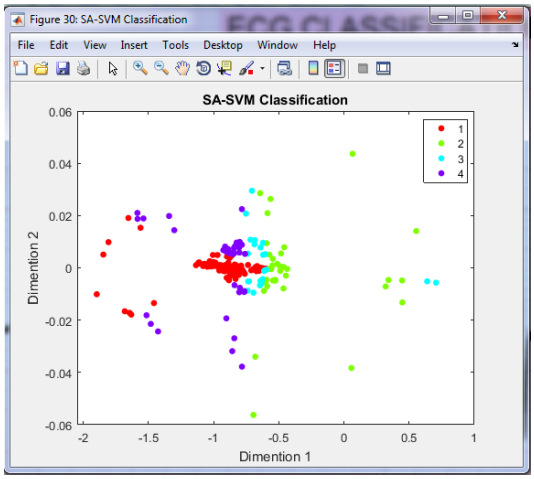

Figure 14 : Self- Advising Support Vector Machine classification for the selected features

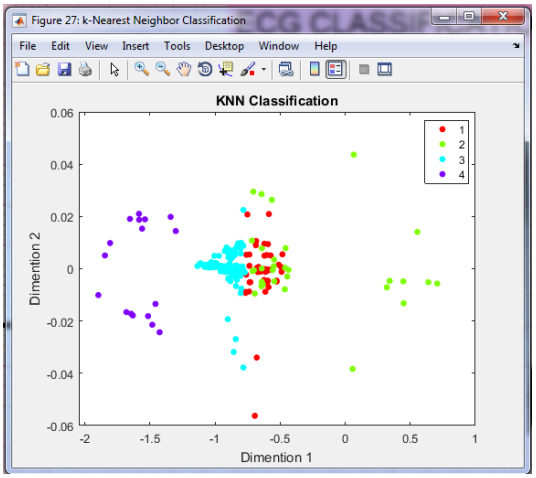

Figure 15 : K Nearest Neighbor Classification

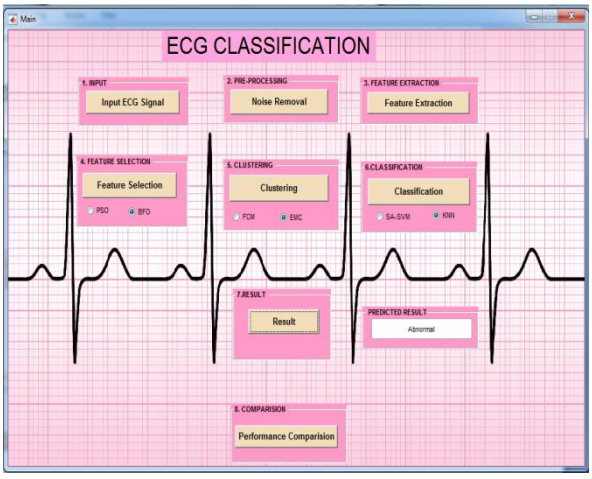

Figure 16 : Framework for the final Predicted result.

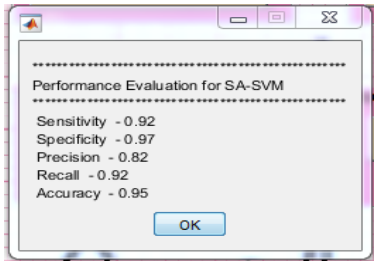

Figure 17 : Performance Evaluation for SA-SVM Classification Algorithm

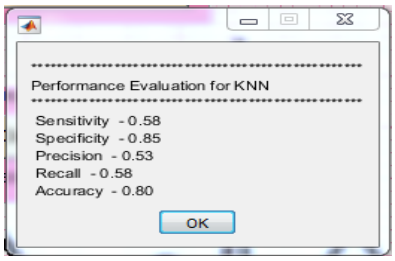

Figure 18 : Performance Evaluation for the KNN Classification Algorithm

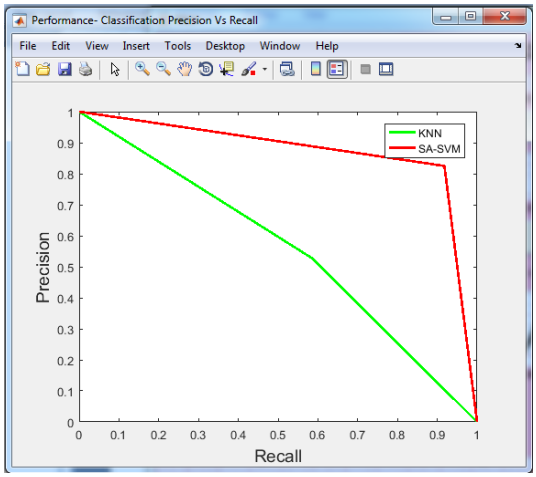

Figure 19 : Precision and Recall comparison for SA-SVM and KNN

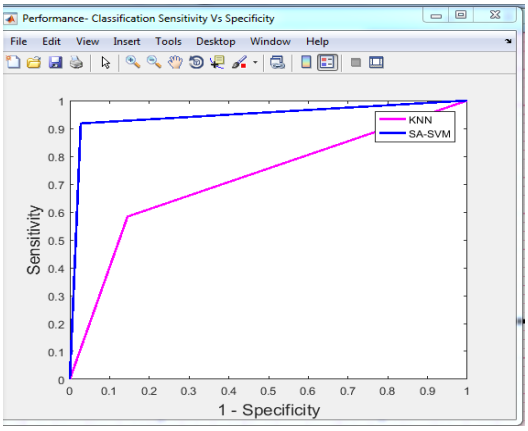

Figure 20 : Sensitivity and Specificity Comparison for SA-SVM and KNN

5. CONCLUSION

In this paper, a novel methodology for ECG signal classification was proposed. The sample ECG signal dataset was taken from the MIT-BIH Arrhythmia data base. To diagnosis the cardiac vascular diseases, ECG Signal Analysis plays an important role in it, so, the researchers put forth some additional contribution towards the classification of ECG signals. This paper also contributed a novel methodology to find the abnormalities in the ECG using computer Aided diagnosis system. To achieve a good performance in the proposed ECG classification system, the following techniques were adopted, such that, Feature Extraction, Feature Selection, Clustering and Classification. From the results mentioned in the Experimental Results section, the combination of Particle Swarm Optimization and Fuzzy C Means Clustering along with Self Advising Support Vector Machine yields a good result of Accuracy 95% while compared with the Bacterial Foraging Optimization and Expectation Maximization Clustering along with K Nearest Neighbor algorithm of Accuracy 80 %.

REFERENCES

[1] Sheng-Wu Xiong, Hong-Bing Liu and Xiao-Xiao Niu. (2005). "Fuzzy support vector machines based on FCM clustering". Proceedings of 2005 International Conference on Machine Learning and Cybernetics, vol. 5, pp. 2608- 2613, 2005.

[2] Alshraideh H, Otoom M, Al-Araida A, Bawaneh H, ve Bravo J. "A web based cardiovascular disease detection system". J Med Syst, Vol: 39, P: 122, 2015.

[3] Melgani F, Bazi Y. "Classification of electrocardiogram signals with support vector machines and particle swarm optimization". IEEE Trans Inf Technol Biomed, Vol:12, P: 667-677, 2008.

[4] Güler İ, Übeylı ED. "ECG beat classifier designed by combined neural network model". Pattern Recognition, Vol.38, P:199-208, 2005.

[5] Khalaf AF, Owis MI, Yassine IA. "A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines". Expert Syst Appl; Vol.42, P: 8361-8368, 2015.

[6] Escalona-Morán MA, Soriano MC, Fischer I, ve Mirasso CR. "Electrocardiogram classification using reservoir computing with logistic regression". IEEE Journal of Biomed Heal Information, Vol. 19: P: 892-898, 2015.

[7] Dilmac S, ve Korurek M. "ECG heart beat classification method based on modified ABC algorithm", Application Soft Computing Journal, Vol:36, P:641-655, 2015.

[8] Rashad Ahmed, Samer Arafat, "Cardiac Arrhythmia Classification Using Hierarchical Classification Model", 6th International Conference on Computer Science and Information Technology, 2014.

[9] Y. N. Singh, S. K. Singh, and A. K. Ray, "Bioelectrical signals as emerging biometrics: Issues and challenges," ISRN Signal Processing, pp. 1-13, 2012.

[10] M. E. A. Bashir et al., "Highlighting the current issues with pride suggestions for improving the performance of real time cardiac health monitoring," Inform. Technology in Bio-and Medical Informatics, ITBAM, Springer Berlin Heidelberg, pp. 226-233, 2010.

[11] Sandeep Raj , Kailash Chandra Ray, " ECG Signal Analysis Using DCT-Based DOST and PSO Optimized SVM", IEEE Transactions on Instrumentation and Measurement (Volume: 66 , Issue: 3 , March 2017.

[12] Das S., Biswas A., Dasgupta S., Abraham A. "Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications", Foundations of Computational Intelligence Volume 203, Springer, 2009.

[13] YasharMaali, AdelAl-Jumaily, "Self-advising support vector machine", Knowledge-Based Systems Vol: 52, P: 214-222, 2013.

[14] Zhu, C.J., Yang, S., Zhao, Q., Cui, S., Wen, N., "Robust semi-supervised Kernel-FCM algorithm incorporating local spatial information for remote sensing image classification", Journal of Indian Society in Remote Sensors, Vol: 42(1),P: 35–49, 2014.

[15] M.Akhil jabbar* B.L Deekshatulua Priti Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm", International Conference on Computational Intelligence: Modeling Techniques and Applications, Procedia Technology, Vol:10, P:85 – 94, 2013.