

Classification of Heart Rate Data Using BFO-KFCM Clustering and Improved Extreme Learning Machine Classifier

R.kavitha¹, T.Christopher²

¹Department of computer science, PSGR Krishnammal College for women, Coimbatore, India

²Department of Research & PG computer science, Government Arts College, Coimbatore, India

Article Info

Article history:

Received Jan 15, 2016

Revised Feb 15, 2017

Accepted Feb 22, 2017

Keyword:

ECG signals classification
Improved Extreme Learning
Machine (ELM)
Kernel Fuzzy C-Means
Support Vector Machine
(SVM)

ABSTRACT

An Electrocardiogram or ECG is an electrical recording of the heart and is used in the investigation of heart disease. The heart rate varies not only in relation to the cardiac demand but is also affected by the presence of cardiac disease and diabetes. Furthermore, it has been shown that Heart Rate Variability (HRV) may be used as an early indicator of cardiac disease susceptibility and the presence of diabetes. Therefore, the heart rate variability may be used for early clinical screening of these diseases. The generalization performance of the SVM classifier is not sufficient for the correct classification of heart rate data. To overcome this problem the Improved Extreme Learning Machine (IELM) classifier is used which works by searching for the best value of the parameters, and upstream by looking for the best subset of features using Bacterial Foraging Optimization (BFO) that feed the classifier. In this work, nine linear and nonlinear features are extracted from the HRV signals. After the preprocessing, feature extraction is done along with feature selection using BFO for data reduction. Then, proposed a scheme to integrate Kernel Fuzzy C-Means (KFCM) clustering and Classifier to improve the accuracy result for ECG beat classification. The results show that the proposed method is effective for classification of heart rate data, with an acceptable high accuracy.

*Copyright © 2017 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

R.kavitha,
Department of computer science,
Krishnammal College for women,
Coimbatore, India.
Email: rkavitha2509@gmail.com

1. INTRODUCTION

ECG is a method to measure and record different electrical potentials of the heart. Willem Einthoven developed the ECG method in the early 1900s. The origin of the electrical activity measured by ECG is in the muscle fibers of different parts of the heart. The electrical potential generated by electrical activity in cardiac tissue is measured on the surface of the human body. Current flow, in the form of ions, signals contraction of cardiac muscle fibers leading to the heart's pumping action. It is a non persistent recording produced by an electrocardiographic device. The recognition and classification of the ECG beats is a very important task in the coronary intensive unit, where the classification of the ECG beats is essential tool for the diagnosis. ECG offers cardiologists with useful information about the rhythm and functioning of the heart.

Therefore, its analysis represents an efficient way to detect and treat different kinds of cardiac diseases up to now; many algorithms have been developed for the recognition and classification of ECG signal. Some of them use time and some use frequency domain for depiction. Based on that many specific attributes are defined, allowing the recognition between the beats belonging to different pathological classes.

The ECG waveforms may be different for the same patient to such extent that they are unlike each other and at the same time alike for different types of beats [1].

Heart Rate Variability is used to measure the variations in heart signals and more specifically variations per unit time of the number of heartbeats. The Electrocardiogram is the electrical activity of heart and generates electrical signals which are called as PQRST waves. The most important wave is the QRS complex. Heart beat depends on the time interval between two QRS complex waves which is called as R-R interval [2]. Normally Healthy persons present large values of HRV. Prediction of HRV analysis is one of the major research topics from last two decades. Several methods for heart rate variability were proposed among them, Spectral Methods based on FFT, nonlinear approach, including markov chain model are widely used [3, 4].

Neural Network and Machine Learning methods are one of the powerful methods to classify and predict the Heart rate variability patterns. Various studies have been done using different classification methods like Support Vector Machine, Neural Network, and Wavelet Transform PCA etc.

In this work, IELM is used for Classification of HRV data. By using IELM, not only reduce the classification time, but also improve the classification accuracy significantly. In this work after feature extraction the selection of best features is done using BFO which is used to reduce the dimension of features in Heartbeat classification with clustering using KFCM. The remaining part of the paper is organized as follows. Section 2 describes the preprocessing and feature extraction along with feature selection process. Section 4 describes the IELM system for HRV classification. Section 5 describes the experimental design including the evaluation results of this system. Finally, Section 7 concludes the paper.

2. MATERIALS AND METHODS

In this paper, the HRV signal explored as the basic signal to classify cardiac arrhythmias into five classes: Normal Sinus Rhythm (NSR), Premature Ventricular Contraction (PVC), Atrial Fibrillation (AF), Ventricular Fibrillation (VF) and and 2° Heart Block (BII). The HRV arrhythmia data, obtained using the ECG data from the MIT-BIH Arrhythmia Database. The analysis is carried out in three stages. First a preprocessing procedure is used to remove the interfering signals from the ECGs. In the second stage, time and frequency domain and nonlinear methods are applied to extract corresponding features. In the third stage the extracted features are reduced to attain the optimal features using BFO and based on these the clustering process is done using KFCM. These results are given as input to IELM classifier. The overall architecture diagram is illustrated in Fig.1. Next materials and methods are described. Then the different steps of the proposed algorithm are explained. Finally results obtained on the MIT-BIH arrhythmia database are presented.

a. Preprocessing And Feature Extraction Along With Feature Selection Process

2.1.1 Preprocessing the Signal

The HRV data used in this project is generated from the ECG signals provided by the MIT-BIH database. At first, it is necessary to extract the HRV signals from the ECG signals. Generally, many interfering signals such as the mains 50 Hz, the electromyogram (EMG) signals and also the baseline wandering can affect the extraction process. Hence, these interfering signals are removed from the input ECG signal using a 5-15 Hz band pass filter. Then the signal is processed using thresholding and the resultant signal is identified as HRV signal.

2.1.2 Feature Extraction

The next step in the block diagram is the feature extraction. Generally, the cardiovascular system demonstrates both linear and nonlinear behavior. Therefore, in this work, combinations of linear and nonlinear features are considered.

Linear analysis:

Time domain features

Four commonly used time domain parameters of the HRV signal which are directly extracted from the RR interval time series are:

- **Mean HR:** The mean value of the heart rate within one minute in each segment. Instantaneous heart rate (beat per minute) is equal to 60 divided by each R-R interval (second).
- **STD HR:** The standard deviation of Instantaneous heart rate in each segment.
- **pNN50:** The number of successive difference of 64 R-R intervals that differs more than 50 ms, respectively, divided by 64.
- **HRV triangular index:** This refers to the integral of the histogram (i.e. Total number of RR intervals) divided by the height of the histogram. A bin width of 1/128 is selected.

Frequency domain features

Although the time domain parameters are effective, they do not have the ability of discrimination between the sympathetic and parasympathetic contents of the HRV signal. High-frequency (HF) band (0.2-0.5 Hz) of HRV signal shows the cardiac vagal activities such as Respiratory Sinus Arrhythmia (RSA). In fact, HF components are considered as the origin of parasympathetic activities of the cardiovascular system. On the other hand, the low-frequency (LF) band (0-0.2 Hz) is related to the baroreceptor control and is mediated by sympathetic systems. In this project, the power spectral density (PSD) for the HF and LF bands are calculated and the ratio of the LF and HF bands power (LF/HF) is considered as the Frequency domain feature of the HRV signal.

Nonlinear analysis

The HRV signal analysis by use of methods on nonlinear dynamics leads to very valuable information for physiological interpretation of the heart. Hence, four different nonlinear parameters of the HRV signal are used in this work.

- **SD1/SD2:** Poincare plot is a graphical representation of the correlation between successive RR intervals. This is obtained plotting each RR interval (RR (n+1)) as a function of the previous interval (RR (n)) in RR interval time series. This plot is quantitatively analyzed calculating the standard deviation of the distances of the time series points from the lines $y = x$ and $y = x + 2RR_m$, in which RR_m is the mean of all values of RR interval time series. These values are named SD1 and SD2 respectively. In fact, SD1 represents the fast beat-to-beat variability, while SD2 describes the relatively long-term variability in the HRV signal. In this work, SD1/SD2 is used as the first nonlinear feature which is extracted from HRV segments.
- **LLE:** The Largest Lyapunov Exponent provides useful information about the dependency of system on initial conditions and a positive Lyapunov exponent confirms the existence of chaos in the system. For calculating LLE, a point is selected in the reconstructed phase space of the system and all neighbor points residing within a predefined radius ϵ are determined. As the system evolves, the mean distances between the trajectory of the initial point and the trajectories of the neighbor points are calculated. Then the logarithm of these mean values plots against the time and the slope of the resulting line are considered as LLE. The embedding dimension and the lag are selected to be $m = 10$ and $t = 1$, respectively. The threshold distance ϵ is selected to be mSD , where SD is the standard deviation of the RR time series.
- **SpEn:** The Spectral Entropy shows the complexity of the input time series (HRV segment) in the frequency domain. Large values of SpEn show high irregularity and smaller values of it indicate more regular time series. The Shannon's channel entropy is used to estimate the spectral entropy of the process as:

$$SpEn = - \sum_f p_f \log p_f$$

where p_f is the PDF (probability density function) value at frequency f . Heuristically, the entropy is interpreted as a measure of uncertainty about the event at f . Thus entropy can be used as a measure of system complexity. The spectral entropy H describes the complexity of the HRV.

- **D2:** The Correlation Dimension is a measure of complexity of the time series and determines the minimum number of dynamic variables which can model the system.

2.1.3 Feature Selection Using BFO

In this section, features of heart rate are extracted. The extracted features are reduced further by using Bacteria Foraging Optimization to remove redundancy and irrelevant features. The resulting feature subset (obtained by BFO) is the most representative subset and is used to improve the classification result.

Bacteria Representation

Each bacteria's position represent one possible solution (feature subset) required for face recognition. The number of dimensions of search space is m where m is the length of feature vector (FV) extracted from section 2.1.2. In each dimension of search space, bacteria position is 1 or 0, where 1 or 0 indicates that this feature is selected or not selected, respectively, as required feature for next generation. In the each iteration of chemo taxis step, each bacteria tumbles to the new random position. Position of i th bacteria in j th chemo taxis and k th reproduction step is defined as:

$$\theta^i(j, k) = F_1 F_2 \dots F_m \quad (1)$$

Where, m is the length of feature vector extracted. Each $F_z = 1$ or 0 ($z = 1, 2, \dots, m$) Depending upon whether z th feature is selected or not for the next iteration.

Fitness Function: In each generation, each bacterium is evaluated, and a value of goodness or fitness is returned by a fitness function. This evolution is driven by the fitness function F [5]. Let w_1, w_2, \dots, w_L and S_1, S_2, \dots, S_n denote the classes and number of signals within each class, respectively. Let M_1, M_2, M_L and M_0 be the means of corresponding classes and the grand mean in the feature space, M_i can be calculated as:

$$M_i = \frac{1}{S_i} \sum_{j=1}^{S_i} W_j^{(i)}, \quad \forall i = 1, 2, \dots, L \quad (2)$$

Where (i) $W_j, j = 1, 2, \dots, S_i$, represents the sample heart rate signal from class w_i and grand mean M_0 is:

$$M_0 = \frac{1}{n} \sum_{i=1}^L n_i M_i \quad (3)$$

Where n is the total number of heart rate data signal of all the classes. Thus the between class scatter fitness function F is computed as follows:

$$F = \sqrt{\sum_{i=1}^L (M_i - M_0)^t (M_i - M_0)} \quad (4)$$

The algorithm proposed for feature extraction using BFO. There are certain variations in BFO algorithm used in this work. Firstly, step 6.6 of the proposed algorithm moves the bacteria back to its previous position if current position is less suitable (checked using fitness function). So in this algorithm, bacteria have “memory” as they remember their previous position. Secondly, as there are chances that bacteria may get stuck in local optima, elimination dispersal removes bacteria from its current position and moves it to “random” new position. In the proposed algorithm, position of bacteria is decided randomly in the each iteration. There is no need of using Elimination Dispersal.

2.2. Kernelized Fuzzy C-Means Algorithm

The well-known Cover's theorem states that if a dataset is not linearly separable; transform it into a higher dimensional space non-linearly. Thus, the newly obtained dataset is more likely to be linearly separable. Here, the non-linear transformation of the dataset into a higher dimensional space is performed with some non-linear kernel functions. The concept of “kernel” is first attracted great attention with the introduction of support vector machines (SVM) by Cortes and Vapnik [6]. This idea is then utilized by Zhang and Chen [7] in fuzzy c-means clustering algorithm and they proposed the kernelized fuzzy c-means algorithm (KFCM). The proposed algorithm is realized by restating the distance function in FCM algorithm with a kernel-based function. The objective function of the proposed algorithm then can be given as follows:

$$J_k = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|\phi(x_j) - \phi(c_i)\|^2 = 2 \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m (1 - k(x_j, c_i)) \quad (5)$$

Where ϕ is the mapping function. In [7] used a Gaussian kernel function for non-linear mapping of the dataset. A Gaussian kernel function can be given with the following formula:

$$k(x, y) = \exp\left(-\frac{d(x, y)^2}{\sigma^2}\right) \quad (6)$$

Then by substituting Eq.(6) into Eq.(5)

$$J_k = 2 \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \left(1 - \exp\left(-\frac{d(x, y)^2}{\sigma^2}\right)\right) \quad (7)$$

By setting the derivative of the objective function J_k to zero, u_{ij} and c_i can be found as follows:

$$u_{ij} = \frac{(1 - k(x_j, c_i))^{-\frac{1}{(m-1)}}}{\sum_{i=1}^c (1 - k(x_j, c_i))^{-\frac{1}{(m-1)}}} \quad c_i = \frac{\sum_{j=1}^n u_{ij}^m k(x_j, c_i) x_j}{\sum_{j=1}^n u_{ij}^m k(x_j, c_i)} \quad (8)$$

Thus, by applying KFCM clustering on HRV, obtained clusters with inactive state and active state result.

2.3. Heart Rate Data classification Using IELM

Extreme learning machine (ELM) is an efficient algorithm for single-hidden layer feedforward neural networks (SLFNs), which can produce good generalization performance in most cases and learn thousands of times faster than conventional popular algorithms. However, the performance of ELM is sensitive to the initialized number of hidden neurons. In some traditional methods, the number of hidden nodes is gradually increased by a fixed interval to select the nearly optimal number of nodes for ELM, whereas these methods are of a little bit of complexity and quite time-consuming. This work proposes an improved ELM based on KFCM clustering, which does not need to define the number of hidden nodes in advance manually and randomly. The proposed algorithm automatically determines the number of hidden nodes for different data sets. Empirical study of KFCM-based ELM on several commonly used classification benchmark problems shows that it achieves better performance compared with the standard ELM.

The main idea of the proposed method lies in determining the number of hidden neurons in the SLFNs. In ELM [8] algorithm, the number of hidden neurons is randomly chosen in the beginning of learning, resulting in uncertain network architecture. Usually is not known that how appropriate it is to randomly determine the number of hidden nodes for different data sets. In most cases, the number of hidden nodes is gradually increased by fixed interval and the nearly optimal number of nodes for ELM is then selected based on cross-validation method, which, however, is quite time-consuming in the experiment. Huang et al. [9] proposed an incremental algorithm referred to as incremental extreme learning machine (I-ELM) by increasing hidden nodes one by one until the number of hidden nodes has exceeded the predefined maximum number or the residual error becomes less than the expected one. Subsequently, a convex incremental extreme learning machine (CI-ELM) was proposed in Huang et al. [10]. Different from I-ELM, CI-ELM recalculates the output weights of the existing hidden node after a new hidden node is added. It obtains a faster convergence rate while remaining the I-ELM's efficiency. In order to decrease the network complexity and obtain more compact network architecture, Huang et al. [11] proposed an enhanced method for I-ELM (called EI-ELM). At each learning step, several hidden nodes are randomly generated and among them the hidden node leading to the largest residual error decreasing will only be added to the existing network.

Feng et al. [12] later proposed an approach referred to as error minimized extreme learning machine (EM-ELM), which adds random hidden nodes to SLFNs one by one or group by group and incrementally updates the output weights during the growth of the network. However, there exists more or less complexity in the algorithms above, all of which gradually increase the number of hidden nodes during the data processing. In order to reduce the complexity of determining the number of hidden neurons, a KFCM-based ELM algorithm is proposed in this section in view of the advantages of AP clustering. Firstly, KFCM used to cluster the given samples. Secondly, the number of hidden nodes in the SLFN is determined by the number of clusters after the KFCM clustering. Summarizing, the proposed KFCM-based ELM algorithm carries out the following processing:

KFCM-based ELM Algorithm: Given a training set N and activation function $g(x) = \frac{1}{(1+\exp(-x))}$

Step 1: Separate the data sets into two non-overlapping subsets including equal number of samples for training and testing.

Step 2: All the inputs (attributes) in the training subset are normalized into the range [0, 1].

Step 3: KFCM clustering. Calculate the number of clusters of the training subset and determine the hidden node size in SLFN according to it.

Step 4: Randomly assign hidden node parameters: input weight vector a_i and hidden node bias b_i .

Step 5: Calculate the hidden layer output matrix H using the learning subset.

Step 6: Calculate the output weight matrix β .

3. EXPERIMENTAL RESULTS AND DISCUSSION

The MIT-BIH arrhythmia database is used as the data source in this study which is developed by Mark & Moody (1997) [Available: <http://ecg.mit.edu/dbinfo.html>]. The database contains 48 recordings. Each has duration of 30 minutes and includes two leads; the modified limb lead II and one of the modified leads V1, V2, V4 or V5. The sampling frequency is 360 Hz, the data are bandpass filtered at 0.1–100 Hz and the resolution is 200 samples per mV. Twenty-three of the recordings are intended to serve as a representative sample of routine clinical recordings and 25 recordings contain complex ventricular, junctional, and supraventricular arrhythmias. There are over 109,000 labelled ventricular beats from 15 different heartbeat types. There is a large difference in the number of examples in each heart beat type. The

largest class is —Normal beatl with about 75,000 examples and the smallest class is —Supraventricular premature beatl (SP) with just two examples. The database is indexed both in timing information and beat classification. For more details about MITBIH Arrhythmia database see [14].

A total of 12 records are used from the database and a total of 30,873 beats; 22,476 normal beats, 5,394 abnormal PVC arrhythmia beats, and 3,003 other arrhythmic beats are extracted. The database index files used from database to locate beats in ECG signals. For example the proposed work classifies with these five classes such as normal sinus rhythm (NSR), premature ventricular contraction (PVC), atrial fibrillation (AF), ventricular fibrillation (VF) and and 2° heart block (BII). If for example a segment of HRV with the VF arrhythmia is classified as the VF, then it is said that the segment is classified TP. On the other hand if a non-VF segment is classified as non-VF, then it is said that the segment is classified TN. Any non-VF segment which is classified a VF segment by mistake will produce a FP, while any VF segment which is classified a non-VF segment by mistake will produce a FN result. For the evaluation of proposed classifier, a total of 1317 segments, which are obtained on the MIT-BIH arrhythmia database, were used and it consisted of 835 NSR segments, 57 PVC segments, 322 AF segments, 78 VF segments and 25 BII segments. The implementation was experimented on a variety of datasets

a. Accuracy Comparison

The proposed IELM-PSO algorithm produces better accuracy rate which is much greater accuracy results than existing algorithms such as SVM, ESVM-PSO. When the number of features increases the accuracy of the result is increases. This approach produces high accuracy rate when compared to existing system.

b. F1- measure Comparison

F1- measure is defined as the harmonic mean of precision and recall. A good classifier is assumed to have a high F1- measure, which indicates that the classifier performs well with respect to both precision (P) and recall (R). $F1 - measure = \frac{2 \times PR}{P+R}$. The proposed IELM-BFO based FS algorithm produces high F1- measure is higher than the existing algorithms such as SVM, ESVM-PSO. When the number of features increases the F1-measure of the result is increases. This approach produces effective F1-measure rate when compared to existing system.

c. Execution Time Comparison

The best accuracy rate is achieved using the linear and non-linear feature vector and the BFO-based feature selection algorithm using only less number of selected features which is significantly improves the accuracy and reduces the time complexity.

d. Mean Absolute Error

In statistics, the mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by;

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i|$$

As the name suggests, the mean absolute error is an average of the absolute errors $|e_i| = |f_i - y_i|$, where f_i is the prediction and y_i the true value.

Fig.5 shows the No. of selected features for different feature vector dimensions and classification of heart rate data is done using methods such as the IELM-BFO based FS, ESVM-PSO based FS and SVM without feature selection algorithms. The best accuracy rate is achieved using the linear and non-linear feature vector and the BFO-based feature selection algorithm using only less number of selected features and the classification is done using KFCM based IELM is significantly reduces the MAE and reduces the computational complexity.

4. CONCLUSION

In this work, a KFCM-based ELM algorithm is proposed to automatically determine the number of neurons in the hidden layer of SLFN. KFCM clustering method applied in the algorithm provides the relatively reasonable number of hidden nodes before training the given heart rate data. The BFO has been used for the features selection. A performance comparison of the proposed algorithm with the ESVM-PSO and SVM methods has been carried out on heart rate data. The experimental results indicate that the proposed method achieves improved testing accuracy compared to the ESVM-PSO and SVM.

REFERENCES

- [1] Osowski, S., Linh, T.H. (2001). ECG beat recognition using fuzzy hybrid neural network. *IEEE Trans. Biomed. Eng.* 48 (11), 1265–1271.
- [2] Argyro Kampouraki, George Manis, and Christophoros Nikou, Member, IEEE, "Heartbeat Time Series Classification With Support Vector Machines", *IEEE Transactions On Information Technology In Biomedicine*, VOL. 13, NO 4 July 2009
- [3] M. Kamath and E. Fallen, "Power spectral analysis of HRV: A noninvasive signature of cardiac autonomic functions," *Crit. Rev. Biomed. Eng.*, vol. 21, pp. 245–311, 1993.
- [4] R. Silipo, G. Deco, R. Vergassola, and C. Gremigni, "A characterization of HRV's nonlinear hidden dynamics by means of Markov models," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 8, pp. 978–986, Aug. 1999.
- [5] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, 2000
- [6] C. Cortes, V. Vapnik, *Machine Learning* 20 (1995) 273–293.
- [7] Dao-Qiang Zhang, Song-Can Chen, A novel kernelized fuzzy C-means algorithm with application in medical image segmentation, *Artificial Intelligence in Medicine* 32 (September (1)) (2004) 37–50,
- [8] Karpagachelvi, S., Arthanari, M., & Sivakumar, M. (2011). Classification of ECG signals using extreme Learning Machine. *computer and information science*, 4(1), p42.
- [9] Guang-Bin Huang, Lei Chen, Chee-Kheong Siew, "Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes", *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879-892, 2006.
- [10] Guang-Bin Huang, Lei Chen, "Convex Incremental Extreme Learning Machine", *Neurocomputing*, vol. 70, no. 16-18, pp. 3056-3062, 2007.
- [11] Guang-Bin Huang, Lei Chen, "Enhanced Random Search Based Incremental Extreme Learning Machine", *Neurocomputing*, vol. 71, no. 16-18, pp.3460-3468, 2008.
- [12] Guo-Rui Feng, Guang-Bin Huang, Qing-Ping Lin, Robert Gay. "Error Minimized Extreme Learning Machine with Growth of Hidden Nodes and Incremental Learning", *IEEE Transactions on Neural Networks*, vol. 20, no. 8, pp. 1352-1357, 2009.
- [13] G. B. Moody, and R. G. Mark, The impact of the MIT/BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May- Jun. 2001.