Congestive heart failure detection using random forest classifier

Zerina Masetic[a], Abdulhamit Subasi[b,*]

[a] International Burch University, Faculty of Engineering and Information Technologies, 71000 Sarajevo, Bosnia and Herzegovina
[b] Effat University, College of Engineering, Computer Science Department, Jeddah 21478, Saudi Arabia

ARTICLE INFO

ABSTRACT

Background and objectives: Automatic electrocardiogram (ECG) heartbeat classification is substantial for diagnosing heart failure. The aim of this paper is to evaluate the effect of machine learning methods in creating the model which classifies normal and congestive heart failure (CHF) on the long-term ECG time series.

Methods: The study was performed in two phases: feature extraction and classification phase. In feature extraction phase, autoregressive (AR) Burg method is applied for extracting features. In classification phase, five different classifiers are examined namely, C4.5 decision tree, *k*-nearest neighbor, support vector machine, artificial neural networks and random forest classifier. The ECG signals were acquired from BIDMC Congestive Heart Failure and PTB Diagnostic ECG databases and classified by applying various experiments.

Results: The experimental results are evaluated in several statistical measures (sensitivity, specificity, accuracy, *F*-measure and ROC curve) and showed that the random forest method gives 100% classification accuracy.

Conclusions: Impressive performance of random forest method proves that it plays significant role in detecting congestive heart failure (CHF) and can be valuable in expressing knowledge useful in medicine.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Human heart is the most important and hardest working muscle in human body, that together with blood vessels compose cardiovascular system. It pumps the blood into the every cell of the human body. Furthermore, heart muscle is the engine of the human body [1]. Heart failure is the common syndrome that progresses slowly but causes cardiac dysfunction resulted from the inability of the heart to pump the blood to all the cells of the human body efficiently. The heart weakens by heart attacks, long-term high blood pressure or an anomaly of one of the heart valves. Yet, heart failure is generally not recognized until it comes to the more progressive phase, entitled congestive heart failure, which causes fluid to flow to lungs, feet and abdominal cavity. Congestive heart failure is a condition that can be caused by heart diseases, such as coronary artery disease, damage of the heart after the heart attack, high blood pressure, valvular heart disease, diabetes and even the alcoholism [2]. According to European Heart Network and European Society of Cardiology [3], over 4 million people die from cardiovascular diseases in

---

* Corresponding author. Tel.: +966 590129599.
E-mail addresses: zmasetic@ibu.edu.ba (Z. Masetic), absubasi@effatuniversity.edu.sa (A. Subasi).

Europe and 1.9 million in European Union (EU) which is 47% deaths in Europe and 40% in EU.

Considering that there is no definite diagnosis of heart failure, medical diagnosis such as history or physical examinations, electrocardiography (ECG), chest radiography or echocardiography is crucial for detecting the congestive heart failure. The electrocardiogram (ECG) is noninvasive tool that records electrical activity of the heart and shows irregularities of the heartbeats. It safely examines and records the electrical impulses of the heart and show possible damages of the heart or irregularities of the heartbeats [4]. Thus, ECG is an important tool for determining the function and the health of the cardiovascular system. Moreover, it is significant to define accurate and timely diagnosis of physicians to avoid more damage and to determine proper methods and approaches [5]. Still, the problem occurs when there is insufficient number of physicians to meet the needs of patients. Therefore, it is necessary to develop an effective and automated diagnostic systems based on ECG recordings, combined with application of machine learning techniques for classification of heart diseases. These diagnostic systems will aid medical experts in detecting the irregularities in the cardiovascular system. The diagnostic system will firstly process the ECG signals taken from different subjects and hence decomposed into few features performing feature extraction. Extracted ECG signals are used to detect different types of heart failure by using various machine learning techniques [6].

According to the number of researches done, the field of heartbeats classification using different techniques is very popular. Beth Israel Deaconess Medical Center (BIDMC) congestive heart failure database were used in different studies. Baim et al. [7] were used this database to show the effect of treatment with oral milrinone. The larger group of 100 patients, with severe congestive heart failure belonging to NYHA3 and NYHA4 groups was treated with oral milrinone with an average initial dose of 27 ± 8 mg/day. Causes of congestive heart failure were different (e.g. ischemic heart disease, dilated cardiomyopathy, valve replacement, etc.). Thuraisingham used [8] BIDMC congestive heart failure database in detecting CHF from normal heartbeats, using k-nearest neighbor algorithm and features from the second-order difference plot (SODP) obtained from Holter monitor cardiac RR intervals. Six features are obtained from second-order difference plot and k-nearest neighbor algorithm is applied with the value of $k = 1$. Authors obtained a success rate of 100% in separating CHF and normal heartbeats. The same database was used by Kuntamalla and Reddy [9] and they applied sequential trend analysis to differentiate CHF patients from patients with normal heart beats. The accuracy of the proposed method is 96.68%. Furthermore, Kuntamalla and Reddy [10] applied multiscale entropy (MSE) to HRV signals, to differentiate healthy young and elderly subjects from CHF patients. They applied Reduced Data Dualscale Entropy Analysis method to reduce the data size for clear differentiation of subjects. They achieved 100% accuracy. Hossen and Al-Ghunaimi [11] used technique based on recognition of power spectral densities pattern of decomposed sub-bands of R–R interval to identify patients with congestive heart failure. They used 12 subjects from BIDMC congestive heart failure database

and reached accuracy of 90%. The same authors, Hossen and Al-Ghunaimi [12] used different wavelet decomposition filters with soft decision algorithm to estimate power spectral density of RR interval data for screening patients with congestive heart failure. The accuracy value of test data with almost all wavelet filters used was 88.6%. Yu and Lee [13] focused on selecting the best feature selector to get high accuracy in recognizing the CHF. The authors proposed conditional mutual information feature selector (CMIFS) with support vector machine (SVM) classifier. The combination achieved 97.59% accuracy using only 15 features. Işler and Kuntalp [14] combined classical HRV and wavelet entropy measures to distinguish healthy patients from patients with congestive heart failure. They applied genetic algorithm to select the best ones among all possible combinations of measures. Furthermore, they used k-nearest neighbor classifier with different values of k. Finally, they got accuracy value of 96.39% with value of $k = 5$ and 7, using 8–10 features. Similar study was done by Asyali [15] and author used linear discriminant analysis to examine the discrimination power of 9 long-term HRV measures and validated the result by applying Bayesian classifier. Sensitivity and specificity value of the stated classifier are 81.8% and 98.1%, respectively. Pecchia et al. [16] investigated discrimination power of heart rate variability (HRV) in distinguishing normal subject and subjects with congestive heart failure (CHF). They performed time and frequency analysis in order to measure HRV features. The result is evaluated by applying classification and regression tree (CART) classifier. Furthermore, authors introduced two non-standard features: average of normal intervals and low/high frequencies for recording over the 24 h and reached the success rate of 89.7% for sensitivity and 100% for specificity.

Feature extraction methods are important in ECG signal classification. In this paper, features are extracted by use of autoregressive Burg method. Consequently, specified signal is categorized as normal or with congestive heart failure by using machine learning techniques. In previous works [8,9,11–15,17], various machine learning techniques were applied. However, k-NN method is not capable of dealing with high dimensional data without dimension reduction, where SVM is not strong enough to handle large number of trivial data without data selection [18]. Decision tree method might not perform well with more complex interactions and overfitting might lead to the instability of the model [19]. Method efficient to achieve excellent performance, but not applied in previously mentioned studies [8–15] is random forest (RF). In this paper, RF method is used to investigate the performance of classifier for ECG signals classification in diagnosing the congestive heart failure. Moreover, performance of different machine learning techniques such as C4.5 decision tree, ANN, k-NN and SVM in ECG signal classification is compared (see Fig. 1.).

The rest of the paper is organized as follows: Section 2 gives explanation about the data used and present methods applied in each step of the ECG signal classification process. Experimental results are presented and discussed in Section 3. Finally, conclusions derived from the results are summarized in Section 4.
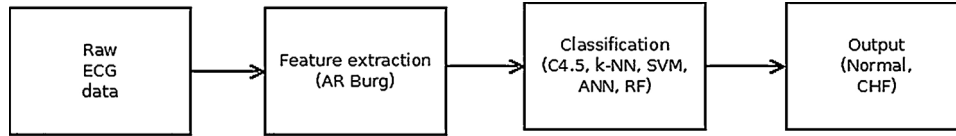
Fig. 1 – Block diagram of the classification system.

2.   Materials and methods

2.1.   Data acquisition

To evaluate the performance of the proposed methods, ECG signals are taken from the group of PhysioNet databases. The paper includes ECG signals from 15 subjects from Beth Israel Deaconess Medical Center (BIDMC) Congestive Heart Failure (CHF) database. The BIDMC CHF database contains recordings of 11 male patients, aged 22–71 and 4 female patients, aged 54–63 with severe congestive heart failure, belonging to New York Heart Association (NYHA) class 3 and 4. This group of subjects is part of larger study group, treated with the oral inotropic agent, milrinone. Duration of each recording, containing two ECG signals is about 20 h, sampled at 250 samples per second in a range of ±10 mV, with 12-bit resolution. Each ECG recording contains around 17 million sample intervals. The larger study is based on 100 patients with severe congestive heart failure treated with digitalis glycosides, diuretics and one or more oral vasodilators [20].

Second data set include ECG signals from 3 patients suffering from congestive heart failure, obtained from PTB Diagnostic ECG database. First patient is 74-years old female, belonging to NYHA 2, second patient is 70-years old male, belonging to NYHA 3 and third patient is 61-year old female, belonging to NYHA 4. These ECG signals are part of larger data group, collected from healthy volunteers and patients with various heart diseases, at the Department of Cardiology of University Clinic Benjamin Franklin, in Germany. Each ECG record includes 15 simultaneously measured signals: the conventional 12 leads with 3 Frank lead ECGs. Signal form lead I is taken for purpose of this study. Original signals are digitized at 1000 samples per second, with 16 bit resolution over a range of ±16.384 mV [20].

Normal heartbeats are taken from 13 subjects from MIT–BIH Arrhythmia database which totally contains 48 ECG recordings lasting for 30 min, obtained from 47 subjects by the BIH Arrhythmia Laboratory, between 1975 and 1979. These recordings are randomly chosen from a larger set of recordings at Boston's Beth Israel Hospital, collected from a mixed population of inpatients and outpatients. These recordings are taken using recording bandwidth of 0.1–100 Hz. Each of these recordings is sampled at 360 samples per second over the 10 mV range, with 11-bit resolution. This database is freely available on http://physionet.org/physiobank/database/mitdb/ [20].

2.2.   The Burg method for autoregressive (AR) parameter estimation

The ECG signals from both databases are preprocessed and features are extracted using autoregressive (AR) Burg algorithm. The autoregressive model is well-known feature extraction method, especially for biological signals. The important step in AR modeling is selection of the optimal order *p*. A process of model order *p* in autoregressive model is given by following formula [21]:

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + e[n] \tag{1}$$

where $x[n]$ represents data of the signal at point n, p represents model of order, a_k is an autoregressive coefficient and $e[n]$ represents noise error [21]. The Burg method is used for estimating a real valued autoregressive coefficient a_k recursively using a_k of previous order p − 1. It is the form of order-recursive least squares method used to estimate parameters by minimizing the forward and backward errors of the linear system. The Burg method is precise because it uses many data points at the time, minimizing the backward and forward error [21]. However, Burg algorithm involves prediction error powers defined by formula [22]:

$$\delta_k^2 = \delta_{k-1}^2 (1 - |a_k|^2) \tag{2}$$

where δ_k^2 is prediction error power which decreases when model of order p is increasing, producing the stationary model.

There is no straightforward way to determine the correct model order. As we increase the order of the model, the root mean square (RMS) error usually decreases quickly up to some order and then more slowly. An order just after the point at which the RMS error flattens out is usually an appropriate order. Much research has been done to discover good model order criteria. One of the criterion most often used is called the *final prediction error* (FPE). It is related to the variance of the modeling error.

$$\text{FPE} = s_p^2 \frac{N+p+1}{N-p-1}, \tag{3}$$

where s_p^2 is the variance of the modeling error and N is the number of points in the signal. Model of order used in this study is 32. The length of each feature vector extracted using AR Burg is 33.

Each signal is divided into subsections of signals, consisting of 1000 data points (window length of 1000). Each subsection was used to extract features using AR Burg. Figs. 2 and 3 shows input to feature extraction and output data after *AR Burg* has been applied, respectively.

2.3.   C4.5 decision tree algorithm

In decision tree method, instances are classified by arranging them down the tree, from root to leaf node. Each internal
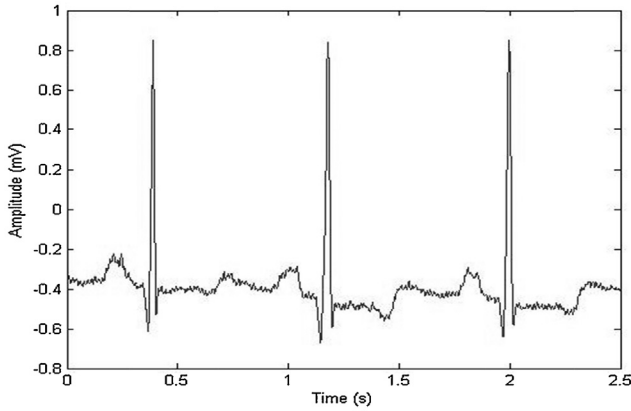
Fig. 2 – Raw ECG signal.

node is a test for some attribute of the tree. The root node is the starting point of the classification process and is without incoming edges. The other nodes have exactly one incoming edge and are called leaves. The classification starts at the root node with testing the attribute specified by this node and continue down the tree branch according to the value of the attribute in the example. When a leaf node is reached, the instance is classified according to the class of the leaf [19,23].

C4.5 algorithm is based on ID3 algorithm, a very simple decision tree algorithm, presented by Quinlan [24]. This algorithm passes through decision tree, visits each node and select optimal split. It is achieved by using the gain ratio, represented by following formula [19]:

$$\text{GainRatio } (S, A) = \frac{\text{InformationGain } (S, A)}{\text{Entropy } (S, A)} \qquad (4)$$

where *entropy* is the term which describes how equally the attribute splits the data [19,23,24] and is calculated by formula:

$$\text{Entropy } (y, S) = \sum_{c_j \in dom(y)} - \frac{\sigma_y = c_j S}{|S|} \cdot \log_2 \frac{\sigma_y = c_j S}{|S|} \qquad (5)$$

Information gain is the impurity-based criterion which uses an entropy measure as the impurity measure, for some

Fig. 3 – ECG data after AR Burg feature extraction is applied.

training set S with respect to the attribute A and is presented by formula [19]:

$$\text{InformationGain } (a_i, S) = \text{Entropy } (y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i} = v_{i,j} S|}{|S|}$$
$$\cdot \text{Entropy } (y, \sigma_{a_i} = v_{i,j} S) \qquad (6)$$

The optimal accuracy is reached when the size of tree was 15 and number of leaves 8.

2.4. k-Nearest neighbor (k-NN) algorithm

Fix and Hodges [25] proposed a classification method which is easy to implement and give high accuracy. k-NN is non-parametric method for pattern classification based on finding the "nearest" training set T. The first step is to reduce the dimension of feature space by using distance function (commonly used Euclidian distance) between a test set and specified training set [26]:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \qquad (7)$$

where a_r is the value of rth attribute of instance x.

A mathematical expression of k-NN for Y is expressed below:

$$Y = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \qquad (8)$$

where Y is a local mean vector and $N_k(x)$ is the neighborhood of x defined by the k closest points x_i in the training set [27]. In k-nearest neighbor, the effective number of parameters is N/k, which is greater than p that represents parameters in least-square fits, and is getting lower value when increasing the value of k. So, there would be N/k neighborhoods and one parameter would fit in each of them, if the neighborhood were not overlapping [28]. Different values of k were applied to the dataset, such as k = 1, 3, 5, 7 and 10, but accuracy value reached the highest number when using 3-nearest neighbor for classification.

2.5. Artificial neural network (ANN) algorithm

Concept of artificial neural network was presented by McCulloch and Pitts in 1943 [29], where artificial neuron is based on biological neuron. Neural network consist of huge number of processing units called neurons, nodes, processing units or cells. Each is connected to another by related weight. Weight is a term which represents synapses in biological neural network. Neural network consists of input nodes which receives data from outside and form input layer, output nodes which send data out of the network and form output layer and hidden node whose signals remain in the network and form hidden layer. In this study, 2800 instances are classified into two classes of output (normal and congestive heart failure). Network is created with one input layer, consisting of 33 inputs
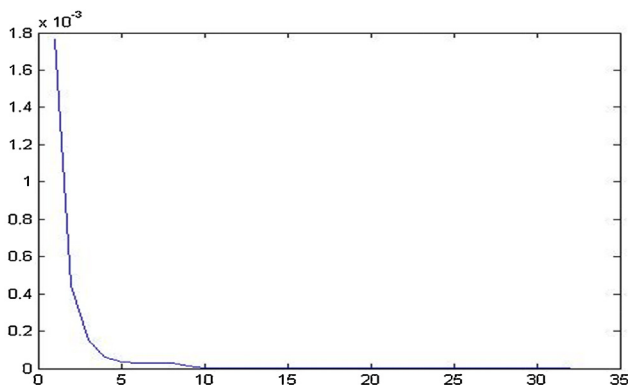
and 10 nodes in the hidden layer, as these parameters give the optimal classification accuracy. Every artificial neuron has internal state, called activation function of the inputs received. Many activation functions tested but sigmoid function gave the best accuracy. A sigmoid (S shaped curves) function presented by formula below [30]:

$$f(x) = \frac{1}{1 + e^{-x}} \qquad (9)$$

and gives only two outputs. Network architecture is the way nodes are organized in layers and interconnected. In this study, feed-forward backpropagation architecture is applied to the dataset [30–32].

Backpropagation neural network architecture is proposed by Rumelhart et al. [32]. It is the classification algorithm that uses gradient descent method to adjust the connection weights in network model. In backpropagation algorithm, output of each neuron will be found by aggregating the neurons of previous level and multiplied by determined weight [33].

2.6. Support vector machines (SVM) algorithm

Support vector machines (SVM) concept is first presented by Boser et al. [34]. They presented a training algorithm that will maximize the margin between training patterns and decision boundaries [28]. SVM is a hyperplane that splits positive and negative sets of examples, with maximal margin. SVMs belong to the category of *kernel methods*, used in high dimensional feature space for computing a dot product. Kernel methods solve the problems of quadratic increase in memory when storing the features and time required calculating the classifier's discriminant function by avoiding mapping the data into the high-dimensional feature space [35]. When considering the dataset mentioned above, consisting of pairs (x_1, y_1), $(x_2, y_2), \ldots, (x_N, y_N)$, where $x_i \in \Re^P$ and $y_i \in \{-1, 1\}$, hyperplane is defined by [28]:

$$\{x : f(x) = x^T \beta + \beta_0 = 0\} \qquad (10)$$

where β is a unit vector: $||\beta||=1$.

A classification rule created by $f(x)$ is

$$G(x) = sign[x^T \beta + \beta_0] \qquad (11)$$

Now, when classes are separable, a function $f(x)$ with $y_i f(x_i) > 0$ for all i values and hyperplane that produces the biggest margin between the training points for classes -1 and 1 can be determined. The optimization problem for this concept is presented by following formula [28]:

$$\max_{\beta, \beta_0, ||\beta||=1} M$$
$$\text{subject to} \quad y_i(x_i^T \beta + \beta_0) \geq M, \quad i = 1, \ldots, N, \qquad (12)$$

So, margin is created and it is M unit away from hyperplane from both sides [28]. In this research, we used Puk kernel for SVM classifier. Linear, nonlinear and RBF kernels were also tested, but computational time increased in nonlinear RBF kernels, besides accuracies were significantly reduced. Hence, we used Puk kernel for SVM classifier which gives

best accuracy among other kernels with $\omega = 1$ and $\sigma = 1$. These values give optimal classification accuracy with reasonable computational time. However, based on literature [36], we tested different values of ω and σ such as 0.5 and 2, 5 and 0.5, respectively. Classification accuracy decreased, whereas computational time increased.

2.7. Random forest (RF) algorithm

Random forest is a classification method which combines multiple tree predictors in that way that each tree depends on a value of randomly chosen vector distributed among all trees in forest in the same way. So, in the random forest algorithm, a random vector θ_k is produced, independent from the previous random vectors and distributed to all trees, and each tree is grown using training set and random vector θ_k, which results in collection of tree-structured classifiers $\{h(x, \theta_k), k = 1, \ldots\}$ at input vector x. In random forest algorithm, generalization error is given by [37]:

$$PE* = P_{X,Y}(mg(X, Y) < 0) \qquad (13)$$

where subscripts X and Y are random vectors that indicate the probability is over the X, Y space and mg is the margin function which measures the extent to which the average number of votes at random vectors for the right output exceeds the average vote for any other output. Margin function is defined as

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \qquad (14)$$

where $I(\cdot)$ is the indicator function [37].

Two parameters that measure accuracy of individual classifier and dependence between classifiers are strength and correlation, respectively. Random forest with random features is formed by choosing a small group of input variables on each node, randomly [37]. In this research, random forest consisted of 20 trees, each constructed while considering 6 random features. Optimal values for trees and random features are obtained by inserting different values of trees and random features, using the classification accuracy as fitting function. Therefore, 20 as the number of trees gave the optimal accuracy results.

3.   Results and discussion

In this study, two automated system are designed to classify normal (N) heart beats and heart beats with congestive heart failure (CHF). In first case, the ECG signals used in the study are taken from BIDMC congestive heart failure and MIT–BIH Arrhythmia databases, both freely available on PhysioNet [20]. A total number of 2800 feature vectors, where 1500 belong to 15 patients with congestive heart failure and 1300 belong to 13 patients that have normal heartbeats are extracted using AR Burg method. Consequently, different machine learning techniques are applied to evaluate the performance of the proposed system. In second case, the same methodology is applied on second data set, composed of 171 congestive heart failure samples from PTB Diagnostic ECG database, combined

with 1300 normal heart beats samples from MIT–BIH Arrhythmia database [20].

3.1. Performance evaluation criteria

In determining the performance of the machine learning techniques, different approaches are used. The common approach is to divide the whole dataset into three subsets: training subset, validation subset and testing subset. All of them should be chosen independently from each other. Training set is then used to train the data, validation set to optimize parameters of classifiers, and finally test data is used to evaluate error rate of optimized classifier. Use of this method is not practical, because of limited number of data for both, training and testing. Moreover, samples in the training or testing subsets may not be representative, because one class should be represented in both subsets in right proportion. [38].

More efficient statistical technique is cross-validation. Dataset is divided into equal number of folds (partitions), where one fold is used for testing and remaining number of folds for training. It will be repeated as many times as number of fold is where each fold is used for testing and each fold is in group of training set. In this study, 10-fold cross validation is applied on the whole dataset, where dataset is divided into 10 folds of (roughly) equal size. The model is trained and tested 10 times; each time is tested on one of k folds and trained using the remaining 9 folds. Then, average cross validation accuracy is found of the k individual accuracies [39]:

$$\text{CVA} = \frac{1}{k}\sum_{i=1}^{k} A_i \tag{15}$$

where k is the number of folds and A_i accuracy of each fold.

When evaluating the performance of the classifiers, two more terms should be introduced: sensitivity and specificity. Sensitivity specifies how good the classifier can recognize positive samples and is defined by [40]:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{16}$$

where TP is number of true positive samples and FN is the number of false negative samples. Sensitivity defines number of people having congestive heart failure and having positive test results. Specificity specifies how good the classifier can recognize negative samples and is defined by [40]:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \tag{17}$$

where TN is number of true negative samples and FP is the number of false positive samples. Specificity defines number of people without heart failure and having negative test results. The accuracy is then calculated by [40]:

$$\text{Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{18}$$

Another statistical index is receiver operating characteristics (ROC) curve, which is evaluation metrics of observer performance, where learner is trying to select the samples

with higher portion of positive values [38]. It is used to show discrimination ability of different statistical methods that combine different results to predict something [41]. ROC curve is created by plotting the number of true positive values on vertical axis (sensitivity) and false positive on the horizontal axis (1-specificity). The ideal point on the ROC curve would be (0, 1), meaning that all positive instances are classified as positive and none of negative instances are misclassified as positive. For each fold in 10-fold cross validation, data are weighted for choosing different cost ratios, the system on each weighted set is trained, true positive and false positive values are calculated, and the points on the ROC axes are plotted. The classification performance is measured by mean area under the curve (AUC), which is useful metric of classifier performance. The bigger area it is, the better classifier model is [40].

Third statistical index used to show the performance of the classification model is F-measure which shows the performance and efficiency of the model created, and looks for potential imbalance problems. F-measure is calculated as [42]:

$$F\text{-measure} = \frac{2\text{TP}_i}{2\text{TP}_i + \text{FP}_i + \text{FN}_i} \tag{19}$$

where TP_i is number of correctly classified data of ith class; FP_i is number of heartbeats incorrectly classified to ith class; FN_i is number of heartbeats of class i, classified in another class [42].

3.2. Experimental results

An automated classifier is designed to classify heartbeat signals belonging to two categories: N (normal heartbeats) and CHF (congestive heart failure). In both cases, the whole data set is divided into 10 subsets of (roughly) equal size and 10-fold cross validation presented by Salzberg [43] was applied to have accurate validation of the results. The results of each fold (subset) are summed and averaged and average accuracy is found. The signals were normalized before classification. Furthermore, experiments in this research are done to examine the capability of classification methods in classifying and detecting heartbeat signals. To be precise, those classification methods are C4.5 decision tree, k-nearest neighbor (k-NN), artificial neural networks (ANN), support vector machines (SVM) and random forest (RF) algorithms.

Three different statistical indices are used to show the performance of the above mentioned algorithms in classifying the ECG heartbeat signals: receiver operating characteristics (ROC), F-measure and accuracy for both classes (N, CHF). All statistical parameters are given in Tables 1 and 2, and Figs. 4–6.

Generally, all the classifiers performed well with the all statistical indices of high value. However, performance of random forest classifier stands out. It reached classification accuracy of 100% with ROC curve and F-measure of value 1 in both cases. Moreover, SVM classifier reached quite high performance, with accuracy value of 99.96%, ROC with value 1 and F-measure with value 1 in first case. Only one instance belonging to heartbeats with congestive heart failure was misclassified. However, performance of SVM is not quite significant in second case. The lowest accuracy value showed the

Table 1 – Performance evaluation of different machine learning methods (BIDMC congestive heart failure + MIT BIH Arrhythmia databases).

	C4.5 decision tree			k-NN			SVM			ANN			RF		
	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)
Normal	0.998	0.998	99.92	0.999	0.997	100	1	1	100	0.999	0.993	99.46	1	1	100
CHF	0.998	0.999	99.80	0.999	0.998	99.53	1	1	99.93	0.999	0.994	99.2	1	1	100
Average	0.998	0.999	99.86	0.999	0.998	99.75	1	1	99.96	0.999	0.993	99.32	1	1	100
Comput. time		0.19 s			0.07 s			0.43 s			12.01 s			0.77 s	

Table 2 – Performance evaluation of different machine learning methods (PTB Diagnostic ECG + MIT BIH Arrhythmia databases).

|	C4.5 decision tree			k-NN			SVM			ANN			RF		
	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)	ROC area	F-measure	Accuracy (%)
Normal	1	1	99.92	0.997	0.998	100	0.833	0.979	100	0.824	0.977	100	1	1	100
CHF	1	0.997	100	0.997	0.988	97.66	0.833	0.8	66.67	0.824	0.787	64.91	1	1	100
Average	1	0.999	99.93	0.997	0.997	99.73	0.833	0.961	96.13	0.824	0.955	95.92	1	1	100
Comput. time		0.08 s			0.04 s			0.30 s			9.86 s			0.28 s	
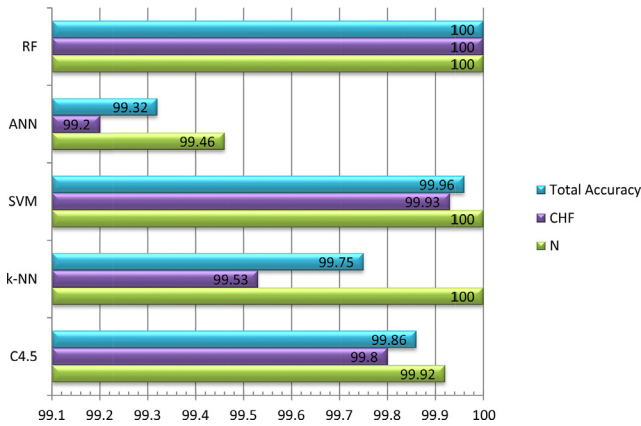
Fig. 4 – Graphical representation of performance evaluation of classifiers (BIDMC congestive heart failure + MIT BIH Arrhythmia databases).
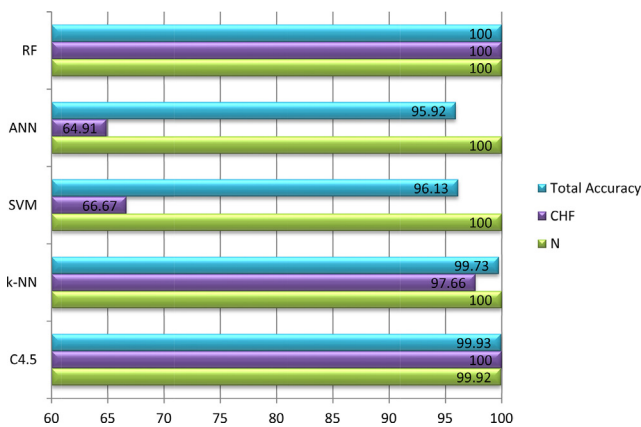

Fig. 5 – Graphical representation of performance evaluation of classifiers (PTB Diagnostic ECG + MIT BIH Arrhythmia databases).
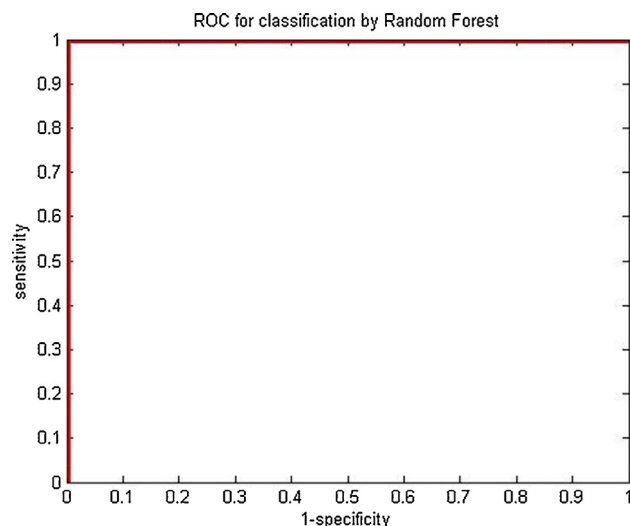

Fig. 6 – ROC curve for test data with N and CHF combined.

ANN model, in both cases. The performance of C4.5 decision tree and k-NN is significant, as well. With accuracy values of 99.86 and 99.75 in first case, and 99.93 and 99.73 for C4.5 and k-NN, respectively, it is indicated that both of the classifiers are capable to separate different classes of heartbeats. ROC and F-measure are 0.998 and 0.999 in first case and 1 and 0.999 in second case for C4.5 classifier and 0.999 and 0.998 in first case and 0.997 and 0.997 in second case for k-NN classifier. In addition, it can be seen in Tables 1 and 2 that all classifiers do not significantly outperform each other.

4.   Discussion

The results in previous section confirmed the advantage of the proposed RF method over the C4.5, SVM, ANN and k-NN methods. As it can be seen from the Tables 1 and 2, RF method gives the best results when considering all three statistical indices, meaning that all instances are correctly classified. Number of trees and random features affects the performance of the random forest classifier. The classification performance of the SVM in Table 1 shows that it can be applicable method for identification of the heart failure. Only one instance belonging to the group of congestive heart failure was misclassified; however SVM classifier did not perform fine in second case, shown in Table 2. Noticeable performance is achieved by C4.5 decision tree in both cases. C4.5 misclassified only 3 instances belonging to CHF in first case and none of the CHF samples in second case. This implies that C4.5 is suitable for classification of the heart beat signals as well.

Obviously, the characteristics of the classifiers and its combination with feature selection algorithm determine the classifier's performance. However, the accuracy of the ECG signal classification may be affected by numerous different factors, like the quality of the signals obtained, performance of the feature extraction algorithm as well as the subsets used for purpose of training and testing [6].

Random forest classifier achieved significant advantage over the classifiers applied in previous papers, noticeable in Table 3. Thuraisingham [8] combined SODP with k-NN classifier, which resulted in accuracy of 100%. The author achieved remarkable success with k-NN classifier in distinguishing ECG signals. When we compared k-NN and RF in speed and complexity, it is obvious that k-NN modeled the classification model faster than RF in this study (Tables 1 and 2), and it is less complex than RF. Still, RF achieved higher classification accuracy, ROC and F-measure values. Considering the overall performance of both classifiers, RF is preferred over k-NN. However, Hossen and Al-Ghunaimi [11,12] did not achieved quite remarkable success when compared to our result. Yu and Lee [13] pointed out that SVM classifier can truly achieve reasonable success.

Moreover, we investigated the success of random forest classifier in previous paper and various databases, presented in Table 4. Thus, Emanet obtained signals from MIT–BIH Arrhythmia database according to five types of ECG beats [44]. The classifier achieved remarkable success in differentiating the beats.

Belgacem et al. [45] obtained ECG signals from four PhysioNet ECG databases (MIT–BIH, ST-T, NSR, PTB) and one

Table 3 – Comparison of the results achieved in this study with previous studies.

Author(s)	Method(s)	Feature selection	Accuracy (%)
Thuraisingham [8]	k-Nearest neighbor classifier	Second-order difference plot (SODP)	100
Kuntamalla and Reddy [9]	Sequential trend analysis	Electrocardiogram (ECG)	96.68
Kuntamalla and Reddy [10]	Multiscale entropy	Empirical mode decomposition (EMD)	100
Hossen and Al-Ghunaimi [11]	Soft decision technique	Frequency-domain analysis	90
Hossen and Al-Ghunaimi [12]	Soft decision technique	Wavelet decomposition	88.6
Yu and Lee [13]	Support vector machine (SVM)	Conditional mutual information feature selector (CMIFS)	97.59
Isler and Kuntalp [14]	k-Nearest neighbor classifier	Genetic algorithm	96.39
Asyali [13,15]	Bayesian classifier	Linear discriminant analysis	93.24
Pecchia et al. [16]	Classification and regression tree (CART)	Short-term HRV analysis	94.85
This study	Random forest classifier	Autoregressive Burg method	100

Table 4 – Comparison of the achievements of random forest classifier applied to different databases.

Author(s)	Feature selection	Accuracy (%)
Emanet [44]	Discrete wavelet transform (DWT)	99.8
Belgacem et al. [45]	Discrete wavelet transform (DWT)	100
Kumar and Kumaraswamy [46]	Discrete cosine transform (DCT)	92.2
Jovic and Bogunovic [47]	Combination of linear and nonlinear features	93
This study	Autoregressive Burg method	100

collected from their students. They applied random forest classifier to recognize healthy subjects and achieve 100% success.

MIT–BIH Arrhythmia database was used by Kumar and Kumaraswamy [46]. They tried to differentiate two arrhythmias (LBBB and RBBB) and normal beats. They applied random forest classifier to differentiate heartbeats and achieved 92.2% classification success.

For purpose of their study, Jovic and Bogunovic [47] used signals obtained from four PhysioNet database to differentiate normal, any arrhythmia, supraventricular arrhythmia and congestive heart failure by applying random forest classifier. The classifier achieved 93% success in differentiating normal and supraventricular arrhythmia.

All above emphasizes the strong power of random forest classifier in correctly differentiating ECG signals. Moreover, our proposed systems reached 100% of sensitivity and specificity, implying that all normal signals were classified correctly with false negative (FN) rate of 0 and all signals with congestive heart failure were classified correctly with false positive (FP) rate of 0. The system might help in early diagnosis of the failure and right treatment. Moreover, the proposed system can be applied in detecting different classes of ECG signals.

According to the results of the present study and understanding of ECG signals classification problems, the following can be emphasized:

- High classification accuracy of RF classifier gives the insight into the feature that defines the ECG signals. The results in the study demonstrated that AR Burg coefficients are the features that represented ECG signals and facilitate the aim of classifier in distinguishing two classes of signals.
- C4.5 decision tree, k-NN, SVM and ANN are suitable classifiers for diagnosis and identification of heart failure; but RF classifier has an advantage over mentioned classifiers based

on the high classification accuracy and values of ROC curve and F-measure.

- For training the RF classifier, appropriate number of trees and random features is selected. Optimal values for trees and random features are obtained by inserting different values of trees and random features, using the classification accuracy as fitting function.
- The impressive testing performance of the RF classifier indicated that the model can be used in clinical studies and practice, to support doctors' decisions.

5.   Conclusion

In this paper, automated heartbeat classification system is developed for detecting congestive heart failure. ECG signals with congestive heart failure are obtained from BIDMC database and ECG signals with normal heartbeats are obtained from MIT–BIH Arrhythmia database. The feature vectors of datasets are extracted by using autoregressive (AR) Burg method. Dataset is extracted into 2800 ECG data segments where 1500 belongs to heartbeats with congestive heart failure and remaining 1300 data segments belongs to normal heart beats. The same dataset was used with five different classification methods, namely C4.5 decision tree, k-NN, SVM, ANN and RF. Among all classifiers, which achieved high statistical measures (accuracy, ROC curve, F-measure), random forest algorithm has significant role in identification and classification of ECG signals, demonstrated by accuracy of 100%, ROC curve of value 1 and F-measure of value 1, as well by success rate of the classifier presented. However, impressive performance of RF is proved by applying it to the second data set, composed of congestive heart failure, obtained from PTB Diagnostic ECG database. Autoregressive (AR) Burg method is applied to extract features and five

classification methods to classify features into normal and congestive heart failure, respectively. Random forest proved its impressive performance by achieving 100% classification accuracy, ROC curve of value 1 and F-measure of value 1. Performance of RF classifier demonstrated that this classifier can be valuable to express knowledge useful in medicine.

REFERENCES

[1] W. Horobin (Ed.), Diseases and Disorders, Marshall Cavendish Corporation, 2008.

[2] C.E. Quinn, 100 Questions and Answers About Congestive Heart Failure, Jones and Bartlett Publishers, Inc., 2006.

[3] N. Townsend, R. Luengo-Fernandez, J. Leal, A. Gray, M. Nichols, European Cardiovascular Disease Statistics 2012, 2012.

[4] C. Passanisi, Electrocardiography, Delmar Learning, New York, 2004.

[5] C.-S. Son, Y.-N. Kim, H.-S. Kim, H.-S. Park, M.-S. Kim, Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches, J. Biomed. Inform. (2012) 999–1008.

[6] A. Subasi, Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders, Comput. Biol. Med. (43) (2013) 576–586.

[7] D.S. Baim, W.S. Colucci, E.S. Monrad, H.S. Smith, R.F. Wright, A. Lanoue, F.D. Gauthier, B.J. Ransil, W. Grossman, E. Braunwald, Survival of patients with severe congestive heart failure treated with oral milrinone, J. Am. Coll. Cardiol. 7 (March (3)) (1986) 661–670.

[8] R.A. Thuraisingham, A classification system to detect congestive heart failure using second-order difference plot of RR intervals, Cardiol. Res. Pract. (November) (2009).

[9] S. Kuntamalla, L.R.G. Reddy, Detecting congestive heart failure using heart rate sequential trend analysis plot, Int. J. Eng. Sci. Technol. 2 (12) (2010) 7329–7334.

[10] S. Kuntamalla, L.R.G. Reddy, Reduced data dualscale entropy analysis of HRV signals for improved congestive heart failure detection, Meas. Sci. Rev. 14 (5) (2014) 294–301.

[11] A. Hossen, B. Al-Ghunaimi, Identification of patients with congestive heart failure by recognition of sub-bands spectral patterns, World Acad. Sci. Eng. Technol. (2008) 21–24.

[12] A. Hossen, B. Al-Ghunaimi, A wavelet-based soft decision technique for screening of patients with congestive heart failure, Biomed. Signal Process. Control 2 (2007) 135–143.

[13] S.-N. Yu, M.-Y. Lee, Conditional mutual information-based feature selection for congestive heart failure recognition using heart rate variability, Comput. Methods Programs Biomed. I (8) (2012) 299–309.

[14] Y. Isler, M. Kuntalp, Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure, Comput. Biol. Med. 37 (2007) 1502–1510.

[15] M.H. Asyali, Discrimination power of long-term heart rate variability measures, in: Engineering in Medicine and Biology Society: Proceeding of the 25th Annual International Conference of the IEEE EMBS, Cancum, 2003.

[16] L. Pecchia, P. Melillo, M. Sansone, M. Bracale, Discrimination power of short-term heart rate variability measures for CHF assessment, IEEE Trans. Inf. Technol. Biomed. 15 (1) (2011) 40–46.

[17] Z. Masetic, A. Subasi, Detection of congestive heart failure using C4.5 decision tree, Southeast Eur. J. Soft Comput. 2 (2) (2013) 74–77.

[18] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (6) (2003) 1947–1958.

[19] O. Maimon, L. Rokach, Data Mining and Knowledge Discovery Handbook, Springer Science + Business Media, New York, 2005.

[20] A.L. Goldberger, L.A. N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiological Signals, 2000.

[21] R. Palaniappan, Biological Signal Analysis, Ventus Publishing, 2010.

[22] W.J. Emery, R.E. Thomson, Data Analysis Methods in Physical Oceanography, Elsevier, Amsterdam, 2004.

[23] T.M. Mitchell, Machine Learning, McGraw-Hill Science/Engineering/Math, 1997.

[24] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, Inc., 1993.

[25] E. Fix, J.L. Hodges, Discriminatory Analysis. Nonparametric Discrimination; Consistency Properties, Texas, 1951.

[26] L.E. Peterson, K-Nearest Neighbor, 2009, Available at: http://www.scholarpedia.org/article/K-nearest_neighbor.

[27] E.M. Blalock (Ed.), A Beginner's Guide to Microarrays, Kluwer Academic Publishers, 2003.

[28] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, 2nd ed., Springer Science, 2009.

[29] W.S. McCulloch, W.H. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. (5) (1943) 115–133.

[30] B. Krose, P. van der Smagt, An introduction to Neural Networks, 8th ed., University of Amsterdam, 1996.

[31] L. Fausett, Fundamentals of Neural Networks, Architectures, Algorithms, and Applications, 1993.

[32] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533–536.

[33] Z.-G. Che, T.-A. Chiang, Z.-H. Che, Feed-forward neural networks training: a comparison between genetic algorithm and back-propagation learning algorithm, Int. J. Innov. Comput. Inf. Control 7 (October) (2011) 5839–5850.

[34] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: Fifth Annual ACM Conference on Computational Learning Theory, 1992.

[35] A. Ben-Hur, J. Weston, A user's guide to support vector machine, Methods Mol. Biol. (2010).

[36] B. Ustun, W.J. Melssen, L.M.C. Buydens, Facilitating the application of support vector regression by using a universal Pearson VII function based kernel, Chemom. Intell. Lab. Syst. (2006) 29–40.

[37] L. Breiman, Random forests, Mach. Learn. 45 (October (1)) (2001) 5–32.

[38] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed., Elsevier Inc., 2005.

[39] R. Zhang, G. McAllister, B. Scotney, S. McClean, G. Houston, Combining wavelet analysis and Bayesian networks for the classification auditory brainstem response, IEEE Trans. Inf. Technol. Biomed. 10 (3) (2006) 458–467.

[40] L. Rokach, O. Maimon, Data Mining with Decision Trees: Theory and Applications, World Scientific Publishing Co. Pte. Ltd., 2008.

[41] J.A. Hanley, B.J. McNeil, The meaning and the use of the area under a receiver operating characteristic curve, Radiol. Soc. N. Am. 143 (April (1)) (1982) 29–36.

[42] I. Jekova, G. Bortolan, I. Christov, Assessment and comparison of different methods for heartbeat classification, Med. Eng. Phys. 30 (2008) 248–257.

[43] S. Salzberg, On comparing classifiers: pitfalls to avoid and a recommended approach, Data Min. Knowl. Discov. (2007) 317–328.

[44] N. Emanet, ECG beat classification by using discrete wavelet transform and random forest algorithm, in: Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, Famagusta, 2009.

[45] N. Belgacem, A. Nait-Ali, R. Fournier, F. Bereksi-Reguig, ECG based human authentication using wavelets and random forests, Int. J. Cryptogr. Inf. Secur. 2 (June (2)) (2012).

[46] R.G. Kumar, Y.S. Kumaraswamy, Investigating cardiac arrhythmia in ECG using random forest classification, Int. J. Comput. Appl. 37 (January (4)) (2012) 31–34.

[47] A. Jovic, N. Bogunovic, Random forest-based classification of heart rate variability signals by using combinations of linear and nonlinear features, in: XII Mediterranean Conference on Medical and Biological Engineering and Computing, Chalkidiki, Greece, 2010.