International Journal of Computer Applications (0975 – 8887)
Volume 57– No.17, November 2012

Performance Evaluation of Boosting Techniques for Cardiac Arrhythmia Prediction

V.S.R.Kumari
Research Scholar
Dept of ECE, Andhra university
Visakhapatnam India

P.Rajesh Kumar, PhD.
Professor in ECE
Assit.Principal, College of Engineeirng,Andhra University

ABSTRACT
Cardiac Arrhythmia is assessed using Electrocardiogram (ECG). Different types of arrhythmia are determined by accurate detection of beats leading to diagnosis of heart disease. Visual inspection of ECG for arrhythmia is tedious and time consuming process. With the advent of image processing techniques, automatic assessment of arrhythmia is widely studied. Various algorithms were developed for detection and classification of ECG signals. This paper investigates ECG classification method for arrhythmic beat classification based on RR interval. The methodology is based on extraction of RR interval of the beat using Symlet on ECG data. The extracted RR data are used as feature for classification. The beats are classified using boosting algorithm. MIT-BIH arrhythmia database was used for evaluating the classification efficiency.

General Terms
Pattern Recognition, Algorithms, Classification Accuracy.

Keywords
Electrocardiogram (ECG), Arrhythmia classification, MIT-BIH ECG data, RR interval, Symlet, Boosting.

1. INTRODUCTION
Electrocardiogram (ECG) records the heart's electrical activity through skin electrodes and it is used for diagnosis of heart diseases. ECG is a plot of voltage measured by leads against time. It is a non-invasive technique where the signal is measured on the human body's surface. Heart rate or rhythm disorder or morphological pattern change indicates cardiac arrhythmia, which can be detected by analysis of a recorded ECG waveform [1, 2]. High mortality rates due to heart diseases, has resulted in a need for proper detection and classification of ECG arrhythmias to ensure proper patient treatment. Visual inspection of ECG for arrhythmia is tedious and time consuming procedure. Recent advances in image processing techniques have led to research in automatic assessment of arrhythmia [3]. Recently, various algorithms were developed for automatic detection and classification of ECG signals.

Amplitude and duration of a P-QRS-T wave has information about the disease which afflicts the heart. ECG signals provide the following about the human heart [4]:

- Heart position and relative chamber size
- Impulse origin/propagation
- Heart rhythm and conduction disturbances
- Extent/location of myocardial ischemia
- Change in electrolyte concentration
- Effects of drugs on the heart.

ECGs do not provide data on cardiac contraction/ pumping function. The position and magnitudes of PR interval and segment, ST interval and segment, QRS interval and QT interval are used for the diagnoses of cardiac diseases [5]. Figure 1 shows a normal ECG pattern and its various components.

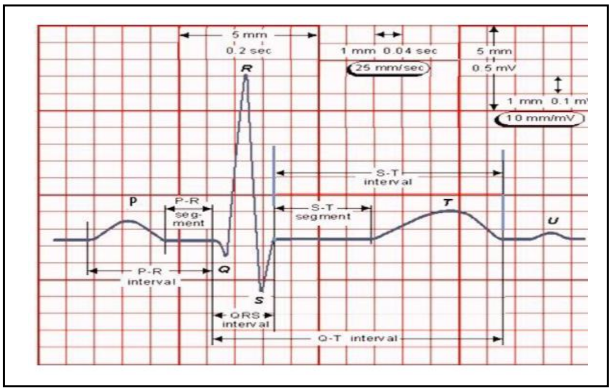

Fig 1: Normal ECG Signal and its various components

ECG features are extracted in time and frequency domains or represented as statistical measures. Though these methods provide impressive results in a few classification tasks, they generally fail to show discrimination power in all ECG beats. Wavelet transformation (WT) is another method that represents signals in various translations and scales [6]. Approaches were developed to classify cardiac arrhythmias based on ECG signals, but they are weak and not totally accurate and thus, ECG features are extracted to classify cardiac arrhythmias. Many Machine learning and data mining methods were sought to improve the accuracy in ECG arrhythmia detection.

Automated classification of heartbeats has been previously studied by other researchers [7–12] based on a variety of features representing the ECG and different classification methods. Features include frequency-based features [7], ECG morphology [8, 9], heartbeat interval features [8 - 11]. Classification methods used include backpropagation neural networks [8–10], self-organizing maps with learning vector quantization [11], and self-organizing networks [12].

In this paper, ECG classification method for arrhythmic beat classification based on RR interval is investigated. The methodology is based on extraction of RR interval of the beat using Symlet conversion on ECG data. The extracted RR data are used as feature for classification. The beats are classified using boosting algorithm. MIT-BIH arrhythmia database was used for evaluating the classification efficiency. The rest of the paper is organized as follows: section 2 reviews some related work available in the literature, section 3 details the methodology. Section 4 gives the results and discussion, and section 5 concludes the paper.

2. RELATED WORK

For arrhythmic beat classification and arrhythmic episode detection and classification employing only the RR-interval signal extracted from ECG recordings, M.G. Tsipouras et al., [13] proposed a knowledge-based method. In the arrhythmic beat classification algorithm, a three RR-interval sliding window is used. For four categories of beats: normal, premature ventricular contractions, ventricular flutter/fibrillation and 28 heart block, the classification is performed. An input used is beat classification of a knowledge-based deterministic automation in order to obtain classification and arrhythmic episode detection. There are six rhythm types classified: ventricular couplet, ventricular bigeminy, ventricular tachycardia, ventricular trigeminy, ventricular flutter/fibrillation and 28 heart block. Using the MIT-BIH arrhythmia database, the proposed method is evaluated that accomplishes 94% accuracy for arrhythmic episode detection and classification and 98% accuracy for arrhythmic beat classification. As the RR-interval signal is used for arrhythmia beat and episode classification, the proposed method is more advantageous when compared to other complicated techniques.

Khoureich Ka [14] proposed an ECG beat classification mechanism based on waveform similarity and RR interval. Six types of heart beats (normal beat, atrial premature beat, paced beat, premature ventricular beat, left bundle branch block beat and right bundle branch block beat) was classified using the proposed method. The ECG signal is denoised using wavelet transform based techniques. RR intervals extracted were used as feature. A training database of annotated beats is compiled for the classifier which is used for waveform comparison of unknown beats. Evaluations were conducted using 46 records in the MIT/BIH arrhythmia database; the proposed method achieved a classification rate of 97.52%.

Bashir, et al., [15] proposed a nested ensemble technique for classifying cardiac arrhythmia in real time. The proposed method includes manipulation of the training dataset and selection of features. The training dataset is manipulated for learning of the classifier by updating training data, and the selection of features to improve the accuracy and performance during classification. Experimental results demonstrated the need of considering all the ECG features for evaluation and good accuracy achieved by the proposed model.

3. METHODOLOGY

3.1 Symlet Wavelet

Wavelets are waveforms bound in both frequency and time. Wavelet analysis splits the signal into shifted and scaled versions of the original (or mother) wavelet. The Continuous Wavelet Transform (CWT) is given by the wavelet function ψ by adding all time of the signal multiplied by scaled, shifted versions. Mathematically the continuous wavelet is defined by

$$C(scale, position) = \int_{-\infty}^{\infty} f(t)\psi(scale, position, t)\,dt$$

Many wavelet coefficients C, which are a function of scale and position, are resulted due to the CWT. The constituent wavelets of the original signal are obtained by multiplying each coefficient by the applicable scaled and shifted wavelet.

Daubechies proposed the symlets which are nearly symmetrical wavelets. The symlets are modifications of the db family [16]. The two wavelet families are similar with the difference of db wavelets have maximal phase whereas symlets have minimal phase. Symlets are compactly supported wavelets with slightest asymmetry. The wavelet coefficient for symlet can be any positive even number and highest number of vanishing moments for a given support width.

3.2 Boosting

In boosting, set of *k* classifiers is iteratively learned and each training tuple are assigned weights [17]. The process starts with learning of a classifier *Mi*, and its weights are updated. This aids the succeeding classifier, *Mi+1*, to assign more weightage to the misclassified training tuples of *Mi*. After all the k classifiers are learned iteratively, the final boosted classifier, *M**, sums the votes of each individual classifier.

The Boosting allots higher weight to the classifier's vote which have a lower classifier's error rate. Thus, accurate classifiers are given higher weightage. The weight of classifier *Mi*'s vote is computed as follows:

$$\log \frac{1 - error(M_i)}{error(M_i)}$$

The class of a tuple is returned based on the highest sum of the votes.

3.2.1 Boosting with decision stump

A decision stump consists of a decision tree with only one split. Decision stump on its own is a weak learner but AdaBoost algorithm combines those weak learners into an accurate classifier [18]. The decision stumps combines to form a committee. The base classifiers are built on weighted examples and the committee takes decisions based on majority vote. At the start of the process, all the examples are assigned equal weights. In the next round, the weights are increased for the misclassified examples and decreased for the correctly classified examples by the first decision stump. This procedure is repeated for a defined number of stumps to be created. Each member in the committee has its own specialty due to its special training. Thus, the committee due to its diversity performs better than any single member.

3.2.2 Boosting with REP tree

Reduced Error Pruning (REP) tree

In decision tree learning algorithm [19], the tree that is created is the best fit on the training data. Generally, the trees built are most likely overfits the training samples. The trees are pruned to reduce its dependency on the training data and also to generalize so that it fits other examples. Reduced Error Pruning (REP) trees are a method to prune decision trees. REP is an easy and effective pruning method. It tends to over prune the tree. The major disadvantage of REP trees is that a separate pruning dataset is required. Though, with a large training dataset or with boosting, REP is extremely powerful. The REP pruning method replaces a sub tree by a leaf which represents the majority of all examples. But modifications are done only if it reduces the error either by having equal or lower number of misclassifications.

3.2.3 Boosting with J48

J48 algorithm is an implementation of the C4.5 decision tree learner. J48 uses greedy algorithm to prompt decision trees for classification. Unseen data are classified based on the decision-tree built using the training dataset. The nodes in J48 generated decision trees, evaluate the significance of each feature, e.g., heart rate.

The ECG features are classified to the appropriate class of arrhythmia by following a path from the root to the leaves of the tree, resulting in a decision about the class. The decision trees are built in a top-down method, choosing the most suitable attribute at each level. Features are evaluated using an information-theoretic measure. This measure gives the "classification power" of each feature. On selection of the feature, the training dataset is split into subsets conforming to the values of the selected feature. This process is repeated for each subset, till majority of the instances in each subset belong to a single class.

4. EXPERIMENTAL SETUP AND RESULTS

The proposed method is evaluated from a dataset created from MIT-BIH arrhythmia database. The dataset used for evaluation consists of 165 instance; 55 events each of Right bunch bundle block, Left bunch bundle block and Normal RR interval. Continuous wavelet transforms using symlet2 filters applied. Figures 1 – 3 show the output of the event on application of Continuous wavelet transforms using symlet2.

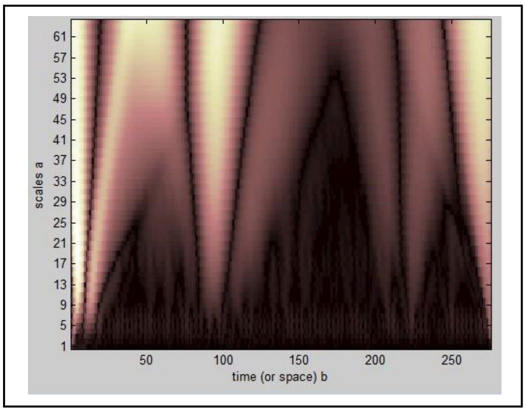

Figure 1: Left bunch bundle block output on applying Continuous wavelet transforms using symlet2

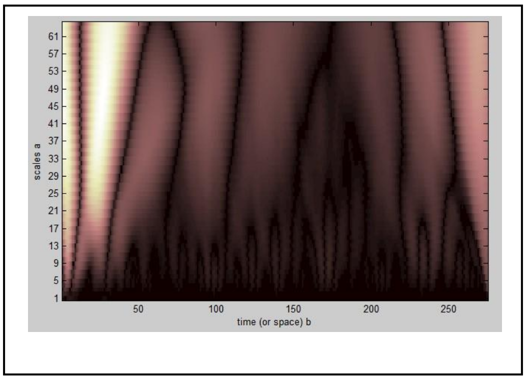

Figure 2: Right bunch bundle block output on applying Continuous wavelet transforms using symlet2

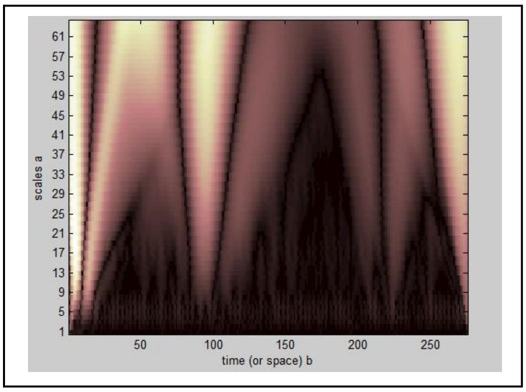

Figure 3: Normal RR interval output on applying Continuous wavelet transforms using symlet2.

5. RESULTS AND DISCUSSION

The instances are classified using Boosting with decision stump, boosting with J48 and boosting with REP tree. The experimental results for the classification accuracy are given in the following Tables and Figures. Table 1 tabulates the summary of the results for different techniques. Figure 4 shows the plotted classification accuracy. Figure 5 shows the Root mean squared Error.

Table 1. Summary of the Results

	Boosting with decision stump	**Boosting with J48**	**Boosting with REP tree**
Classification accuracy	92.125%	91.52%	91.52%
Root mean squared error	0.211 (21.1%)	0.1947 (19.47%)	0.1891 (18.91%)

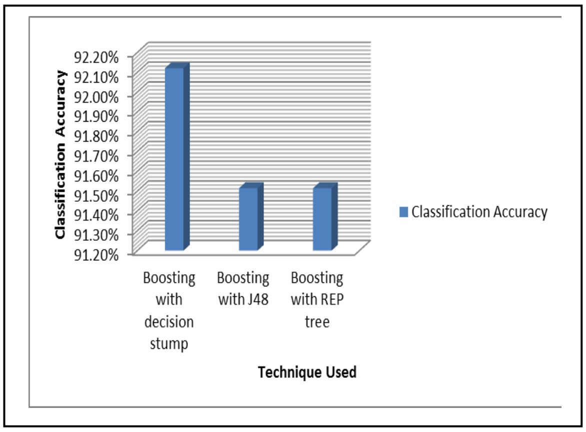

Figure 4: Classification Accuracy

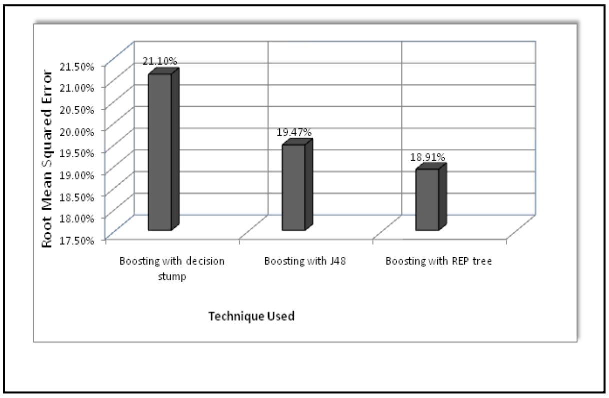

Figure 5: Root Mean Squared Error

Table 2 tabulates the precision, recall and f-Measure. The precision and recall is computed as

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in the Database}}$$

$$f \text{ measure} = 2\text{x}\,\frac{\text{precision x recall}}{\text{precision + recall}}$$

For a good classification system the values of precision and recall should be high which is seen in the proposed technique.

Table 2 Precision, recall and f Measure

	Precision	Recall	F-Measure
Boosting with decision stump	0.911	0.921	0.916
Boosting with J48	0.91	0.915	0.912
Boosting with REP tree	0.915	0.915	0.914

6. CONCLUSION

In this paper, ECG classification method for arrhythmic beat classification based on RR interval is investigated. The methodology is based on extraction of RR interval of the beat using Symlet conversion on ECG data. The extracted RR data is used as feature for classification. The beats are classified using boosting algorithm. MIT-BIH arrhythmia database was used for evaluating the classification efficiency. The instances are classified as Right bunch bundle block, Left bunch bundle block and Normal RR interval. The instances are classified using Boosting with decision stump, boosting with J48 and boosting with REP tree. The classification accuracy of 92.12% was obtained.

7. REFERENCES

[1] K. Polat and S. Günes, Detection of ECG arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine, Applied Mathematics and Computation, pp. 898–906, 2007.

[2] K. Minami, H. Nakajima, and T. Toyoshima, “Real-Time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network,” IEEE Trans. Biomed. Eng., vol. 46, no. 2, pp. 179–185, Feb. 1999.

[3] S.C. Saxena, V. Kumar, and S.T. Hamde, Feature extraction from ECG signals using wavelet transforms for disease diagnostics, Int. J. Syst. Sci. 33 (13), pp.1073–1085, 2002.

[4] S. Osowski, T. H. Linh, and T. Markiewicz, “Support vector machinebased expert system for reliable heart beat recognition,” IEEE Trans. Biomed. Eng., vol. 51, no. 4, pp. 582–589, Apr. 2004.

[5] P. E. McSharry, G.D. Clifford, L. Terassenko and L.A. Smith, “A dynamical model for generating synthetic electrocardiogram signals,” IEEE Trans. Biomedical Engineering, Vol. 50, Issue 3, pp.289–294, Mar. 2003.

[6] Al-Fahoum, A.S., Howitt, I., 1999,” Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias”, Med. Biol. Eng. Comput. (37), 566–573.

[7] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, “Comparing wavelet transforms for recognizing cardiac patterns,” IEEE Eng. Med. Biol. Mag., vol. 14, pp. 167–173, Mar.–Apr. 1995.

[8] T. H. Yeap, F. Johnson, and M. Rachniowski, “ECG beat classification by a neural network,” in Proc. Annu. Int. Conf. IEEE Engineering Medicine and Biology Soc., 1990, pp. 1457–1458.

[9] Y. H. Hu,W. J. Tompkins, J. L. Urrusti, and V. X. Afonso, “Applications of artificial neural networks for ECG signal detection and classification,” J. Electrocardiol., vol. 26, pp. 66–73, 1993.

[10] S. Osowski and T. L. Linh, “ECG beat recognition using fuzzy hybrid neural network,” IEEE Trans. Biomed. Eng., vol. 48, pp. 1265–1271, Nov. 2001.

[11] Y. H. Hu, S. Palreddy, and W. J. Tompkins, “A patient-adaptable ECG beat classifier using a mixture of experts approach,” IEEE Trans. Biomed. Eng., vol. 44, pp. 891–900, Sept. 1997.

[12] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, “Clustering ECG complexes using hermite functions and self-organizing maps,” IEEE Trans. Biomed. Eng., vol. 47, pp. 838–848, July 2000.

[13] M. G. Tsipouras, D. I. Fotiadis, and D. Sideris, “An arrhythmia classification system based on the RR interval signal,” Artif. Intell. Med., vol. 33, pp. 237–250, 2005.

[14] Ahmad Khoureich Ka, “ECG Beats Classification using Waveform similarity and RR Interval”, Arxiv preprint arXiv: 1101.1836, 2011.

[15] Mohamed Ezzeldin A. Bashir , Makki Akasha, D. G. Lee, Min Yi, K. H. Ryu , E. J. Bae, M. Cho, and C. Yoo, “Nested Ensemble Technique for Excellence Real Time Cardiac Health Monitoring”, BioComp, lasvegas USA, 2010.

[16] Vinay U. Kale & Nikkoo N.Khalsa "Performance Evaluation of Various Wavelets for Image Compression of Natural and Artificial Images" International Journal of Computer Science & Communication,Vol. 1, No. 1, January-June 2010, pp. 179-184.

[17] Freund Y, Schapire R. A decision-theoretical generalization of on-line learning and an application to boosting. J Computer Syst Sci 1997;55:119-139.

[18] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Stat 2000;28:337-407.

[19] Mitchell, T. M. (1997). Machine Learning. McGraw-Hill, New York.

[20] Quinlan, J. R. (1987). Simplifying decision trees. In B. Gaines and J. Boose, editors, Knowledge Acquisition for Knowledge-Based Systems, pages 239–252. Academic Press, London.