IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 16-22

Clustering the ECG Signals Using Fuzzy C Means Clustering Technique

D. Raj Balaji [1], Dr. A. Marimuthu [2]
[1](Ph.D Research Scholar, Manonmaniam Sundaranar University, Tirunelveli.)
[2](Associate Professor and Head, Department of Computer Science, Government Arts College, Coimbatore)

Abstract: *The Electrocardiogram is a measurement used to evaluate of the electrical activity of the heart. An Electrocardiogram, mentioned as ECG, signals becomes a vital role in diagnosing Cardiovascular Diseases. A variety of Cardiovascular Diseases like, Arrhythmia, Atrial Fibrillation, Atrio Ventricular Dysfunctions, and Coronary Arterial Disease, etc. may be detected non-invasively by means of ECG monitoring devices. Among these types of cardiovascular diseases, here Cardiac Arrhythmias is particularly taken as a problem domain. The techniques that classify cardiac arrhythmias based on the features that present in ECG measures still have poor accuracy and need much more learning time. With the arrival of contemporary Signal Processing, Data Mining and Machine Learning techniques, the diagnostic power of the ECG has prolonged exponentially. In this paper, a framework was proposed to find the finest feature extraction technique that fabricate a perfect model for clustering P-QRS-T waves present in the ECG signals that find the cardiac arrhythmias. The framework constructed contains four phases, first is the pre-processing step that is initiated to clear the noise that present in the ECG signal while recording. Next is the feature extraction techniques which uses six techniques to extract the hidden features that present in the ECG measures. Then Feature selection process is carried out to select the important feature that predicts the status of the heart beat as well as to minimize the load in features. finally Clustering technique is used to group the ECG signals based on the selected features to give accurate result. Our results proved that the clusters grouped by the Fuzzy C Means Clustering gives more prediction, while comparing with the Expectation Maximization Clustering technique. Thus Fuzzy C Means can give clear feature sets for further classification of ECG.*

Keywords: *Electrocardiogram, P-QRS-T wave, Clustering, FCM, EM.*

I. Introduction

Heart is consider as one of the most vital organ of Human body. Cardiovascular diseases, simply called as CVD, are the preferential reason for the common death around worldwide. Ancient techniques of Visual analysis of ECG for doctors are sophisticated and time consuming task. Visual analysis wants experience to ascertain the problems in ECG [9]. It is an excellent and low price methodology to find the abnormalities that present in the heart.

Electrocardiogram pattern and Hear rate variability may got to be determined over several hours. so the quantity of the information being large, the study is tedious and time overwhelming. Therefore, computer-based investigation and classification of diseases may be really helpful in diagnosing the abnormalities present in the heart [10]. Since visual analysis of long-run recordings of the hearts activity, in a very single recording, is tough to diagnose and can be particularly error prone, automatic computer aided analysis is of major significance [11]. it is a challenge to extract the foremost common and most important theme from amorphous raw Electrocardiogram data.

There are numerous contributions are created and implemented in technological field for electrocardiogram signal classification and pattern recognition. The graph (ECG) is that the record of movement away from the P-QRS-T wave signal induced by the beats produced in the human heart. The ECG Signal detects and shows the activity of the heart and based on that one can give the appropriate treatment to safeguard the patient. it is the most vital and useful methodology used in identification of the condition of the cardiovascular diseases.

Electrocardiography (ECG) manages the electrical action of the focal of the blood circulatory framework, i.e. the heart. Observed by putting sensors at the appendage furthest points of the subject, electrocardiogram (ECG) is a record of the birthplace and the engendering of the electrical potential through cardiovascular muscles [12]. In this way, ECG is a critical non-intrusive clinical device for the analysis of heart illnesses.

The remaining section of this paper was arranged based on the following segment. In section 2, the literature review about the Feature Selection, Feature Extraction ECG Clustering and its related study done by the previous researcher was discussed. In section 3, the proposed framework, Phase divisions, Six feature

extraction techniques that retrieved from the ECG signal. The number of features extracted were reduced by using the feature selection methodology such as Particle Swarm Optimization and Bacterial Foraging Optimization to speed up the clustering process. The reduced features of ECG signals are clustered by using Fuzzy C Means Clustering algorithm and Expectation Maximization algorithm. 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.

II. Literature Review

Fuzzy C Mean clustering is one amongst the powerful and user friendly technique used to group the information supported by some conditions. it is part of an empire in image process and signal process applications [1].
Marius et al. used FCM in feeling recognition through voice signal. They achieved a most mean recognition rate of 63% [2].
Murugappan et al. has enforced FCM in his sensation study through Electrocardiogram signal [3].
Yang et al. used fined FCM clustering technique in image segmentation associated, managed to realize and produced a methodology, which is more strong to noise[4].
In spite of incredible success of the Expectation Maximization algorithmic process in performance because of its simplicity and quick initial progress, some authors [5] have argued that the speed of EM convergence is extraordinarily slow, which a lot of sophisticated second-order strategies ought to usually be favored to EM.
Several strategies are projected to reinforce the convergence speed of the EM algorithmic process, by using largely supported standard optimization theory [6, 7].
Many authors [5, 8] have additionally projected hybrid approaches for Machine learning, recommending change to a Newton or Quasi-Newton methodology once performing many EM iterations.

III. Methodology

In this paper, a framework for clustering the ECG signals based on the selected features were discussed and for that, we selected Fuzzy C Means Clustering technique and Expectation Maximization algorithm, which was mentioned in the overall architecture of the proposed framework depicted in figure 1.

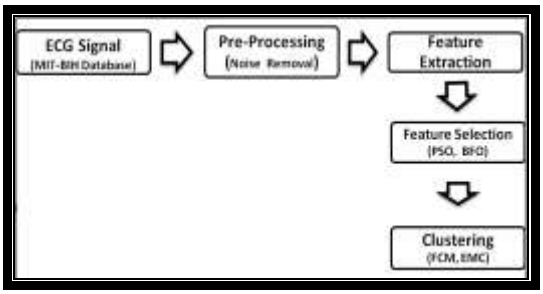

Figure 1: Overall Architecture of the Proposed ECG Classification System using Clustering Technique.

The Phase division for the proposed framework was mentioned in figure 2. Totally, four phases were derived in this paper, such as, Pre-processing, Feature Extraction, Feature Selection and Clustering.

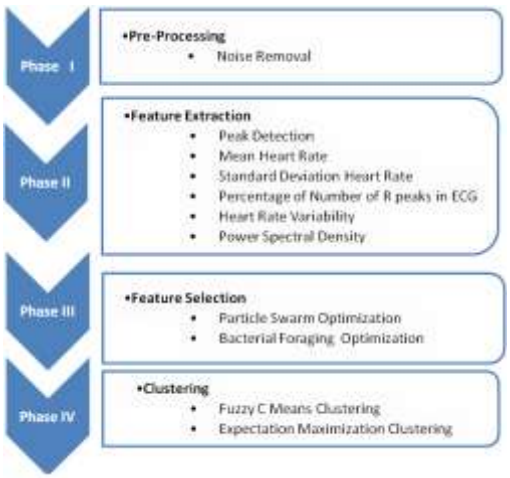

Figure 2: Phases of the proposed ECG Classification system using Clustering Technique .

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[13,14]. 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[15]. 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[16]. 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.

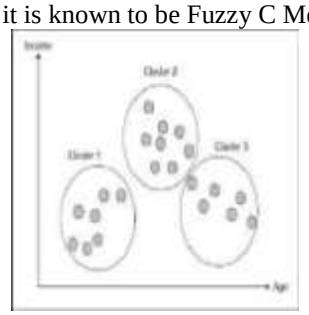

Figure 3 : Fuzzy C Means Clustering

The FCM approach utilizes a Fuzzy participation which allots a degree of membership for each class[17]. The significance of degree of membership in fuzzy clustering is like the pixel likelihood in a blend reproduction statement.

Fuzzy C Means Algorithm

Step-1: Randomly initialize the membership matrix using this equation.

$$\sum_{j=1}^{k} \mu_j(x_i) = 1 \qquad i = 1, 2 \ldots k$$

Step-2: Calculate the Centroid using equation.

$$Cj = \frac{\sum_i [\mu_j(x_i)]^m x_i}{\sum_i [\mu_j(x_i)]^m}$$

Step-3: Calculate dissimilarly between the data points and Centroid using the Euclidean distance.

$$D_i = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Step-4: Update the New membership matrix using the equation.

$$\mu_j(x_i) = \frac{[\frac{1}{d_{ji}}]^{1/m-1}}{\sum_{k=1}^{c} [\frac{1}{d_{ki}}]^{1/m-1}}$$

Here **m** is a fuzzification parameter.
The range **m** is always [1.25, 2]

Step -5: Go back to Step 2, unless the centroids are not changing.

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[18]. The EM calculation comprises of two noteworthy phases. The first one is i) An Expectation Step, ii) A Maximization Step.

Expectation Maximization (EM) Algorithm

log of expectation of P(x|z)

Goal: $\hat{\theta} = \arg\max_{\theta} \log\left(\sum_{z} p(\mathbf{x}, \mathbf{z} \mid \theta)\right) \qquad f(\mathrm{E}[X]) \geq \mathrm{E}[f(X)]$

1. E-step: compute

expectation of log of P(x|z)

$$\mathrm{E}_{z|x,\theta^{(t)}}[\log(p(\mathbf{x}, \mathbf{z} \mid \theta))] = \sum_{z} \log(p(\mathbf{x}, \mathbf{z} \mid \theta)) p(\mathbf{z} \mid \mathbf{x}, \theta^{(t)})$$

2. M-step: solve

$$\theta^{(t+1)} = \arg\max_{\theta} \sum_{z} \log(p(\mathbf{x}, \mathbf{z} \mid \theta)) p(\mathbf{z} \mid \mathbf{x}, \theta^{(t)})$$

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.

IV. Experiments And Result

A. Dataset

Electrocardiograms (ECGs) signals usually utilized in perceptive of the functioning of the heart, as a result of the working effect, one can safe guard the lives by finding the abnormalities that present in the heart early. In 1961, Holter records the ECG for period of twenty four hours, such long run ECG paves the way to establish the arrhythmias database. additionally thereto, 1975, Massachusetts Institute of Technology's laboratory supported to supply the MIT-BIH heart disease information, that was the primary offered set of ordinary data set for the heart disease analysis. The MIT-BIH heart disease information consists of forty eight annotated records collected from forty seven subjects studied at the BIH capital of Massachusetts. Among that 60% of datasets were collected from inpatients and 40% datasets were collected from outpatients.

B. Performance Measures

The performance of the ECG signal Clustering system was depend on numerous important factors including filtering process for noise removal, Methods selected for clustering, 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 5: Proposed ECG signal Clustering Framework

Figure 6 : Input file for ECG signal Clustering

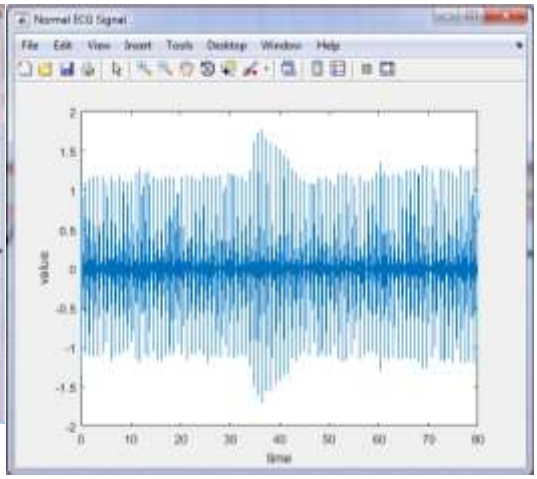

Figure 7: Normal ECG Signal

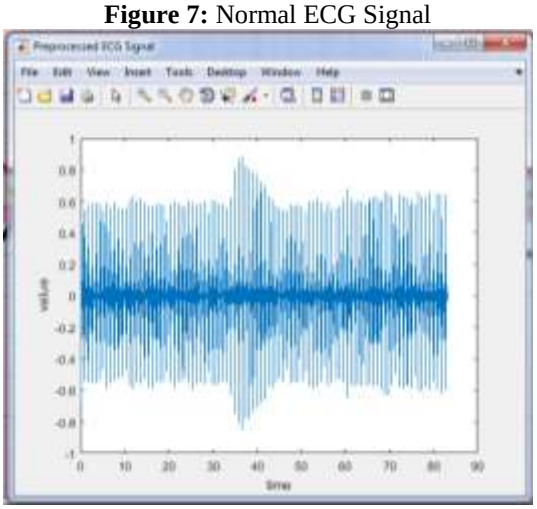

Figure 8 : Pre-processed ECG Signal

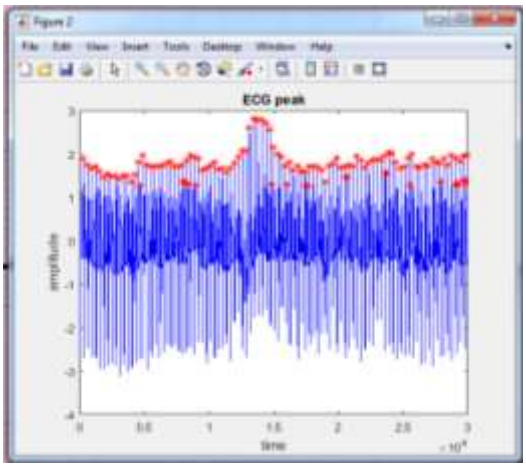

Figure 9: R - Peak Detection for ECG Signal

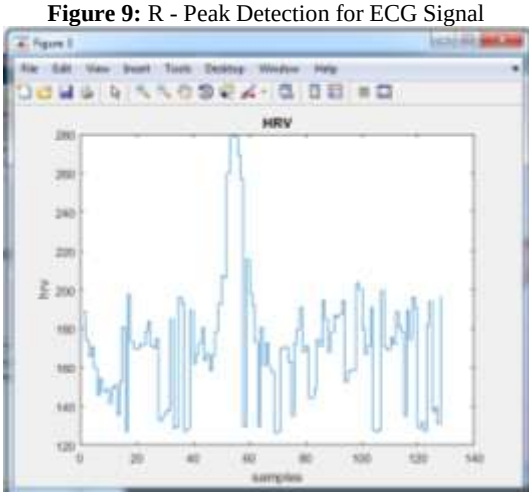

Figure 10 : Heart Rate variability for the given ECG Signal

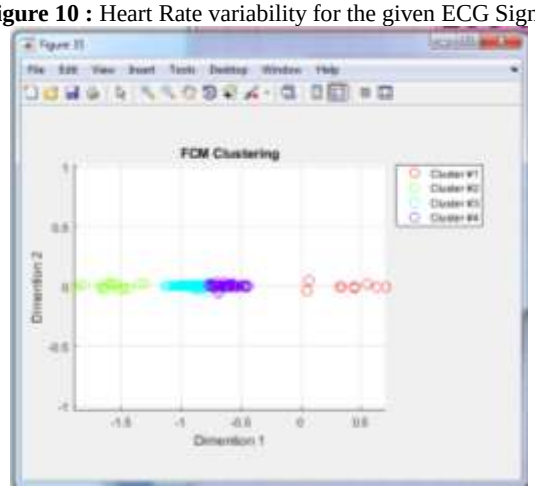

Figure 11: Fuzzy C Means Clustering for the selected features

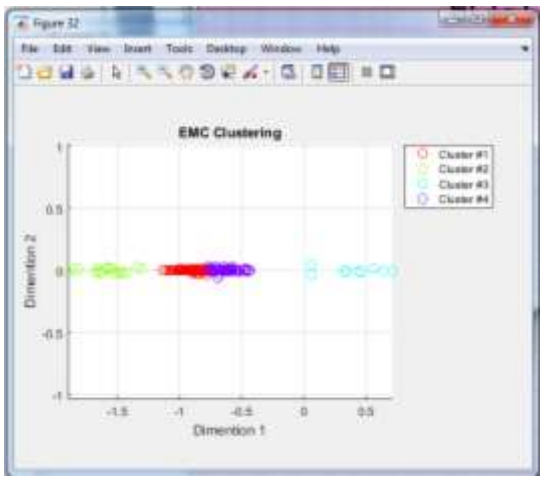

Figure 12 : Expecation Maximization Clustering for the selected features

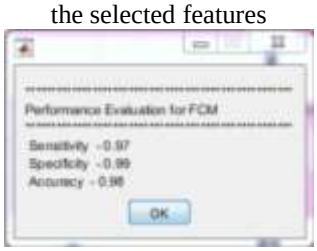

Figure 13 : Performance Evaluation for FCM Clustering Algorithm

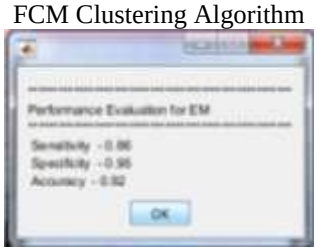

Figure 14 : Performance Evaluation for EM Clustering Algorithm

V. Conclusion

As a result by developing a system for predicting abnormalities using clustering techniques based on the features extracted from the ECG signal. The Clustering techniques like Fuzzy C Means Clustering algorithm and Expectation Maximization Clustering Algorithm was used in this paper. Six features were extracted from the ECG signal and from that six features, selective features were used for the clustering the signals that present in the ECG signal. We have a tendency to compare the results between the two systems to urge that one is the best than the other. Based on the result and discussion section, the Fuzzy C Means Clustering technique having 98 % of accuracy shows better result when comparing with the Expectation Maximization algorithm having 92% of accuracy.

References

[1]. J. C. Bezdec. "Pattern recognition with fuzzy objective function algorithms", Plenum Press, New York, 1981.

[2]. M. Zbancioc and M. Feraru, "A study about the statistical parameters used in the emotion recognition", *Proc. of the 10th International Conference on Development and Application Systems, Suceava, Romania, ISSN: 1844-5020, pp. 187-190, 2012.*

[3]. M. Murugappan, M. Rizon, R. Nagarajan and S. Yaacob. 'EEG Feature Extraction for Classifying Emotions using FCM and FKM', *Proc. of the 7th WSEAS Int. Conf. on Applied Computer and Applied Computational Science 1, pp. 299-304, 2008.*

[4]. Y. Yang, Ch. Zheng and P. Lin, "Fuzzy C-means clustering algorithm with a novel penalty term for image segmentation," *Opto Electronics Rev., vol. 13, no. 4, 2005.*

[5]. Richard A. Redner and Homer F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review, 26(2):195-239, 1984.*

[6]. Mortaza Jamshidian and Robert I. Jennrich. Acceleration of the EM algorithm by using quasi-newton methods. J. of the RS Society series B, *49, 1997.*

[7]. Meng X. L. and van Dyk D. Fast EM-type implementations for mixed e_ects models. J. of the Royal Statistical Society series B, *60:559-578, 1998.*

[8]. S.E. Atkinson. The performance of standard and hybrid EM algorithms for ML estimates of the normal mixture model with censoring. J. of Stat. Computation and Simulation, *44, 1992*.

[9]. O.Sayadi and M. Shamsollahi, "Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a Bayesian filter," *IEEE Transactions on Biomedical Engineering, vol. 58, no. 10, pp. 2748–2757, 2011.*

[10]. Dipti Patra and Smita Pradhan, "Integration of FCM, ICA and Neural networks for ECG signal classification," *In Proceedings of International Conference on Soft Computing (ICSC 2008), 2008.*

[11]. E. M. Tamil, N. H. Kamarudin, R. Salleh, M. Yamani Idna Idris, M. N. Noorzaily, and A. M. Tamil, "Heartbeat electrocardiogram (ECG) signal feature extraction using dis-crete wavelet transforms (DWT), in *Proceedings of CSPA, 1112–1117, 2008.*

[12]. B.Anuradha and V.C.Veera reddy, "Cardiac arythmia classification using fuzzy classifiers", *Journal of theoretical and applied information technology, P: 353-357, 2008.*

[13]. 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*.

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

[15]. 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.*

[16]. 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.*

[17]. 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.*

[18]. Osama Moh'd Alia, Rajeswari Mandava, Mohd Ezane Aziz, "A hybrid harmony search algorithm for MRI brain segmentation", *Evolutionary Intelligence, 2011.*