16th International Research/Expert Conference
”Trends in the Development of Machinery and Associated Technology”
TMT 2012, Dubai, UAE, 10-12 September 2012

ECG SIGNAL CLASSIFICATION USING ENSEMBLE DECISION TREE

Ahmet Mert
Piri Reis University, Dept. of Marine Eng.
Tuzla, 34940 Istanbul
Turkey

Niyazi Kilic and Aydın Akan
Istanbul University, Dept. of Electrical and Electronics Eng.
Avcilar, 34320 Istanbul
TURKEY

ABSTRACT
The electrocardiogram (ECG) is a non-invasive method to measure and record the electrical activity of the heart. ECG signal analysis has an important role on the diagnosis of heart diseases especially, abnormal or irregular heartbeats, namely arrhythmia. There are three basic waves; P, QRS and T in healthy EGC signal. The detection of these waves and time domain morphological properties represent the information about arrhythmia. Time intervals between waves or duration of a wave such as RR interval (RR) and QRS length are successful and well-studied methods of detecting arrhythmia. In addition, form factor (FF) is another technique to represent ECG waveform complexity in a scalar value. In this paper, arrhythmia beat classification using ensemble decision tree is studied. Bootstrap aggregating (bagging) decision tree is used as a type of ensemble learning. ECG signals from 22 patients including five arrhythmia beats and normal beats are obtained from MIT-BIH arrhythmia database. After the filtering process, 56569 ECG beats are collected and feature are extracted based on morphological properties including RR, FF, RR and FF ratio to previous values (RRR, FFR), RR and FF differences from mean values (RRM, FFM). 25% of 56569 beats is used as test data for bagged decision tree and the rest for training. The performance measures of bagged decision tree with varying 75 learners and single decision tree are evaluated to compare the effect of bagging decision tree on ECG beat classification. While bagged decision tree gives accuracy of 99.34%, decision tree yields 98.30% accuracy. Finally, we observe that the bagged decision tree for ECG arrhythmia beat classification can be successfully applied to increase the accuracy of ECG arrhythmia detection.
Keywords: ECG classification, ensemble learning, bagged decision tree.

1. INTRODUCTION
Electrical activity of the cardiac is measured on the skin by electrocardiography and the recordings are called electrocardiogram (ECG) signals which are widely analyzed to diagnose the functional abnormalities of the cardiac. However, variations in the ECG signals, recording noise, and that we need long-term recordings make ECG signal analysis a challenging study. Various methods in the fields of signal processing and pattern recognition are applied [1]to detect irregular heart rhythms namely arrhythmia.
There are two main stages to detect arrhythmia heart beats in ECG signal including feature extraction and classifier steps which are decisive effect on successfully recognized heart beat samples [2,3]. At the first stage, RR interval (RR) is the main morphological feature to detect arrhythmia beats, and the

extracted RR in an ECG recording is used to classify the beat type as normal or arrhythmia. However, RR is a noise sensitive method, and the filtering before extraction and combining its property with other features should be well-designed [4].Thus, form factor which is a successfully applied method in electroencephalography(EEG) can be used to present ECG waveform complexity into a coefficient to support RR distinguishing capability [5], and the ratio of the heart beat's extracted feature to previous value can be used to classify arrhythmia due to its higher discriminative power which eliminates the noise effect in ECG signals [6]. Besides, the machine learning algorithms, such as k-nearest neighbor (*k*-NN), decision trees (DT), artificial neural network (ANN), are the second stage as decisive as extracted features to assign the unknown patterns to true classes. For this reason, ensembles of the classifiers are suggested to increase predictive performance by combining several classifiers, and the outputs of individual classifiers are evaluated to make the final prediction.

In this study, we used bagged decision tree (BDT) classifier as a type of ensemble learning for arrhythmia heart beat recognition. ECG recordings of 22 patients from MIT-BIH Arrhythmia Database [7] are applied to feature extraction stage which consists of computing FF, FF ratio to previous FF value (FFR), FF difference from mean value (FFM), RR ratio to previous value (RRR) and RR difference from mean value (RRM) with obtained RR from the PhysioBank ATM[8] after band pass filtering. Totally, 6-dimesional feature vectors for 56569 heart beats including six beat types namely normal (N), left bundle branch block (L), right bundle branch block (R), atrial premature beat (A), premature ventricular contraction (P) and paced beat (PB) are extracted, and 25% of the ECG samples are used as test data for BDT classifier. The trained BDT is evaluated by test data using the performance measures including accuracy.

2. PROPOSED METHOD

2.1. Feature extraction

Firstly, the 22 ECG recordings are band-bass filtered by using different order low-pass (LP) and high-pass (HP) Butterworth IIR filters to remove DC bias and power line interferences in the pre-processing stage. The 10th order LP filter has 53 Hz cut-off frequency while 3rd order HP filter has 0.75 Hz cut-off frequency. After this pre-processing stage, filtered 22 ECG recordings are applied to the proposed feature extraction stage consists of windowing, RRR, RRM, FF, FFR and FFM computing. The block diagram of the proposed feature extraction is given in Figure 1.

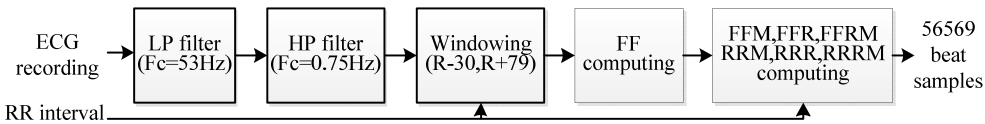

Figure 1. The block diagram of the proposed study

RR of the beats in each ECG recordings are given in the text format in the PhysioBank ATM including interval and the sample numbers of R points, and these are used referee points for windowing. FF is computed for the windowed ECG signal between 30 samples before R point and 79 samples after R point. Finally, RRR, RRM, FFR, FFM are computed and described by

$$RR(i) = R(i) - R(i-1) \quad (1)$$

$$FF(i) = (\sigma_{\ddot{x}} / \sigma_{\dot{x}}) / (\sigma_{\dot{x}} / \sigma_{x}) \quad (2)$$

$$RRR(i) = RR(i) / RR(i-1) \quad (3)$$

$$FFR(i) = FF(i) / FF(i-1) \quad (4)$$

$$RRM(i) = RR(i) - \overline{RR} \quad (5)$$

$$FFM(i) = FF(i) - \overline{FF} \quad (6)$$

where $\sigma_{\ddot{x}}, \sigma_{\dot{x}}$ states the variances of the second and first derivatives of the segmented ECG signal, respectively. Finally, 6 features are extracted for heart beats, and its distribuiton is given in Table 1.

Table 1. Heart beats distribution

Beat Type	**N**	**L**	**R**	**A**	**P**	**PB**	**Total**
Number	39198	5489	5890	782	2895	2315	56569
ECG Recordings	101 103 105 106 107 108 109 111 112 113 114 115 116 118 119 124 201 200 207 209 212 213						

Totally, 56569 heart beats for six classes are used as test and training data for the BDT classifier.

2.2. Bagged decision tree

Bagging or bootstrap aggregating proposed by Breiman in 1996 [9] is a type of ensemble learning which combines classifiers in order to get maximum accuracy that a single classifier cannot provide. In BDT classifier, the training data is divided into subsets by using bootstrap resampling, and each subset is used as training data to construct each decision tree. The number of bootstrapping defines the number of constructed DT, and outputs of DTs trained by different subsets are applied to majority voting stage [10]. The block diagram of BDT is shown in Figure 2.

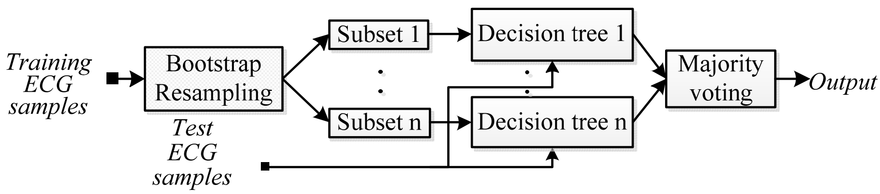

Figure 2. The block diagram of the BDT classifier

Majority voting stage assigns the unknown pattern in test data to most represented class predicted by DTs. 25% of the 56569 heart beats is used as test data to evaluate the performance of BDT computing accuracy.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The output of the BDT classifier is evaluated using the true classes of the test data for different number of base learners to show the effect of BDT on arrhythmia beat classification. Thus, the accuracy graph of the proposed BDT classifier, which represents the overall classification performance is shown in Figure 3 while DT results 98.30% of accuracy.

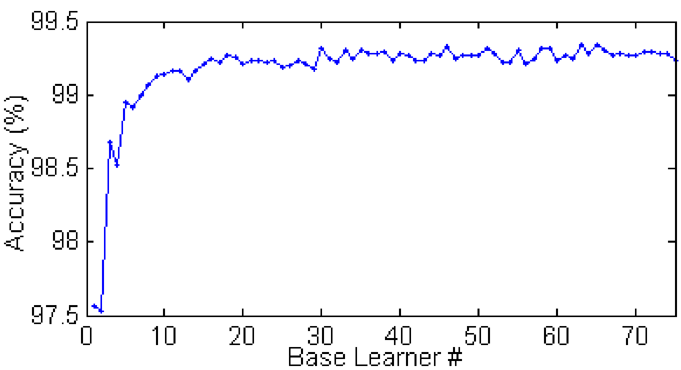

Figure 3. The accuracy graph of the BDT classifier

Referring to Figure 3, the proposed arrhythmia classifier using BDT has 99.34 % accuracy when 63 and 65 numbers of base learners are used.Moreover, the confusion matrices of BDT and DT classifiers are given in Table 2 to show the detailed effect of BDT on recognized heart beat samples.

Table 2 Confusion matrices of the classifiers

		Actual heart beat classes						
		Type	**N**	**L**	**R**	**A**	**P**	**PB**
Classifiers' output	**N**	DT	9720	39	34	16	16	0
		BDT	*9786*	*15*	*17*	*10*	*15*	*0*
	L	DT	27	1319	9	1	5	4
		BDT	*3*	*1356*	*2*	*1*	*4*	*3*
	R	DT	23	8	1425	1	4	0
		BDT	*3*	*1*	*1451*	*0*	*1*	*0*
	A	DT	15	0	1	169	2	0
		BDT	*3*	*0*	*0*	*180*	*1*	*0*
	P	DT	15	5	3	9	695	2
		BDT	*5*	*0*	*2*	*5*	*701*	*2*
	PB	DT	0	1	0	0	1	573
		BDT	*0*	*0*	*0*	*0*	*1*	*574*

The BDT classifier can recognize the misclassified false negative and false positive samples resulted by DT. Therefore, the resulted misclassification decreases, and the accuracy increases when BDT is used as an arrhythmia heart beat classifier.

4. CONCLUSION

The proposed morphological ECG signal feature extraction using RR interval (RR) and form factor (FF) based features are used to investigate the effect of bagged decision tree (BDT) classifier on arrhythmia heart beat classification. The proposed BDT has a higher distinguishing capability with the resulted 99.34% of accuracy when 63 and 65 base learners are used while single decision tree (DT) results 98.30 % of accuracy. Moreover, BDT with at least three base learners has more predictive performance than DT, and it can be successfully applied to ECG signal classification.

5. REFERENCES

[1] Yu S.N., Chen Y.H.: Selection of Higher Order Subband Features For ECG Beat Classification, 16th European Signal Processing Conference EUSIPCO 2008, Lausanne, Switzerland, 2008.

[2] Belhachat1F., IzeboudjenN.: Application Of Discret Wavelets Transform And Linear Prediction Coding For Analysis And Compression Of ECG Signals, 13th International Research/Expert Conference TMT 2009, Hammamet, Tunisia, 2009, pp.421-424.

[3] Belhachat1 F., Izeboudjen N.: Application Of A Probabilistic Neural Network For Classification Of Cardiac Arrhythmias, 13th International Research/Expert Conference TMT 2009, Hammamet, Tunisia, 2009, pp. 397-400.

[4] ChenS.W.: Complexity-Measure –Based Sequential Hypothesis Testing For Real –Time Detection Of Lethal Cardiac Arrhythmias, EURASIP Journal of Advances in Signal Processing , 2007, pp. 1-8.

[5] Rangayyan R.M.: Biomedical Signal Analysis: A Case-Study Approach,USA, Wiley-IEEE Press, 2001.

[6] Kim J., Shin H.S.,Shin K., Lee M.: Robust Algorithm For Arrhythmia Classification In ECG Using Extreme Learning Machine, Biomedical Engineering Online vol.8,2009.

[7] MIT-BIH Arrhythmia Database, http://www.physionet.org/physiobank/database/mitdb/, last accessed April 2012.

[8] Physionet ATM, http://www.physionet.org/cgi-bin/atm/ATM?database=mitdb&tool=plot_waveforms, last accessed April 2012.

[9] Breiman L.:Bagging predictors, Machine Learning Vol. 24,1996, pp.123-140.

[10] Jerez-AragonesJ.M., Gomez-RuizJ.A., Ramaos-JimenezG. , Munoz-PerezJ.:A Combined Neural Network And Decision Tree Model For Prognosis Of Breast Cancer Relapse, Artificial Intelligence in Medicine vol. 27, 2003, pp.45-63.