Neural Comput & Applic (2014) 24:317–326
DOI 10.1007/s00521-012-1232-7

ORIGINAL ARTICLE

Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats

Ahmet Mert · Niyazi Kılıç · Aydın Akan

Received: 18 June 2012 / Accepted: 15 October 2012 / Published online: 26 October 2012

Abstract We explore the effect of using bagged decision tree (BDT) as an ensemble learning method with proposed time-domain feature extraction methods on electrocardiogram (ECG) arrhythmia beat classification comparing with single decision tree (DT) classifier. RR interval is the main property which defines irregular heart rhythm, and its ratio to the previous value and difference from mean value are used as morphological feature extraction methods. Form factor, its ratio to the previous value and difference from mean value are used to express ECG waveform complexity. In addition, skewness and second-order linear predictive coding coefficients are added to the feature vector of 56,569 ECG heart beats obtained from MIT–BIH arrhythmia database as time-domain feature extraction methods. The quarter of ECG heart beat samples are used as test data for DT and BDT. The performance measures of these classifiers are evaluated using the metrics such as accuracy, sensitivity, specificity and Kappa coefficient for both classifiers, and the performance of BDT classifier is examined for number of base learners up to 75. The BDT results in more predictive performance than DT according to the performance measures. BDT with 69 base learners has 99.51 % of accuracy, 97.50 % of sensitivity, 99.80 % of specificity and 0.989 of Kappa coefficient while DT gives 98.78, 96.05, 99.57 and 0.975 %, respectively. These metrics show that the suggested BDT increases the numbers of successfully identified arrhythmia beats. Moreover, BDT with at least three base learners has higher distinguishing capability than DT.

A. Mert (✉)
Department of Marine Engineering,
Piri Reis University, 34940 Istanbul, Turkey
e-mail: amert@pirireis.edu.tr

N. Kılıç · A. Akan
Department of Electrical and Electronics Engineering,
Istanbul University, 34320 Istanbul, Turkey
e-mail: niyazik@istanbul.edu.tr

A. Akan
e-mail: akan@istanbul.edu.tr

Keywords Arrhythmia classification · Ensemble learning · Bagged decision tree · Kappa coefficient

1 Introduction

Heart is a special muscle which its cells (myocytes) control two main functions namely as nervous (electrical) activity and mechanical tension with force feedback. Contraction of the heart is controlled by sino-artrial node (SA node) which is the part of the heart's conduction. Periodicity of electrical signal from SA node and its intrinsic electrical conduction form the heart beat variability and the heart's contraction sequence. Myocytes electrical activity causes potential difference on the skin surface which is non-invasively measured and recorded by electrocardiography [1]. The recording is called electrocardiogram (ECG) which is used to analyze the heart rate and regularity. Since detected electrical activity in ECG represents the regional muscular activities, the electro-mechanical function of myocytes region can be diagnosed [2]. A normal ECG signal consists of three basic waves including P, QRS and T which are induced by electrical activity on the cardiac surface. These waves are formed by the atrial depolarization, the ventricular depolarization and the ventricular repolarization sequentially [3]. A disease caused by described heart conduction system is named arrhythmia which defines an irregular heartbeat or an irregular group of heartbeats [4], and it can be diagnosed effectively based on long-term ECG recordings [5].

Since it is a difficult process to detect arrhythmia heart beats in long-term ECG recording, machine learning algorithms become supportive tools in clinical environments to help physicians improve diagnostic accuracy [6]. In Brause's study, diagnosis with the help of machine learning algorithms increases accuracy to 91.1 % while the accuracy of diagnosis by experienced physicians is 79.97 % [7].

Accurate analysis of ECG signals for arrhythmia diagnosis is the subject of pattern recognition and depends on feature extraction and classifier methods [8]. Both these stages have a definite effect on diagnostic accuracy. That's why, several methods are applied. The first stage, feature extraction for EGC signal can be categorized into three main types namely time, frequency and time-frequency domain analysis. The time-domain features are called morphological and complexity features [9]. The well-known morphological feature extraction method is to find RR interval which is also used to determine heart rate [10]. The other morphological features can be summarized as PR, QRS and ST lengths, amplitude, slopes depend on characteristics of required cardiac disease classification [11, 12]. Morphological and complexity measures [13] are so noise sensitive that preprocessing and filtering should be well designed. However, advanced time-domain feature extraction methods for ECG are principal component analysis (PCA), independent component analysis (ICA), higher-order statistics (HOS), correlation coefficients and linear predictive coding (LPC) which require more complex algorithms and computations, but they can be less sensitive to noise [3, 14–19]. Moreover, form factor (FF) is another time-domain feature extraction method which has been successfully applied to electroencephalography (EEG) classification which is suggested method for diagnosis of normal beat and ectopic beat in ECG [8]. Discrete wavelet transform (DWT) has taken attention in ECG classification [20] and became as one of the most popular and applied methods for time-frequency feature extraction of ECG signal [21]. It decomposes a signal into sub-bands. After determining which sub-bands represent ECG waveform without noise, its coefficients are used as feature vector. In addition to DWT coefficients, HOS methods can be applied to each sub-band of the ECG signal for more effective features [9, 22].

The second stage which has decisive effect on the performance is the machine learning algorithm to assign extracted unknown patterns to true classes. In general, learning approaches are categorized as supervised and unsupervised in pattern recognition. In case of a set of training samples with known class output is used by learning algorithm to predict unknown samples' classes, it is called supervised learning. In unsupervised learning also known as clustering, similarities in samples are used to assign into classes by another algorithm. Many classifiers include a single classifier such as nearest neighbor, decision tree (DT), artificial neural networks (ANN) and support vector machines (SVM) for prediction. These classifiers have been successfully applied to biomedical signal classification [23–26]. However, selecting the best classifier is an open issue because of varying input samples [27]. Therefore, classifier ensemble method has been proposed to decrease prediction error for learning algorithms [28]. There are the four ensemble learning types: bootstrap aggregating (bagging), boosting, random subspace and stacking. Bagging learning is based on assigning unknown data using several classifiers trained by bootstrap sample of training data. The output of each learner is applied to voting stage to assign data to most voted class [29]. Boosting method uses varying weighted input pattern for each learner and optimizes the weights of the lowest prediction error [30, 31]. In contrast, there are different types of machine learning algorithms in stacking ensemble method. The training data are applied to classifiers, and the outputs of them are used as metadata to be classified by a final classifier [2, 32]. Moreover, in random subspace method, the features of input pattern are divided into subsets and applied to individual learners especially in large data sets [33, 34].

We investigate ECG arrhythmia classification using bagged decision tree (BDT) as an ensemble learning method with extracted time-domain features. RR- and FF-based morphological feature extraction method is combined with skewness and LPC coefficients. Totally, nine features are extracted using time-domain methods: RR interval, FF, ratio of RR and FF to the previous values (RRR and FFR), RR and FF differences from mean RR and FF (RRM and FFM) with skewness and second-order LPC coefficients. Therefore, noise sensitivity problem of time-domain methods is removed using ratios and differences from mean value as well as skewness and LPC for robust feature combination. The extracted feature set computed using ECG signals in MIT–BIH arrhythmia database [35] is classified using single DT and BDT in order to investigate the effect of bagging learning method on ECG classification. The performance measures of both classifiers presented in the forms of confusion matrix, accuracy and sensitivity, specificity and Kappa statistic show that the first usage of the proposed BDT with the described time-domain feature extraction methods increases the numbers of successfully diagnosed ECG heart beats. This paper continues with the details of the proposed methods presented in Sect. 2 and the results given in Sect. 3. Discussion and conclusion are given in Sects. 4 and 5, respectively.

2 Materials and method

2.1 MIT–BIH arrhythmia database

MIT–BIH arrhythmia database contains approximately 30 min ECG recordings of 47 patients and generally used as a standard test database for the evaluation of arrhythmia classifiers. The sampling frequency of two channels including a modified limb lead II and one of the modified leads among V1, V2, V4 and V5 is 360 Hz.

In this study, we used six heartbeat types which are normal rhythm (N), left bundle branch block (LBBB), right bundle branch block (RBBB), atrial premature beat (APB), premature ventricular contraction (PVC) and paced beat (PB). Totally 56,569 heart beats are obtained from 22 ECG signals, and the distribution of heart beats is given in Table 1.

The selected 22 ECG recordings including the six types of heart beat with different rates are used in this study. The classification of arrhythmia beat using BDT is investigated using this data set described in the next sections.

2.2 Proposed feature extraction for ECG signals

The normal ECG rhythm defines regular heart rhythm and waveform that can be easily observed. However, the ECG signals of the patients with arrhythmia do not have regular rhythm and waveforms which the points in QRS cannot be observed manually. Therefore, the morphological properties of the heart beat such as RR interval and QRS width are the main rules of arrhythmia detection used by physicians. It is expected that RR intervals in an ECG signal of a healthy patient are almost the same while RR intervals of arrhythmia beats of an unhealthy patient are varying. However, RR interval can be a weak feature for the classification of arrhythmia types, because it does not contain information about the waveform complexity and other segments of the ECG signal. For this reason, another feature extraction method, form factor (FF) is used to represent the waveform complexity of the EGC signal. LBBB and normal heart beats of ECG signals' waveforms are given in Fig. 1 to show the effect of waveform complexity on arrhythmia detection.

Table 1 Composition of the heart beat data set

Beat types	Number of beats	Rate (%)
N	39,198	69.29
LBBB	5,489	9.70
RBBB	5,890	10.41
APB	782	1.38
PVC	2,895	5.11
PB	2,315	4.09
Total	56,569	100
ECG recordings	101 103 105 106 107 108 109 111 112 113 114 115 116 118 119 124 201 200 207 209 212 213	

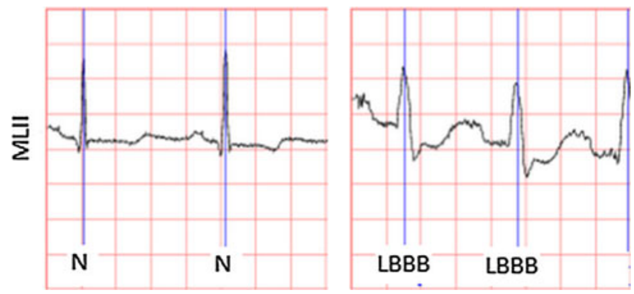

Fig. 1 LBBB and normal beat ECG waveforms [35]

The first parameter of FF called activity is the variance (σ_x^2) of the segmented signal (x_n). The second parameter namely mobility (M_x) is found calculating the square root of the ratio of the activity of the first derivative of the segmented signal ($\sigma_{x'}^2$) to the activity of the segmented original signal (σ_x^2) Finally, FF is the ratio of the mobility of the first derivative of the signal to mobility of the original signal, and these are formulated as follows [8];

$$M_x = \left[\frac{\sigma_{x'}^2}{\sigma_x^2}\right]^{\frac{1}{2}} \tag{1}$$

$$\mathrm{FF} = \frac{M_{x'}}{M_x} = \frac{\sigma_{\ddot{x}}/\sigma_{\dot{x}}}{\sigma_{\dot{x}}/\sigma_x} \tag{2}$$

Higher-order statistics (HOS) is another applied feature extraction method in this study. HOS is described an effective tool to represent waveform complexity of the ECG signal, especially third-order cumulant has most powerful distinguishing capability compared to second- and fourth-order cumulants [36]. For zero mean discrete time signals, third-order cumulant can be determined by,

$$C_{3x}(k,l) = E\{x(n)x(n+k)x(n+l)\} \tag{3}$$

where E states the expectation operator, k and l state time lags. Special form of third-order cumulant with zero lag called as skewness (s) can be described by

$$s = E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] \tag{4}$$

where σ is the standard deviation to normalize output, and μ is the mean value of samples, in case of non-zero mean signals.

Linear prediction coding (LPC) predicts next samples of a signal from a linear combination of previous samples of the original signal. In other words, LPC is an all-pole IIR filter that can be computed by,

$$\tilde{x}(n) = \sum_{i=1}^{p} a_i x(n-i) \quad (5)$$

where $\tilde{x}(n)$ is the predicted sample, a_i is denominator polynomial called LPC coefficient, and p is the order of LPC. It is given that second-order LPC coefficients provide better distinguishing capability for EGC classification [26].

In this study, nine dimensional features of the ECG signals are computed using RR- and FF-based feature extraction, skewness and second-order LPC coefficients which can represent many ECG classes [16] after pre-processing stage. Two Butterworth IIR filters with different orders are used as a low pass (LP) and high pass (HP) filter to remove noise and DC bias. The tenth order LP filter has 53-Hz cut-off frequency, and the third-order HP filter has 0.75-Hz cut-off frequency. ECG recordings and their RR intervals can be found on the web page of PhysioBank ATM [37]. The toolbox on that page also provides R points in the required ECG recording as sample numbers and RR intervals as duration in text file format. The RR intervals in the text file are used directly as features, and R points as sample numbers are used for reference points of other feature extraction methods.

Each beat in ECG signals is segmented between 30 samples before referenced R point and 79 samples after R point for FF computing. Finally, RR, FF, RR and FF ratio to the previous values (RRR, FFR), the differences of RR and FF from mean values (RRM, FFM) are extracted as RR- and FF-based features that can be formulated by,

$$\mathrm{RR}(i) = R(i) - R(i-1) \quad (6)$$

$$\mathrm{FF}(i) = \frac{\sigma_{\ddot{x}}/\sigma_{\dot{x}}}{\sigma_{\dot{x}}/\sigma_{x}} \quad (7)$$

$$\mathrm{RRR}(i) = \mathrm{RR}(i)/\mathrm{RR}(i-1) \quad (8)$$

$$\mathrm{FFR}(i) = \mathrm{FF}(i)/\mathrm{FF}(i-1) \quad (9)$$

$$\mathrm{RRM}(i) = \mathrm{RR}(i) - \overline{\mathrm{RR}} \quad (10)$$

$$\mathrm{FFM}(i) = \mathrm{FF}(i) - \overline{\mathrm{FF}} \quad (11)$$

The other feature extraction methods are skewness and second-order LPC. The windowing is selected between 27 samples before R points and 60 samples after R point to increase classification accuracy after try and trial method. In brief, nine dimensional feature vectors computing RR- and FF-based morphological features, skewness and second-order LPC coefficients are considered. The block diagram of the proposed feature extraction method is given in Fig. 2.

A total of 56,569 heart beats obtained from 22 ECG recordings of MIT–BIH arrhythmia database are applied the feature extraction stages, and 9-dimensional feature vector of each beat is computed and saved for classification.

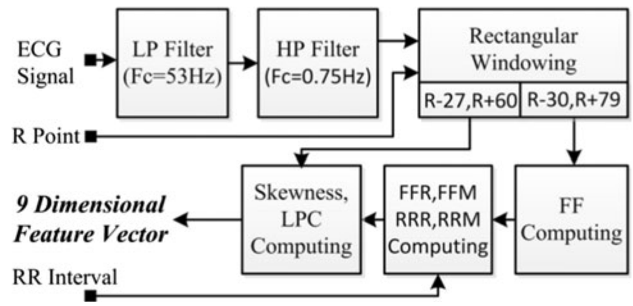

Fig. 2 The block diagram of the proposed feature extraction

2.3 Ensemble learning

Ensemble learning is the method of using multiple learning models to increase predictive performance. The prediction of each learning algorithm is combined in several methods such as majority voting and averaging [38, 39]. The well-known ensemble methods are listed in the literature as bagging, boosting, stacking and random subspacing [40]. Boosting method is a powerful procedure for combining the performance of each weak learner [30]. Each pattern in training data is weighted observing its effect on prediction error. After each iteration, the weights are determined and applied to the classifier. Random subspace method proposed by Ho [34] is another method of ensemble learning. In this method, features divided into random dimensionality subspaces are used to construct classifiers, and output is combined by majority voting. This method has advantages on classification of high dimensional data. However, it is still a problem how to select the optimum feature subspaces. In addition, stacking ensemble method is to add a new classifier to correct the errors of previous classifiers. The outputs of the previous classifiers are used as metadata for the last classifier. Thus, ensemble of various classifiers can be considered to increase predictive performance in the field of pattern recognition.

Bagging or bootstrap aggregating proposed by Breiman in 1996 [28] is a procedure for combining base learners or classifiers using the same training data set. The unknown test pattern is assigned to the class based on majority voting rule. The algorithm of bagging can be described by the following steps:

1. Training data (x_i, z_i) $i = 1,\ldots,n$,
2. For $b = 1,\ldots,B$
    a. *Generate bootstrap samples of training data, some instances will replicated, some will omitted*
    b. *Use bootstrapped data as training data for each classifier,* n_b.
3. *Classify test data using trained each classifier,* n_b and assign to the most represented.

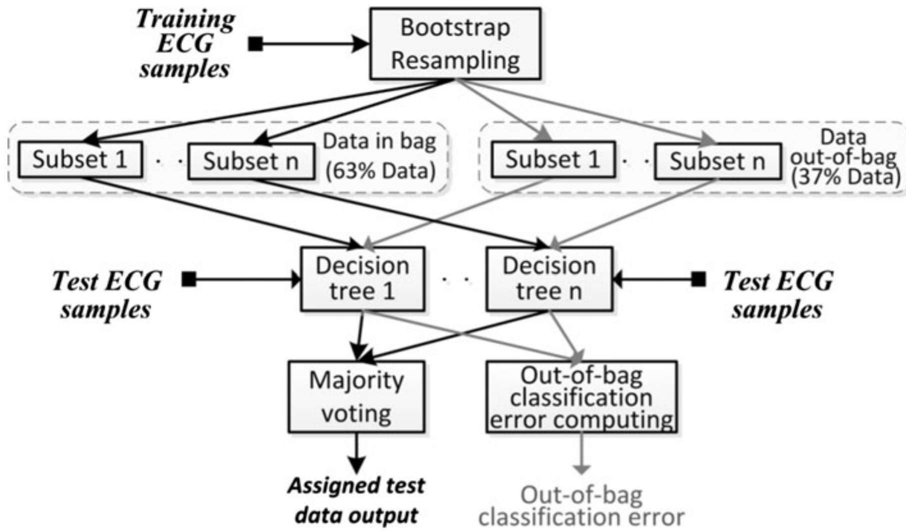

Fig. 3 The diagram of the bagged decision tree applied in this study

A pattern in bootstrap resampled training set has a probability of $1-(1-1/n)^n$ of being selected, and this is approximately $1-1/e=0.63$ for large n values, which means that each bootstrapped sample includes about 63 % unique pattern in the training data, namely data in bag. Thus, this different distribution in each sample causes different numbers of classifiers, and the remained patterns about 37 % of the training set can be used to evaluate ensembles of bagged decision trees before testing procedure, which is called out-of-bag (OOB) classification error. In contrast to this advantage, the bagging ensemble reduces the variance and increases the classification accuracy of only unstable base classifiers such as DTs and ANNs [40, 41]. In other words, k-nearest neighbor and Naïve Bayes classifiers are stable and not effective for bagging procedure [42].

In this study, bagging ensemble method is considered for the classification of arrhythmia heart beats. DT is selected as the base learner of the bagging method, and arrhythmia beat classification using the single DT with BDT is compared. The diagram of the BDT method applied in this study is given in Fig. 3.

Seventy-five percent of extracted 56,569 heart beats using MIT–BIH arrhythmia data set are used as testing instances for the BDT. The number of bootstrap resampling is varied between 2 and 75 to construct the same numbers of base learners namely DTs. Thus, DTs trained by subsets with non-uniform ECG sample distribution are grown up to 75. The effect of the numbers of base learners on bagging ensemble classifiers can be analyzed by observing the OOB error without applying any testing method such as partitioning and k-fold cross-validation because of the bootstrap resampling. In other words, 37 % of ECG observations (5,233 ECG heart beat samples) which are omitted from the training data in bootstrapping can be a practical way to examine bagging method. However, the test data are applied to construct DTs by bootstrap resampled training subsets to make final decision about the arrhythmia heart beat classification evaluating performance measures. Finally, 25 % of the samples are applied to the constructed DTs, and the final class assignment is decided based on majority voting rule.

2.4 Performance measures

The performance of a proposed classifier is evaluated comparing the actual value with predicted value. For this reason, the first step is to store the actual and assigned class attributes in the form of the confusion matrix given in Table 2.

Accuracy is the common method which indicates the overall performance of the proposed classification. Sensitivity and specificity are the other measures which indicate correctly identified actual positive samples and correctly identified negative samples, respectively. These measures can be found by

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{12}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{13}$$

Table 2 A confusion matrix

Actual	Assigned value	
Value	Positive	Negative
Positive	TP	FN
Negative	FP	TN

TP, TN, FP and FN refer true positive, true negative, false positive and false negative, respectively

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (14)$$

Kappa statistic [43, 44] is also a measure of the agreement between two raters, which is thought to be more robust because it eliminates agreements which can be attributed to chance [45]. Kappa value (k) is formulated as

$$k = \frac{p_0 - p_c}{1 - p_c} \quad (15)$$

where p_0 denotes the observed proportions of agreements, and p_c denotes the expected proportion of agreement. These are defined as

$$p_0 = \frac{\sum_{i=1}^{k} n_{ii}}{N} \quad (16)$$

$$p_c = \sum_{i=1,j=1,i=j}^{k} p_i p_j \quad (17)$$

where k is the number of classification categories, n_{ii} is the number of cases that comparison pair agrees as to classification in category i, N is the total number of cases, $p_{i.}$ and $p_{.j}$ are the marginal probabilities. The computed Kappa value defines the agreement level given in Table 3.

According to Kappa value, maximum value is one and defines total agreement. The agreement level of two raters decreases when its value decreases. It is desired to get maximum Kappa value when a classifiers output is compared with actual classes.

3 Experimental results

Arrhythmia hear beat classification is examined using the proposed time-domain feature extraction methods and BDT as described in the previous sections. The feature extraction process consists of time-domain methods based on RR interval and FF, higher-order statistics including skewness, and second-order LPC coefficients of ECG signal. Totally, 56,569 ECG heart beats obtained from MIT–BIH arrhythmia database are extracted for classification of six heart beat types namely normal, LBBB, RBBB, APB, PVC and PB. One of the ensembles learning method, BDT, is used as the classifier and compared to ECG classification using single DT. The two classifications and feature extraction algorithms of which details of each stage are given in the previous sections are written using MATLAB®, and the block diagram is given in Fig. 4.

Table 3 Interpretation values of Kappa statistic

Kappa value	Interpretation
1.00	Total agreement
0.75–1.00	Excellent agreement
0.40–0.75	Fairly good agreement
0.00–0.40	Poor level of agreement
0.00	Agreement entirely due to chance
<0.00	Lower than that expected by chance

25 % of 56,569 heart beat samples are remained as test data for BDT and DT, and the rest of the samples are used as training data for DT and BDT with varying numbers of base learners. Before testing, the OOB error is observed to evaluate the effect of grown DTs on the prediction error. This is a useful indicator to estimate BDT performance before complex testing computations. For this reason, the OOB error of the proposed BDT is given in Fig. 5.

The proposed BDT with up to 75 base learners has minimum OOB error (0.006058) when 69, 73, 74 and 75 base learners are used. This useful OOB error information resulted by classifying patterns in bag with out-of-bag patterns shows that the numbers of grown trees increase the predictive performance. To extend performance measures of the arrhythmia classification, test data are applied to the trained the BDT, and the effect of bagging method on arrhythmia detection is evaluated using metrics including

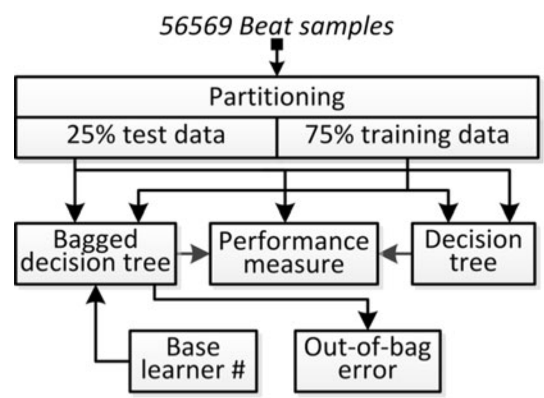

Fig. 4 Proposed arrhythmia heart beat classification

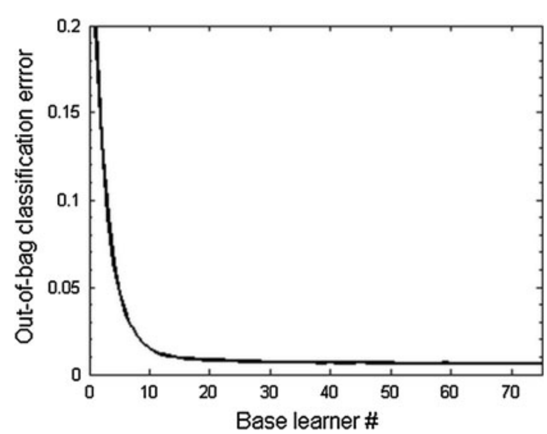

Fig. 5 OOB error of the proposed BDT classifier

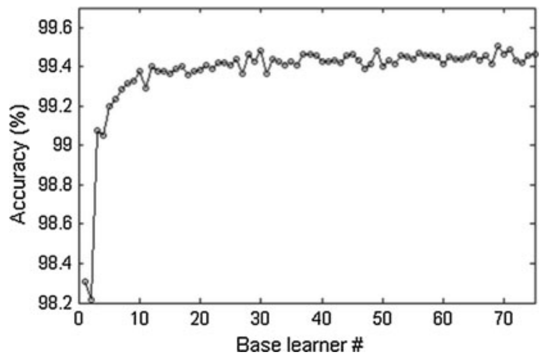

Fig. 6 The accuracy graph of arrhythmia classification using BDT

accuracy, sensitivity, specificity and Kappa value. Thus, the accuracy graph of bagging DT varying numbers of base learners is given in Fig. 6, while single DT results 98.78 % of accuracy.

The proposed arrhythmia classification using BDT trained by 25 % of extracted 56,569 ECG heart beat results 99.51 % of the maximum accuracy with only 69 base learners, although the OOB error rate is minimum when 69, 73, 74 and 75 base learners are used. That's why, different test data ratios affect the accuracy slightly, but the OOB error is an effective indicator to estimate the BDT's performance before testing procedure. Moreover, BDT results higher accuracy after the numbers of base learners are three (99.07 %) or more when compared to single DT, and the confusion matrices of the arrhythmia classification using BDT with 69 base learners and single DT are given in Table 4 to show the differences between successfully recognized heart beats using BDT and DT.

Referring Table 4, the counts of correctly predicted samples in the confusion matrix of the BDT classifier are higher, while misclassified samples are lower when compared to DT classifier. To extend the investigation of the effect of the proposed arrhythmia classification using BDT on the distinguishing capability for each class, TP, TN, FP, FN, sensitivity and specificity values of each class are computed and given in Table 5 using the counts in the confusion matrices of arrhythmia classification using BDT and DT.

Generally, TP and TN counts which state successfully recognized positive and negative samples for the BDT classifier are higher than DT, while FP and FN counts are lower. This results higher sensitivity and specificity values of each beat type classification. In addition, the suggested

Table 4 Confusion matrices of the classifiers

	Actual heart beat classes							
			N	LBBB	RBBB	APB	PVC	PB
Classifier's output	*N*	DT	9,747	16	29	19	16	0
		BDT	9,787	6	10	15	11	0
	LBBB	DT	13	1,349	4	1	7	0
		BDT	3	1,365	1	1	2	0
	RBBB	DT	17	5	1,437	0	4	0
		BDT	2	1	1,460	0	1	0
	APB	DT	14	0	0	168	3	0
		BDT	4	0	0	174	0	0
	PVC	DT	9	2	2	8	692	3
		BDT	4	0	1	6	709	2
	PB	DT	0	0	0	0	1	576
		BDT	0	0	0	0	0	577

Table 5 Results of arrhythmia classifications using suggested BDT and DT

Beat type	Sample #	TP		TN		FP		FN		Sensitivity (%)			Specificity (%)		
		DT	BDT	DT	BDT	DT	BDT	DT	BDT	DT	BDT	Increase	DT	BDT	Increase
N	9,800	9,747	9,787	4,262	4,300	80	42	53	13	99.46	99.87	0.41	98.16	99.03	0.87
LBBB	1,372	1,349	1,365	12,745	12,763	25	7	23	7	98.32	99.49	1.17	99.80	99.95	0.15
RBBB	1,472	1,437	1,460	12,644	12,666	26	4	35	12	97.62	99.18	1.56	99.79	99.97	0.18
APC	196	168	174	13,929	13,942	17	4	28	22	85.71	88.78	3.07	99.88	99.97	0.09
PVC	723	692	709	13,395	13,406	24	13	31	14	95.71	98.06	2.35	99.82	99.90	0.08
PB	579	576	577	13,562	13,563	1	0	3	2	99.48	99.65	0.17	99.99	100	0.01
Total	14,142							Average		96.05	97.50	1.45	99.57	99.80	0.23

Table 6 Overall performance measures of the BDT and DT

Measures	BDT	DT
Sensitivity (%)	97.50	96.05
Specificity (%)	99.80	99.57
Accuracy (%)	99.51	98.78
Kappa coefficient	0.990	0.975

BDT has more increasing effect on resulted lower sensitivities of DT classifier. For example, APC classification using DT has the lowest sensitivity (88.78 %), and this ratio is increased by 3.07 % and resulted 88.78 %. Specificity has similar behavior to the sensitivity, and it has more increase (0.87 %) for normal beat classification using BDT. Briefly, arrhythmia heart beat classification using the suggested BDT and the feature extraction methods decreases unsuccessfully recognized beat samples, especially APB. Final assessment on arrhythmia classification is given in Table 6 considering overall classification results of both BDT and DT.

Resulted Kappa values of BDT and DT are nearly one, and both classifications are named "excellent agreement", because predicted arrhythmia classes are nearly same to actual classes. However, Kappa value of the BDT classifier with 69 base learners is higher than DT's, which indicates the suggested feature extraction method with BDT has higher predictive performance.

4 Discussion

The main morphological feature extraction method for ECG signal classification is the RR interval. Its powerful distinguishing capability on irregular heart rhythms increases its use as a feature in medical diagnostic decision support systems. However, the detection of RR interval is noise sensitive, and it can cause misclassified hear beat samples. For this reason, filtering methods before feature extraction of RR interval should be well designed or RR interval should be combined to other feature extraction methods. Thus, FF, skewness and second-order LPC coefficients to extract the information of the ECG waveform complexity as well as the reported as successful methods [5], the ratios of RR and FF to the previous values and differences from mean values are used with RR as time-domain methods.

The machine learning methods are as decisive as feature extraction methods for ECG classification as well as for any pattern recognition problem. Various machine learning algorithms such as *k*-NN, ANN, SVM have been studied and successfully applied to ECG heart beat classification. However, ensemble learning methods to combine each learner's predictive performance are rarely applied to this field. Bagging decision tree is used as an ensemble method to increase the numbers of successfully recognized arrhythmia ECG beat samples. Considering accuracy metric indicates overall distinguishing performance of a classifier, single DT results 98.78 % of accuracy, while BDT with 69 base learners results 99.51 % of accuracy. In this study, we observe that BDT with three and more base learner provides higher predictive performance examining given accuracy (99.51 %), sensitivity (97.50 %), specificity (99.80 %), Kappa coefficient (0.989), while single DT results 98.78, 96.05, 99.57 and 0.975 %, respectively. That is the reason why BDT has more predictive performance on misclassified ECG samples of DT especially for APC beat type.

The comparison of this study with previous studies, in terms of the methodology, data set and accuracy is reported in Table 7. Since various methodology and heart beat number and types are used in the previous studies, it is not possible to make definite comparison. However, this suggested feature extraction method using RR interval and FF-based features, third-order cumulant and second-order LPC coefficients with decision tree and bagged decision tree classifiers have higher accuracy rate than the previous studies.

Table 7 Comparison of this study with previous studies

Authors	Methodology	Accuracy (%)
Tsipuaras et al. [10]	RR- and knowledge-based classifier (KBC)	98.20
Karpagachelv et al. [46]	Morphological and temporal features with extreme learning machine	89.78
Osowski and Linh [36]	Cumulants with fuzzy hybrid neural network	96.06
Langerholm et al. [47]	Hermite functions and RR with self-organized map	98.49
Dokur and Olmez [48]	DWT with intersecting spheres network	95.70
Kim [5]	PCA with extreme learning machine	98.72
YU and Chen [22]	DWT and HOS with neural network	97.28
Engin [16]	HOS, DWT and AR modeling with neuro-fuzzy networks	98.00
Proposed study	RR- and FF-based features HOS and second-order LPC coefficients with DT and BDT	98.78 (for DT) 99.51 (for BDT)

Finally, although BDT with 69 base learners has higher performance measures, BDT with at least three base learners can be used to increase the number of successfully recognized ECG beat samples in comparison with a single DT and the previous studies, which make BDT a successful classifier for ECG signals. Moreover, the time consuming of BDT algorithm with three learners takes approximately nine seconds, when DT consumes eight seconds on 64-bit Windows® 7 running Laptop PC with Intel® i3 2.27 GHz processor with 3 GB DDR3 RAM. However, time consumed increases approximately seventy seconds, in case 75 base learners are used.

5 Conclusion

In this study, we used bagged decision tree (BDT) which is the type of ensemble learning method as the arrhythmia heart beat classifier and compared with single decision tree (DT). Twenty-two ECG recordings obtained from MIT–BIH arrhythmia database are used to evaluate these classifiers. Totally, 56,569 heart beats are extracted using RR interval (RR) and form factor (FF)-based features, skewness and second-order linear predictive coding (LPC) coefficients for six types of arrhythmia heart beats. RR which is the main property to detect irregular heart rhythms is compared to previous RR and mean RR value. Thus, RR ratio to previous RR (RRR) and RR difference from mean RR value (RRM) is used to increase powerful morphological properties of RR. FF- and FF-based features; FF ratio to previous one (FFR) and FF difference from mean FF value (FFM) is computed like in RR-based features to represent ECG waveform complexity into a few coefficients. In addition to these, skewness and second-order LPC coefficients are added to features of the ECG signals. Finally, 9-dimensional feature vector for 56,569 heart beats is extracted. The quarter of the extracted ECG samples are used as test data for DT and BDT, and DT results 98.78 % of accuracy while BDT results 99.51 % of accuracy with 69 base learners and the defined feature extraction method. The other performance measures including sensitivity (97.50 %), specificity (99.80 %) and Kappa value (0.990) are higher for BDT classifier when compared to DT results 96.05 %, 99.57 % and 0.975, respectively. Finally, BDT has a higher predictive performance of arrhythmia hear beat classification considering the given performance measures when compared to DT. That is why, the BDT classifier can recognize false-negative samples of each class resulted by DT especially for atrial premature beat. In other words, BDT increases resulted sensitivity rates of DT classifier for each classes. In conclusion, the suggested combination of time-domain feature extraction methods and BDT with at least three base learners can be successfully used for arrhythmia decision support system to increase medical diagnostic accuracy.

Acknowledgments This work was partially supported by The Research Fund of The University of Istanbul. Project numbers: IRP-11824 and UDP-25231.

References

1. Sache FB (2004) Computational cardiology: modeling of anatomy, electrophysiology, and mechanics. Springer, Germany
2. Homaeinezhad MR, Atyabi SA, Tavakkoli E, Toosi HN, Ghaffari A, Ebrahimpour R (2012) ECG arrhythmia recognition via a neuro-SVM-KNN hybrid classifier with virtual QRS image-based geometrical features. Expert Syst Appl 39:2047–2058
3. Zhang H, Zhang LQ (2005) ECG analysis based on PCA and support vector machines. ICNN&B 2:743–747
4. Sandoe E, Sigurd B (1991) Arrhythmia–a guide to clinical electrocardiology. Publishing Partners, Bingen
5. Kim J, Shin HS, Shin K, Lee M (2009) Robust algorithm for arrhythmia classification in ECG using extreme learning machine. Biomed Eng Online 8:31. doi:10.1186/1475-925X-8-31
6. Özçift A (2011) Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. Comput Biol Med 41:265–271
7. Brause RW (2001) Medical analysis and diagnosis by neural networks, In Computer Science Department. Frankfurt a.m, Germany
8. Rangayyan RM (2001) Biomedical signal analysis: a case-study approach. Wiley-IEEE Press, USA
9. Yu SN, Chen YH (2009) Noise-tolerant electrocardiogram beat classification based on higher order statistics of sub-band components. Artif Intell Med 46:165–178
10. Tsipouras MG, Fotiadis DI, Sideris D (2005) An arrhythmia classification system based on the RR-interval signal. Artif Intell Med 33:237–250
11. Jekova I, Bortolan G, Christov I (2008) Assessment and comparison of different methods for heartbeat classification. Med Eng Phys 30:248–257
12. Asl BM, Setarehdan SK, Mohebbi M (2008) Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. Artif Intell Med 44:51–64
13. Chen SW (2007) Complexity-measure—based sequential hypothesis testing for real–time detection of lethal cardiac arrhythmias. Eurasip J Adv Sig Pr 1–8. doi:10.1155/2007/20957
14. Chiu CC, Lin TH, Liau BY (2005) Using correlation coefficient in ECG waveform for arrhythmia detection. Biomed Eng App Bas C 17:37–42
15. He T, Clifford G, Tarassenko L (2006) Application of independent component analysis in removing artefacts from the electrocardiogram. Neural Comput Appl 15:105–116
16. Engin M (2004) ECG beat classification using neuro-fuzzy network. Pattern Recogn Lett 25:1715–1722
17. Chawla MPS (2009) A comparative analysis of principal component and independent component techniques for electrocardiograms. Neural Comput Appl 18:539–556
18. Karimifard S, Ahmadian A (2011) A robust method for diagnosis of morphological arrhythmias based on Hermitian model of higher order statistics. Biomed Eng Online 10:1–18
19. Park KS, Cho BH, Lee DH, Song SH, Lee JS, Chee YJ, Kim IY, Kim SI (2008) Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function. Comput Cardiol 35:229–232

20. Al-Fahoum AS, Howitt I (1999) Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias. Med Biol Eng Comput 37:566–573
21. Yu SN, Chen YH (2007) Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. Pattern Recognit Lett 28:1142–1150
22. Yu SN, Chen YH (2008) Selection of higher order sub-band features for ECG beat classification. In: 16th EUSIPCO
23. Mert A, Kilic N, Akan A (2011) Support vector machines with reduced dimensionality using independent component analysis for breast cancer classification. In: ELMAR 2011 proceedings, pp 37–40
24. Thanapatay D, Suwansaroj C, Thanawattano C (2010) ECG beat classification method for ECG printout with principle components analysis and support vector machines. ICEIE 1:72–75
25. Mousa R, Munib Q, Moussa A (2005) Breast cancer diagnosis system based on wavelet analysis and fuzz-neural. Expert Syst Appl 28:713–723
26. Xiao Q, Jian CW, Fei GD (2011) ECG signal classification based on BPNN. ICEICE 2:1362–1364
27. Cavalin PR, Sabourin R, Suen CY (2011) Dynamic selection approaches for multiple classifier systems. Neural Comput Appl. doi:10.1007/s00521-011-0737-9
28. Breiman L (1996) Bagging predictors. Mach Learn 24:123–140
29. Zhu X, Yang Y (2008) A lazy bagging approach classification. Pattern Recogn 41:2980–2992
30. Freund Y, Shapire RE (1997) A decision—theoretic generalization of on line learning and an application to boosting. J Comput Syst Sci 55:119–139
31. Das R, Sengur A (2010) Evaluation of ensemble methods for diagnosis of valvular heart disease. Expert Syst Appl 37:5110–5115
32. Hothorn T, Lausen B (2003) Bagging tree classifiers for laser scanning images: a data- and simulation- based strategy. Artif Intell Med 27:65–79
33. Moon H, Ahn H, Kodell RL, Baek S, Lin CJ, Chen JJ (2007) Ensemble methods for classification of patients for personalized medicine with high-dimensional data. Artif Intell Med 41:197–207
34. Ho TK (1998) The random subspace method for constructing decision forests. IEEE T Pattern Anal 20:832–844
35. MIT–BIH arrhythmia database. http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm. Accessed 26 May 2012
36. Osowski S, Linh TH (2001) ECG beat recognition using fuzzy hybrid neural network. IEEE T Bio Med Eng 48:1265–1271
37. PhysioBank ATM. http://physionet.org/cgi-bin/atm/ATM?database=mitdb&tool=plot_waveforms. Accessed 26 May 2012
38. Jerez-Aragones JM, Gomez-Ruiz JA, Ramaos-Jimenez G, Munoz-Perez J (2003) A combined neural network and decision tree model for prognosis of breast cancer relapse. Artif Intell Med 27:45–63
39. Zhang Y, Zhong S (2012) A privacy-preserving algorithm for distributed training of neural network ensembles. Neural Comput Appl. doi:10.1007/s00521-012-1000-8
40. Webb AR, Coppesy KD (2011) Statistical pattern recognition. Wiley, Malvern
41. Tumer K, Ghosh J (1996) Error correlation and error reduction in ensemble classifiers. Connect SCI 8:385–404
42. Freidman JH (1997) On bias, variance, 0/1-loss, and the curse-of-dimensionality. Data Min Knowl Disc 1:55–57
43. Cohen J (1968) Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 70:213–220
44. Fleiss JL (1981) Statistical methods for rates and proportions. Wiley, New York
45. Berdinas BG, Betanzos AA (2002) Empirical evaluation of a hybrid intelligent monitoring system using different measures of effectiveness. Artif Intell Med 24:71–96
46. Karpagachelvi S, Arthanari M, Sivakumar M (2011) Classification of electrocardiogram signals with support vector machines and extreme learning machine. Neural Comput Appl. doi:10.1007/s00521-011-0572-z
47. Lagerholm M, Peterson C, Braccini G, Ebendrandt L, Sornmo L (2000) Clustering ECG complexes using hermite functions and self-organizing maps. IEEE T Bio Med Eng 47:838–848
48. Dokur Z, Olmez T (2001) ECG beat classification by a hybrid neural network. Comput Method Progr Bio 66:167–181