Analysis of Multi-Lead ECG Signals using Decision Tree Algorithms

Smita L. Kasar
Assistant Professor
MGM's Jawaharlal Nehru Engineering College,
Aurangabad.

Madhuri S. Joshi
Professor
MGM's Jawaharlal Nehru Engineering College,
Aurangabad.

ABSTRACT

Electrocardiogram (ECG) is a signal with unique, valuable information about the functional aspects of the heart with respect to time. The automatic analysis of ECG signals is an important application since the early detection of heart diseases/abnormalities can prolong life and enhance the quality of living through appropriate treatment. The ECG is collected using a number of electrodes placed in different positions on the body. Multi-lead ECGs acquired simultaneously helps in better diagnosis of heart diseases. This paper focuses on classification of healthy and Myocardial infarction signals. The identification of acute myocardial infarction with symptoms of Ischemia is critical to delivering appropriate medical care. In this paper decision tree based classifiers are implemented for the classification of ECG signals. The signals were analyzed for 34 normal and 33 myocardial infarction patients in the database PTB from the domain Physionet.org. The classifiers, J48 and Classification and Regression Trees (CART) are compared with respect to accuracy measures. The J48 classifier performs better with the correct classification rate of 98% and 0.9 Kappa statistics.

General Terms

Signal Processing, Classification, Data Mining.

Keywords

ECG, PCA, MI, J48, CART

1. INTRODUCTION

The ECG is a recording of the electric potential generated by the electrical activity of the heart. The ECG thus represents the extracellular electrical behavior of the cardiac muscle tissue [1]. Different electrical phases of a cardiac cycle are described and a summation in time and space of the action potentials are represented. The state of the cardiac health is generally reflected in the shape of ECG waveform and the heart rate. In the Time domain the ECG signal is identified by different waves viz., P, Q, R, S, T. The ECG waveform is as shown in fig. 1.1. The P wave represents atrial depolarization. The Q, R & S waves together make up a complex, QRS complex, which represents ventricular depolarization and T wave corresponding to the period of ventricular repolarisation. The interval between S wave and the beginning of the T wave is called the ST segment. In some ECGs an extra wave can be seen at the end of the T-wave, called as U wave. Its origin is uncertain, though it may represent repolarisation of the papillary muscles. The standard ECG has 12 leads which include 3-bipolar leads, 3-augmented unipolar leads and 3-chest (precordial) leads.
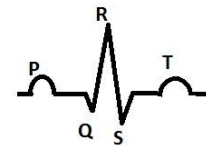

Fig. 1.1 ECG Wave in Time Domain

The ECGs were classified by the investigators into one of three categories, based on the potential clinical importance of the discrepancies. Category I – No discrepancy or ECG with minor discrepancy, Category II – ECGs of indeterminate clinical importance, Category III – ECGs with probable clinical importance. Sinus tachycardia, Sinus Bradycardia, Atrial Enlargement etc. are in the category-I whereas Right bundle branch block, Left bundle branch block, Left ventricular Hypertrophy are in category II. Category III includes Ischemia, Acute Myocardial infarction, Atrial Fibrillation, Atrial Flutter etc. [10]. The American heart association defines myocardial infarction as 'the damaging or death of an area of the heart muscle (myocardium) resulting from blocked blood supply to the area; the medical term for a heart attack'. Myocardial infarction is a major cause of death and disability worldwide.

2. LITERATURE SURVEY

ECG signal preprocessing is the first step towards the classification of heart diseases. ECG is contaminated by various kinds of noises like Power line Interference, Electrode contact noise, Motion artifact, Muscle contractions, Baseline Wander etc. The noise reduction is the important factor since the signal should be accurately represented for further analysis. There are many different methods in the literature for denoising the signals. Feature extraction is the determination of a feature or a feature vector from a pattern vector. In order to make pattern processing problems solvable one needs to convert patterns into features, which become condensed representations of patterns, ideally containing only salient information. Feature extraction methods could be based on either calculating statistical characteristics or producing syntactic descriptions. Various techniques and transformations proposed earlier in the literature for extracting features from an ECG signal and a comparative study of various methods proposed by researchers in extracting the features from ECG is presented [21].

The area of data mining arose over the last decade to address the issue of classification using the predictive and descriptive models. The paper [4] presents a critique of literature concerned with Data mining techniques and its applications. Several other classification techniques based on predictive models have been found in the literature.

Classification of ECGs into normal and arrhythmic is implemented by artificial neural networks and Linear Discriminant Analysis [2]. Independent Component Analysis and wavelet transform are utilized to get an ensemble feature composed of ICA-based features and the QRS complex width feature [3]. Principal component analysis is practiced to reduce dimensionality and for feature extraction of the ECG data prior to or at times after performing ICA in special circumstances [6]. In [7] Ant Colony Optimization based clustering analysis of ECG arrhythmias taken from the MIT–BIH Arrhythmia Database is aimed. Both time domain and discrete wavelet transform based frequency domain characteristics are utilized in the analysis. Since the number of wavelet coefficients is a huge amount as compared to the time domain parameters, Principal Component Analysis (PCA) based compression is applied to them in order to decrease their number to the number of time domain features. In [8] a comprehensive survey of the nonlinear dynamics of ECG signals is introduced. The correlation dimension and largest Lyapunov exponent are used to model the chaotic nature of five different classes of ECG signals. The paper [9] uses automatic extraction of both time interval and morphological features to classify ECGs into normal and arrhythmic. Classification is implemented by Artificial neural networks and Linear Discriminant Analysis.

3. METHODOLOGY

3.1 Database
Physikalisch-Technische Bundesanstalt (PTB), the National Metrology Institute of Germany, has provided this compilation of digitized ECGs for research, algorithmic benchmarking or teaching purposes to the users of PhysioNet. The ECGs were collected from healthy volunteers and patients with different heart diseases by Professor Michael Oeff, M.D., at the Department of Cardiology of University Clinic Benjamin Franklin in Berlin, Germany. The database comprises 549 records from 290. Each record includes 15 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the 3 Frank lead ECGs (vx, vy, vz). Each signal is digitized at 1000 samples per second, with 16 bit resolution over a range of $\pm$ 16.384 mV[23].

3.2 Classification Model
The figure 3.1 represents the detailed architecture of the classification model

Data Preprocessing and Feature Extraction: The 12 lead simultaneously acquired ECG signal is read as input to the scheme. Every lead signal is passed through the notch filter to remove the noise components from the signal. Then the signal is processed with the help of principal component analysis. Principal component analysis (PCA) is a statistical technique whose purpose is to condense the information of a large set of correlated variables into a few variables termed as principal components, while not throwing overboard the variability present in the data set. The principal components are derived as a linear combination of the variables of the data set, with weights chosen so that the principal components become mutually uncorrelated. Each component contains new information about the data set, and is ordered so that the first few components account for most of the variability. The features such as P amplitude, Q amplitude, R amplitude, S amplitude, T amplitude, Q duration is extracted from the signal. These features are used to calculate few other features like the ratio of Q and R amplitudes. Then, with the help of Scoring pattern for myocardial infarction (Okajima et al. 1990), the score is counted. The score is included as a

characteristic in the feature vector for the categorization of the ECG signals. The feature vectors are used for training and testing using decision tree algorithms for the prediction.
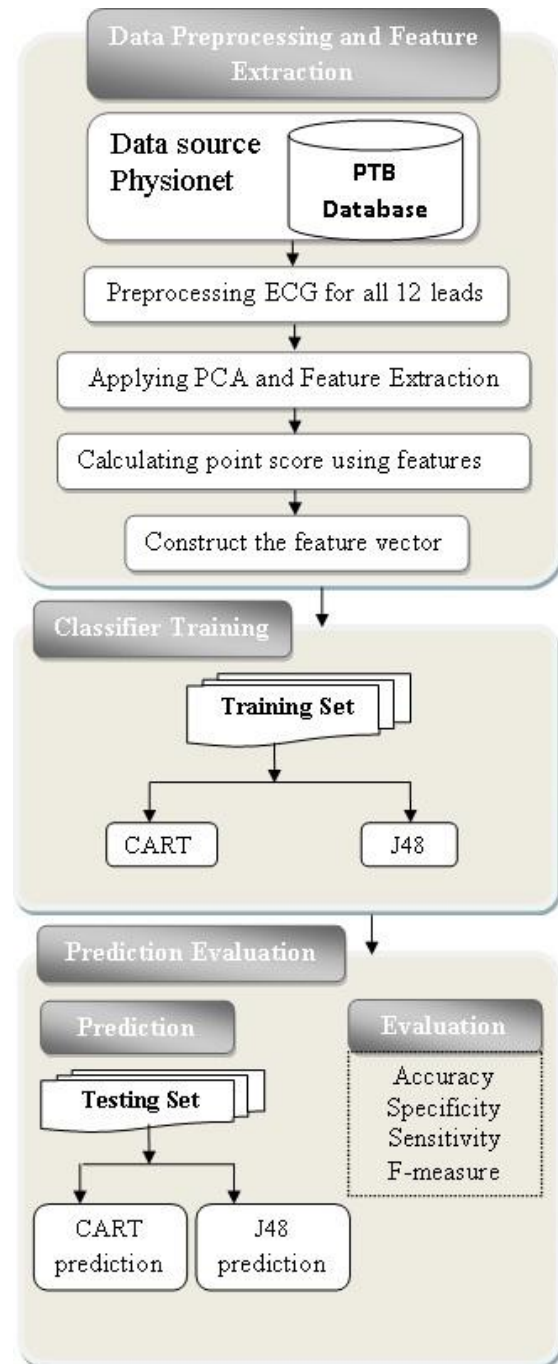

Fig. 1: Detail Classification Model

Classifier Training: The preprocessed feature vector data are passed to the classifier training module. The decision tree is a predictive modeling, data mining technique used in classification, clustering and prediction tasks. This approach is most useful in classification problems. The different algorithms applied for comparison are Classification and Regression technique (CART) and J48. J48 decision tree is simple and easy to implement and this method can also handle high dimensional data.This algorithm uses pruning method to build the tree. Then the J48 Algorithm recursively classifies the data and gives maximum accuracy. CART refers to both,

classification and regression tree. This algorithm generates binary decision tree.

Prediction and Evaluation: The class labels for new test data are predicted. The classification involves diagnosis of two classes, Normal and Myocardial Infarction. The performance of the model is analyzed using different evaluation measures such as accuracy, sensitivity, specificity and F-measure.

4. RESULTS

The Accuracy of the predictive model is computed based on the precision, recall values of classification matrix. The classifier is trained to classify the data set as either Normal signal or MI signal. Table 3.2 shows the detail performance of the implemented classifiers using the CART and J48. The J48 algorithm classifies 66 instances correctly out of 67 cases. But with the same instances CART classifies 62 instances correctly.

Table 3.1: Performance of the Classifiers

Evaluation Criteria	Classifiers	
	CART	J48
Correctly classified instances	62	66
Incorrectly classified instances	05	1
Accuracy (%)	92.5	98.5

Predictive analytics, a table of confusion, is a table with two rows and two columns that report the number of *false positives(FP), false negatives(FN), true positives(TP) and true negatives(TN).* This allows more detailed analysis than the mere proportion of correct guesses. Table 3.2 shows the confusion matrix for the classifiers. The J48 algorithm performs better as compared to the CART algorithm.

Table 3.2: Confusion Matrix

Classifier	Normal	MI
CART	30	4
	1	32
J48	34	0
	1	32

The kappa statistic measures the agreement of prediction with the true class. The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. The MAE is the average over the verification sample of the absolute values of the differences between forecasts and the corresponding observation. The MAE is a linear score, which means that all the individual differences are weighted equally in the average. The RMSE is the difference between forecast and corresponding observed values are each squared and then averaged over the sample. The MAE and the RMSE can be applied together to diagnose the variation in the errors in a set of forecasts. The RMSE will always be larger or equal to the MAE; the greater conflict between them, the greater the variation in the individual errors in the sample. The comparison of all the simulation error parameters is shown in Table 3.3.

Table 3.3. Training and simulation error

Evaluation Criteria	Classifiers	
	CART	J48
Kappa statistics (KS)	0.85	0.97
Mean absolute error(MAE)	0.13	0.02
Root mean squared error (RMSE)	0.25	0.12
Relative absolute error(RAE)	27.01%	5.77%
Root relative squared error (RRSE)	51.97%	24.03%

Table 3.4. Comparison of Evaluation measures

Classifier	TP rate	FP rate	Precision	Recall	Class
CART	0.88	0.03	0.96	0.882	Normal
	0.97	0.118	0.889	0.97	MI
J48	1	0.03	0.971	1	Normal
	0.97	0	1	0.971	MI

Table 3.4 depicts the comparison of different valuation measures. The precision is given by

$$Precision = \frac{TP}{TP + FP}$$

Recall is fraction of relevant instances that are retrieved.

$$Recall = \frac{TP}{TP + FN}$$

The different parameters like precision , recall give better and promising results for J48 algorithm. The parameter F-measure is improved from 0.92 to 0.98 in J48 implementation for both normal and MI signals.

5. CONCLUSION

The rate and risk of Cardiovascular diseases have drastically increased these days. This has contributed to the need for automatic classification of the heart diseases. The major advantage of Electrocardiogram signal is that it is inexpensive and non invasive. In the present model the characteristics are extracted from ECG signal and feature vectors are developed and tested using simple CART and J48 decision tree algorithms. The present work evaluates the CART and J48 decision algorithms using Accuracy, Kappa statistics and different evaluation measures. The classification accuracy is 92.5% using CART but enhanced to 98.5% using J48 algorithm. Also the kappa statistics is improved from 0.85 to 0.97. The ECG signals can also be assessed for different types of infarction.

6. REFERENCES

[1] Yue-Der Lin, Yu Hen Hu, Power-Line Interference Detection and Suppression in ECG Signal Processing, IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, January 2008

[2] Reza Sameni, Mohammad, B. Shamsollahi, Christian Jutten, Gari D. Clifford, A Nonlinear Bayesian Filtering Framework for ECG Denoising, IEEE Transactions on Biomedical Engineering, vol. 54, no. 12, Dec 2007

[3] Zhao Yong, Hong Wenxue, Xu Yonghong, Cui Jianxin, ECG Beats Classification Based on Ensemble Feature Composed of Independent Components and QRS Complex Width, IEEE International Conference on

Computer Science and Software Engineering 2008, pp 868-871

[4] Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao, Data mining techniques and applications-a decade review from 2000 to 2011, Expert systems with applications 39 (2012),Elsevier, pg.11303-11311

[5] T. M. Nazmy, El-messiry, Al-bokhity, Adaptive Neuro-Fuzzy Inference System For Classification of ECG Signals, Journal of Theoretical and Applied Information Technology – 2009

[6] M.P.S. Chawla, PCA and ICA processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison, Applied Soft Computing 11, 2011, pp 2216–2226

[7] Mehmet Korürek , Ali Nizam, Clustering MIT–BIH arrhythmias with Ant Colony Optimization using time domain and PCA compressed wavelet coefficients, Digital Signal Processing, Elsevier, 20 (2010) pp 1050–1060

[8] Mohamed I. Owis, Ahmed H. Abou-Zied, Abou-Bakr M. Youssef , Yasser M. Kadah, Robust Feature Extraction From Ecg Signals Based On Nonlinear Dynamical Modeling, Proceedings of the 23rd Annual EMBS International Conference, October 25-28,2001 Istanbul, Turkey

[9] C Alexakis, HO Nyongesa, R Saatchi, ND Harris, C Davies, C Emery, RH Ireland, SR Heller, Feature Extraction and Classification of Electrocardiogram (ECG) Signals Related to Hypoglycaemia, IEEE Computers in Cardiology 2003, pp 537−540.

[10] E. R. Snoey, B. Housset, P. Guyon, S. Elhaddad, J. Valty, P. Hericod, Analysis of Emergency Department interpretation of Electrocardiogram, Journal of Accident and Emergency Medicine,11, 1994, pp 149-153.

[11] Gary M, Friesen, Thomas C. Jannett, Manal Afify Jadallah, Standford L. Yates, Stephen R.Quint, H.Troy N Nagle, A Comparision of the Noise Sensitivity of Nine QRS Detection Algorithms, IEEE Transactions on Biomedical Engineering,Vol, 37, No. 1, March 1990

[12] Yüksel Özbay , Bekir Karlik,A Recognition of ECG Arrhythmias Using Artificial Neural Networks, Proceedings – 23rd Annual Conference – IEEE/EMBS , Istanbul, Turkey,2001

[13] Fabian Vargas, Djones Lettnin, Maria Cristina, Felippetto de Castro, Marcello Macarthy, Electrocardiogram Pattern Recognition by Means of MLP Network and PCA: A Case Study on Equal Amount of Input Signal Types, IEEE Proceedings of the VII Brazilian Symposium on Neural Networks (SBRN'02) 2002

[14] Inan Gulera, Elif Derya Ubeyl, ECGbeat classifier designed by combined neural network model , ScienceDirect,Pattern Recognition 38 (2005) , Elsevier, pp 199 – 208

[15] P.Sasikala, Dr. R.S.D. Wahida Banu, Extraction of P wave and T wave in Electrocardiogram using Wavelet Transform ,(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (1) , 2010, pp 489-493

[16] Asie Bakhshipour, Mohammad Pooyan, Hojat Mohammadnejad, Alireza Fallahi, Myocardial Ischemia Detection with ECG Analysis, Using Wavelet Transform and Support Vector Machines, Proceedings of the 17th Iranian Conference of Biomedical Engineering (ICBME2010), 3-4 November 2010

[17] Farah Nur Atiqah , Francis Abdullah, Fazly Salleh Abas, Rosli Besar, ECG Classification using Wavelet Transform and Discriminant Analysis, International Conference on Biomedical Engineering ,27-28 February 2012

[18] John R. Hampton, 'The ECG made Easy", Churchill Livingstone Elsevier, ISBN 978-0-443-06826-3

[19] Raphael Twerenbold, Roger Abächerli, Tobias Reichlin, Stefan Osswald, Christian Müller Early Diagnosis of Acute Myocardial Infarction by ST-Segment Deviation Score, IEEE Computing in Cardiology 2012, pp 657-659.

[20] Md. Ashfanoor Kabir, Celia Shahnaz, 'Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains', Biomedical Signal Processing and Control,vol.7 (5),Sept 2012, pp. 481-489.

[21] S. Karpagachelvi, Dr.M.Arthanari, M.Sivakumar,ECG Feature Extraction Techniques - A Survey Approach, (IJCSIS) International Journal of Computer Science and information Security, Vol. 8, No. 1, April 2010

[22] Okajima, M., Okamoto N., Yokoi M., Iwatsuka T., Ohsawa, Methodology of ECG interpretation in the Nagoya Program, Methods in Information Medicine,29,1990,pp 341-345.

[23] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220