Using an Active Learning Semi-supervision Algorithm for Classifying of ECG Signals and Diagnosing Heart Diseases

Javad Kebriaee, Hadi Chahkandi Nejad and Sadegh Seynali

Abstract Diagnosis of various heart defects and arrhythmias based on the ECG signals recorded from the patient has greatly appealed to the medical community. Biological signal processing performed by experts in the field has involved many challenges to be able to present a precise model of the recorded signals and to analyze and diagnose defects and arrhythmias based on the extracted features and to classify them into the normal and abnormal classes. It is an issue that has appealed to researcher for years to make the process of precisely diagnosing heart diseases intelligent. An efficient classification method with active and semi-supervised learning for classification of the ECG signal based on the mRMR feature selection method has been used in this research. The extracted features include the temporal features, AR, and wavelet coefficients. Finally, the indicators of validity, precision, and sensitivity for this set of selected features have also been evaluated through application of the proposed classifier. The results of simulations in the Matlab software environment suggest that the proposed system has 98.64% validity for diagnosis of 6 class types of ECG. Comparison between the obtained precision and that of the previous research demonstrates the proper performance of the proposed method.

Keywords Classification · ECG signals · Active learning · Semi-supervised learning

---

J. Kebriaee · S. Seynali
Computer Engineering Department, Islamic Azad University,
Birjand Branch, Birjand, Iran

H. Chahkandi Nejad (✉)
Electrical Engineering Department, Islamic Azad University,
Birjand Branch, Birjand, Iran
e-mail: Hchahkandin@iaubir.ac.ir

S. Montaser Kouhsari (ed.), *Fundamental Research in Electrical Engineering*, Lecture Notes in Electrical Engineering 480,
https://doi.org/10.1007/978-981-10-8672-4_6

1 Introduction

Diagnosis of different types of cardiac anomaly and arrhythmia based on the ECG signals recorded from the patient has greatly appealed to the medical community. Experts in the field have been confronted with many challenges processing biological signals, so that they can present an accurate model from the recorded signals that can be analyzed making it possible to diagnose anomalies and arrhythmias and to classify them as normal or abnormal based on the extracted features. Making the process of accurate diagnosis of heart diseases intelligent is an issue that has appealed to researchers for years. In this research, an efficient classification method with active semi-supervised learning has been used for classification of the ECG signal based on the mRMR feature selection methods. The extracted features include temporal features, AR, and wavelet coefficients. Finally, the correctness, accuracy, and sensitivity indices are evaluated for this set of selected features through application of the proposed classification. The results of simulation in the MATLAB software environment suggest that the proposed system has an accuracy of 98.64% for diagnosis of 6 ECG class types, which demonstrates the desirable efficiency of the proposed method as compared to the accuracy obtained in previous research.

Diagnosis of different types of cardiac anomaly and arrhythmia based on the ECG signals recorded from the patient has greatly appealed to the medical community. Experts in the field have been confronted with many challenges processing biological signals, so that they can present an accurate model from the recorded signals that can be analyzed making it possible to diagnose anomalies and arrhythmias, to classify them as normal or abnormal, to diagnose the type of arrhythmia for presentation of the source of anomaly, and to diagnose ventricular tachyarrhythmia, atrial fibrillation, and congestive heart failure based on the extracted features. Many methods and algorithms have been presented so far for analysis of ECG biological signals in the field of frequency, which have classified the cardiac signals of a normal individual and those of one with a specific arrhythmia based on the frequency ranges of the signals. Analysis in the field of frequency is affected to a large extent by signal noises resulting from movement of the patient, the electric equipment noise of the device, etc., and it is not possible to predict and analyze beyond the time of recording and for future. Furthermore, it is impossible in the field of frequency to make a distinction between arrhythmia and noises, and primary diagnosis of CHF is impossible. For this reason, there has been concentration to some extent in the few recent years on analysis of these signals in different fields.

In this research, we will detail the different steps of implementing the proposed method, and will also investigate the effects of changing the parameters effective on implementation of the method. The method has been implemented in the MATLAB software environment. The steps of signal preprocessing will first be described, and the steps concerning extraction and selection of features will then be addressed, followed finally by data classification, for classification of the normal signal and the 5 arrhythmias [1].

2 History

It has been debated for years by researchers in all countries to make the process of diagnosing heart diseases intelligent. The process consists of steps during which the ECG signal is selected as the input to the software, and the software is expected to diagnose well-being or disease and even the type of heart disease with an acceptable accuracy. All these pieces of software extract and select the appropriate features of a signal after receiving it, and then diagnose the type of disease. We will examine different methods used in previous research below.

2.1 Classification of the ECG Signal Using a Wavelet, Morphological Properties, and Neural Networks

In this research, 15 temporal features and 15 features of transform of the selected wavelet have been used after the preprocessing and the PCA method for reduction of the feature sizes, which has resulted in selection of 8 of the best features in each class. Classification is made by a combination of the multilayer perceptron neural network and the radial basis neural network. It has been demonstrated in this research that the hybrid structure of the neural network obtains much better results than the multilayer perception (MLP) neural network [2].

2.2 Classification of Cardiac Arrhythmias Using SVM

In this research, the features of the ECG signal have been extracted through its analysis with a combination of wavelet transform and the AR model. The common methods of cardiac disease diagnosis have been optimized with such an integration. Then, a support vector machine classifier with a Gaussian kernel has been used for automatic classification of five types of cardiac arrhythmia.

2.3 Electrocardiogram Signal Processing

Biological signals are processed in 4 steps.

Measuring or recording the signal: Transformers are used for recording and collecting signals from the body.

Transforming the signal: This step is referred to as preprocessing. The purpose is to reduce signal noise and data size, so that the signal features are easier to extract in the third step. Calculating the signal parameters: The step consists of extraction of the appropriate meaningful parameters (signal features) [3]. Interpreting or classifying the signals: The physician or computer introduces the final interpretation using the extracted features.

3 Pattern Recognition

One of the important purposes of recording and processing critical signals is to interpret them and to employ the useful information in them in diagnosis and treatment. The interpretation step occurs in the recognition or classification phase. For instance, it must be specified after the ECG signal is recorded and preprocessed whether or not it concerns an individual with a particular heart disease. Signal classification actually responds to this question.

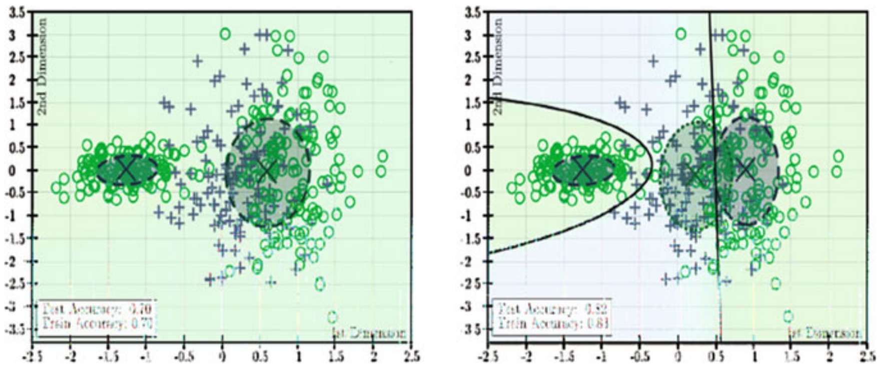
3.1 Pattern Recognition Methods

The methods for pattern recognition fall in general into three groups. Statistical methods: Statistical models are generated for patterns and classes, and classification is made using the concept of probability distribution. Structural (analytical): The pattern classes are specified by figurative structures. These methods are applied most often in cases where the patterns have specific structures [4]. Intelligent networks: Artificial neural networks are networks of units that model the brain neurons. Classification is made using these structures [5].

4 Proposed Method

Classifiers are sometimes based on generating models, since they model observed data (samples), can be trained using partially labeled datasets, and make it possible to integrate expert knowledge easily. However, it is important to identify overlapping processes belonging to different classes and incomplete data distribution matching in order to obtain great classification performance. That is why we use two different methods of training technique for the CMM probabilistic modeling approach (classifier based on probabilistic hybrid models). The first method that we use, titled shared component classifier (CMMsha), determines structural information in an unsupervised way in the first step with a density model of shared components, and is extended to a classifier using class labels [6].

The second classifier, titled separate component classifier (CMMsep), uses class information at present for training, first for construction of the density model of separate classifiers and then for allocation of the components to classes.



(a) Shared-components classifier CMM_{sha}

(b) Separate-components classifier CMM_{sep}

Fig. 1 Comparison of the two types of CMM over an artificial dataset

In Fig. 1, the differences between the two approaches of CMMsha on the left and CMMsep on the right are shown for a two-dimensional continuous (real-value) input space, where we use bivariate Gaussians (normal distributions, in other words) as model components. Each component should model one class of samples in the input space, which can be assumed to result from a process in the actual observed environment. We have 3 such processes that belong to two classes, where the circles are red, and the crosses are blue. The two processes, one blue and one red, overlap each other to a large extent. The center (average) of the Gaussian component describes its location, covariance matrix, and shape. The curves in Fig. 1 are Gaussian curvatures located at the centers represented by large crosses. The decision boundary is shown as a thick black line. In the first step of modeling, CMMsha on the left does not use label information, and is therefore incapable of differentiating between overlapping processes or classes. CMMsha recognizes only two classes, and places both in the green class, and incapability of realizing a high classification rate occurs subsequently. CMMsep on the right, which uses class information in the first step of modeling, is capable of recognizing up to 3 classes, modeling them correctly, and providing us with higher classification accuracy.

4.1 A General View of the Active Learning Approach Using Directed Learning

The standard PAL learning cycle (without our novel directed learning process, in other words) is represented by thick black arrows in Fig. 2a. PAL usually begins with a large repository U of unlabeled samples (the gray box on the left) and a small set of labeled samples L (the gray box on the right), where $X = U \cup L$ and $|L| \ll |U|$. A classifier G is trained based on L. Then, a question-and-answer set of unlabeled samples is specified for labeling based on selection strategy Q, which

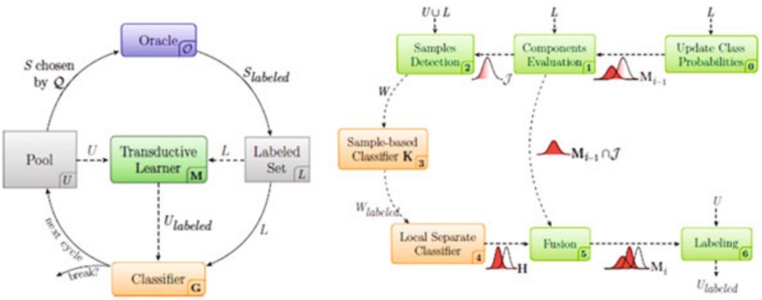


Fig. 2 Graphic description of the active learning process

considers the knowledge contained in G for sample selection, and is presented to predictor O. The labeled samples are added to L, and classifier G is updated. If a termination condition is met, PAL is stopped; otherwise, a new round of questions and answers begins (learning cycle, in other words). Our extension of the standard PAL cycle is shown using the additional dotted arrows in part a of Fig. 2 [7].

4.2 *Our Approach Differs from the Standard Approach from the Following Aspects*

The initial set L is empty.

A directed trainer has been used (part a of Fig. 2, the green box in the middle), which adopts a generating model M in each cycle, and uses M for labeling all the samples in U with a semi-supervised approach.

Classifier G is trained with the labels for all the samples in the dataset.

(a) PAL learning cycle (full arrows) with extensions (dotted arrows).
(b) Directed learning process, extending the standard PAL cycle, concerning the green directed learner box in part a of Fig. 2.

5 Simulation

The method has been implemented in the MATLAB software environment. The steps of signal preprocessing will first be described, and the steps concerning extraction and selection of features will then be addressed, followed finally by data classification, for classification of the normal signal and the 5 arrhythmias.

The ECG signals concerning different patients have been provided by the MIT-BIH standard database. The database has 48 two-channel ECG signals

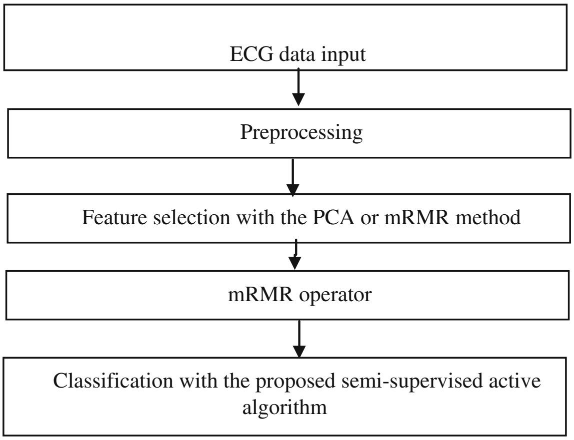


Fig. 3 Flowchart for the order in which the algorithms and techniques applied in the proposed method are run

obtained from 47 case studies at the BIH arrhythmia laboratory between 1975 and 1979. The signals have been stored with a frequency of 360 samples per second and an accuracy of 12 bits in a range of 10 mV, and 20 patients' signals were selected as input data (Fig. 3).

5.1 Signal Shift to the Baseline Deviation

The noise resulting from breathing when the electrocardiography signal is recorded has a low frequency of about 15 Hz. The noise causes the baseline of the electrocardiography signal to change, as a result of which extraction of the temporal properties and features of the signal is confronted with problems. The level-8 wavelet of the ECG signal has a higher amplitude than at the previous wavelet levels, and the ECG signal is corrected if it exhibits a considerable difference from the baseline at some pulses (Figs. 4 and 6) following the removal of the level, so that a uniform, regular signal is achieved.

5.2 Removal of the Noise Resulting from Mains Electricity

The ECG signal at this level includes noise resulting from mains electricity. Through application of a band-pass filter that does not pass signals in the range of 60 Hz, the noise resulting from mains electricity, which has a higher frequency than that of the main signal, can be filtered (Figs. 5 and 6) [8].

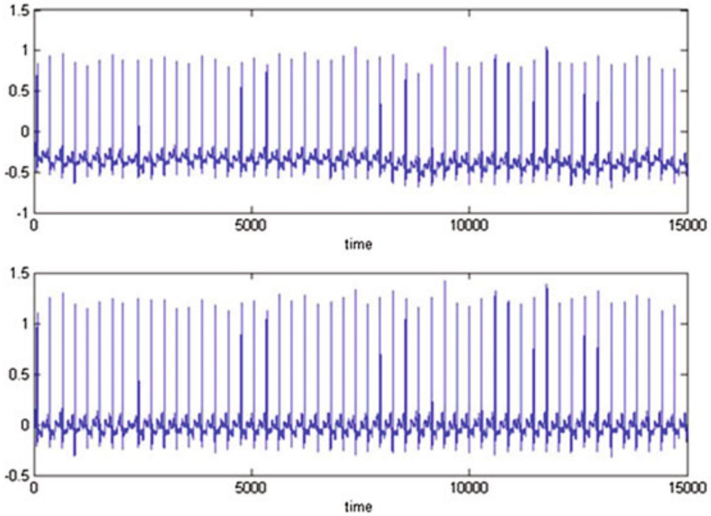
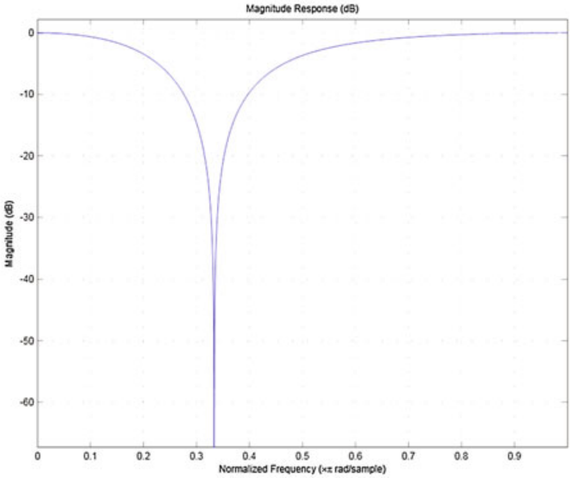


Fig. 4 50 heartbeats where the baseline deviation noise has been removed

Fig. 5 Removal of the 60-Hz frequency with a notch filter



5.3 *Signal Windowing*

Windowing is carried out using the information extracted from the software environment, such as the moment when wave R has occurred and the type of disease diagnosed by the physician for each pulse. For this purpose, 100 samples before and 200 after the moment when wave R has occurred of the smoothed signal is considered as a complete pulse (Fig. 6), and the pulse is placed in the class concerning its disease based on the physician's diagnosis on it.

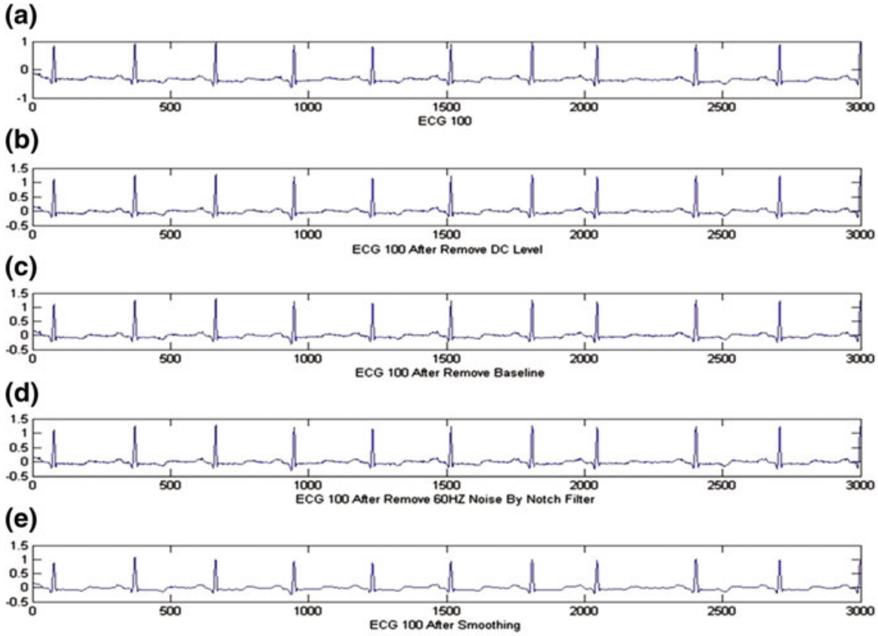


Fig. 6 **a** 100 ECG signal, **b** ECG signal after removal of the DC value, **c** ECG signal after removal of the baseline deviation noise, **d** ECG signal after removal of mains electricity, **e** ECG signal after smoothing

5.4 Selection of the Training and Test Data

At the final step of signal preprocessing, 1060 pulses are fully randomly selected from each of the six classes, out of which 750 pulses are considered as training data and 310 as test data, and are stored in a separate classification (Given the low amount of data available in the database on the disease A signal, 224 pulses have been considered as training data and 97 as test data) [9].

5.5 Feature Extraction

Temporal features, wavelet features, and AR features were selected for the feature extraction step, which were extracted for the training and test data. Each extracted feature vector consists of a total of 64 features, including 9 temporal features, 48 wavelet features, 2 AR features, and 5 PCA features.

Feature selection with PCA.

When the PCA operator is applied, the features are mapped in a new space, and sorted based on importance. The superior features are selected, and stored for

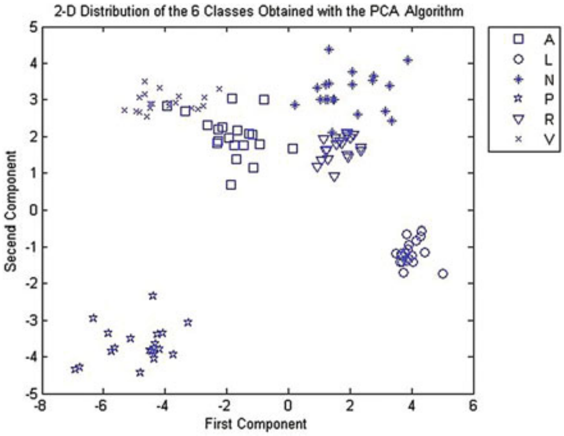


Fig. 7 Distribution of 20 feature vectors from each class after PCA mapping shown in two dimensions

Table 1 Accuracy of classification using 5 classification methods and 40 selected features

Method	A	L	N	P	R	V	Accuracy
PCA.SVM	93.67	81.44	82.58	85.66	84.02	80.30	84.61
mRMR.SVM	94.05	82.55	82.90	86.12	80.26	83.71	84.93
PCA.Porposed	94.11	99.40	98.91	99.64	99.70	99.03	98.47
mRMR.Proposed	94.11	99.05	99.54	99.33	100	99.82	98.64

classification in the following steps. The distribution of 20 feature vectors from each class after PCA mapping is shown in two dimensions in Fig. 7. Table 1 shows the output of the classifier.

Identification of the important points of the signal using PCA.

We consider all the points of the signal as the input to PCA. Given the PCA mapping, the mapped points are ranked based on importance. We select 5 superior mapped points as features to be used in the following steps [10].

6 Principal Component Analysis (PCA)

In the principal component analysis method, new coordinate axes are defined for the data, such that the first axis lies in the direction that maximizes data variance, the second axis is considered perpendicular to the first axis in the direction that maximizes data variance, and the following axes similarly lie perpendicular to all the previous axes to maximize data variance in that direction. Principal component analysis is one of the popular methods of feature extraction, used in many studies due to simplicity and high processing speed. The PCA technique is the best method for linear reduction of data dimensions; that is, less information is lost than in other methods, since the less important coefficients obtained from the transformation are eliminated. Assume that input matrix Xhas N_T samples and n features, and the N_T samples should be placed in C groups. The data mean and covariance are calculated based on the following equations.

Equation (1):

$$\begin{aligned} m_d &= \frac{1}{N_T}\sum_{i=1}^{c}\sum_{j=1}^{N_i} x_{i,j} \\ COV &= \frac{1}{N_T}\sum_{i=1}^{c}\sum_{j=1}^{N_i} \left(x_{i,j} - m_d\right)\left(x_{i,j} - m_d\right)^T \end{aligned} \tag{1}$$

In the next step, the eigenvalues and eigenvectors are calculated based on the covariance matrix. Then, k greater eigenvalues are selected of the total n. Now, input matrix X is transformed under eigenvector matrix P with k features to the principal component analysis space.

Equation (2):

$$Y_{ij} = [P_1, P_2, \ldots, P_k]^T X_{ij} \tag{2}$$

6.1 Feature Selection with mRMR

Through integration of the features and its application to mRMR, the dimensions are sorted in terms of importance. In mRMR, the feature dimensions are sorted in all the classes through selection of the best with the criteria of maximum dependency and minimum redundancy, and no mapping is carried out there.

6.2 Classification Using the Proposed Method

Learning and test: First, all the ECG segments including a particular type of heartbeat are mapped onto the feature space using the wavelet, AR, and temporal features. In the learning phase, the proposed network receives a few samples as input. These patterns are heartbeats shown by m feature parameters that can be seen as points in an m-dimensional space. The network is then capable of obtaining the labels of the new vectors by comparing the samples used in the learning phase. The network receives the training and test data, and carries out the classification. Finally, the network output is compared to the test data, and the error rate of the classifier and classification percentage are obtained by counting the pulses classified incorrectly.

In classifiers where mRMR is used as feature selection, this method of feature extraction achieves the best classification accuracy with the lowest number of selected features.

$$\begin{aligned} ERROR &= \frac{Number\ of\ False\ Classified\ Beats}{Number\ Beats} \\ Accuracy &= (1 - ERROR) \times 100 \end{aligned} \tag{3}$$

Table 1 shows the accuracy of ECG signal classification using 5 classification methods. The classification has been carried out through selection of 40 features and 750 pieces of training data. In this table, the proposed method and the support vector machine method have been compared to each other for two different instances of feature extraction.

Figure 8 shows the training procedure of the proposed network in 21 iterations.

A change in the number of selected features causes a change in classification accuracy. Figure 9 shows classification accuracy for changes in the number of selected features with 750 training data, and the proposed network selects the best solution from among the features.

In this paper, we described the steps in simulation of 5 different methods of ECG signal classification. Classification was divided into the 4 major steps of preprocessing, feature extraction, feature selection, and classification. Among the implemented methods, use of the mRMR. Proposed classification exhibits acceptable accuracy as compared to the other common methods. Table 2 shows the other comparable indicators in regard to this research, such as sensitivity.

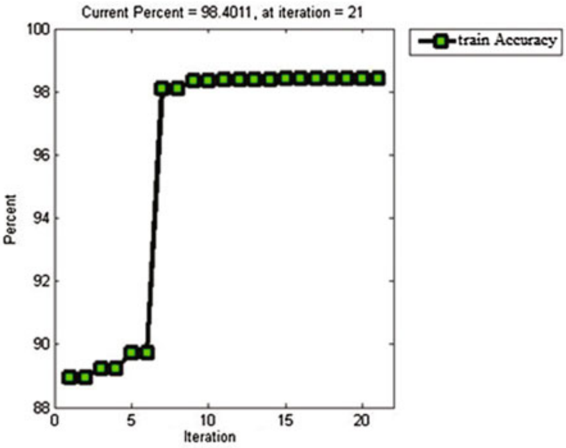


Fig. 8 Selection optimization procedure for the proposed semi-supervised network in 21 iterations of the algorithm

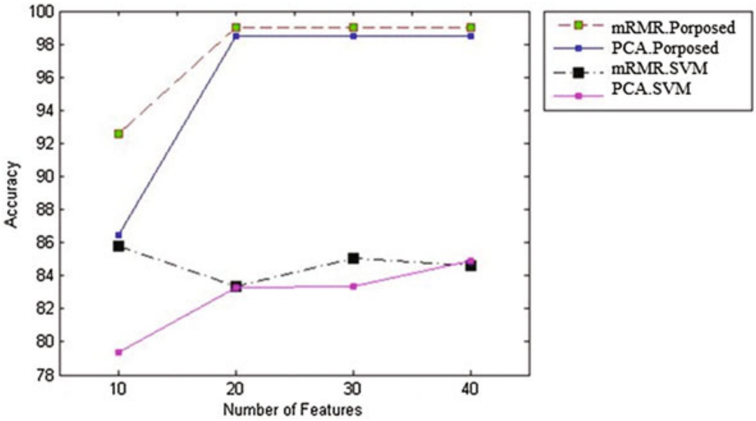


Fig. 9 Accuracy of classification using 4 classification methods for 10, 20, 30, and 40 dimensions of the feature vector

Table 2 Statistical results of implementing the proposed network and support vector machine for two feature selection methods

Feature(s)-network	Statistical parameters		
	Sensitivity (%)	Specificity (%)	Accuracy (%)
PCA-SVM	84.32	85.53	84.61
mRMR-SVM	85.32	86.63	84.93
PCA-Proposed	96.71	96.82	98.47
mRMR-Proposed	97.32	97.73	98.64

7 Conclusion

In this research, the classification of 5 cardiac arrhythmias and the normal signal selected from the MIT-BIH database were investigated. For each arrhythmia, 70% was used for training the system and the remaining 30% for testing the system. As shown, use of mRMR for optimal feature selection and of the proposed classifier network is an appropriate approach for achieving 2 purposes:

(1) increasing classification accuracy
(2) selecting fewer features (superior features).

Selection of fewer features will reduce computation, while higher accuracy was observed where fewer features had been selected than in the case where all the features had been applied to the classifier. Furthermore, selection of wavelet features and features extracted from the signal mapped by PCA has an important role in improvement of the precision of classification, in such a way that of the total of 6 superior features, 1 is a wavelet feature, and 2 are features extracted from the PCA-mapped signal. As compared to previous research with the same patient ECG signals as in this one, it is observed that application of the proposed method to these data obtains favorable results. Thus, an accuracy of 84.93 has been obtained with 750 pieces of training data and the mRMR.SVM method, while implementation of the mRMR. Proposed method with 750 pieces of training data shows an accuracy of 98.64.

8 Features Works

Finally, we make suggestions for achievement of better diagnosis of heart disease based on the results obtained from analysis of the collected dataset.

(1) Addition of further independent features can reduce algorithm error.
(2) Use of combinations of various classifiers can improve the rate of ECG signal recognition.
(3) Use of two-dimensional features using a Hankel matrix.

References

1. Ghorbanian P, Ghaffari A, Jalali A, Nataraj C (2010) Heart arrhythmia detection using continuous wavelet transform and principal component analysis with neural network classifier. In: Computing in cardiology IEEE conference, pp 669–672
2. Khazaee A (2013) Heart beat classification using particle swarm optimization. I.J. Intell Syst Appl 5(6):25–33
3. Rodrigues D, Pereira LAM, Almeida TNS, Papa JP, Souza AN, Ramos CCO, Yang X-S (2013) BCS: a binary cuckoo search algorithm for feature selection, circuits and systems (ISCAS), 2013 IEEE international symposium, pp 465–468
4. Reitmaier T, Calma A, Sick B (2015) Transductive active learning–a new semi-supervised learning approach based on iteratively refined generative models to capture structure in data. Inf Sci 293:275–298
5. Ubeyli ED (2010) Lyapunov exponents/probabilistic neural networks for analysis of EEG signals. Expert Syst Appl 37:985–992
6. Ubeyli ED (2010) Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals. Expert Syst Appl 37:1192–1199
7. Guler NF, Ubeyli ED, Guler I (2005) Recurrent neural networks employing Lyapunov exponents for EEG signals classification. Expert Syst Appl 29:506–514
8. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
9. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley, Chichester
10. Fisch D, Sick B (2009) Training of radial basis function classifiers with resilient propagation and variational Bayesian inference. In: International joint conference on neural networks (IJCNN '09), Atlanta, pp 838–847