Chapter 13

Random Forest Classifier Based ECG Arrhythmia Classification

V. Mahesh  
PSG College of Technology, India

A. Kandaswamy  
PSG College of Technology, India

C. Vimal  
PSG College of Technology, India

B. Sathish  
PSG College of Technology, India

ABSTRACT

Heart Rate Variability (HRV) analysis is a non-invasive tool for assessing the autonomic nervous system and for arrhythmia detection and classification. This paper presents a Random Forest classifier based diagnostic system for detecting cardiac arrhythmias using ECG data. The authors use features extracted from ECG signals using HRV analysis and DWT for classification. The experimental results indicate that a prediction accuracy of more than 98% can be obtained using the proposed method. This system can be further improved and fine-tuned for practical applications.

INTRODUCTION

Electrocardiography is a commonly used, non-invasive procedure for recording the electrical activity of the heart. The record, which is called an electrocardiogram (ECG or EKG), shows the series of waves that relate to the electrical impulses which occur during each beat of the heart. The information present in ECG characteristic wave peaks and time intervals between them are important. The waves in a normal record are named P, Q, R, S, and T. Any abnormality in shape and variation in the time interval of these waves is considered as arrhythmia. Detection of abnormal ECG signals is a critical step in administering aid to patients. Often, patients are hooked up to cardiac monitors in hospital continuously. This requires continuous monitoring by the physicians. Due to the large
Random Forest Classifier Based ECG Arrhythmia Classification

number of patients in intensive care units and the need for continuous observation of them, several methods for automated arrhythmia detection have been developed in the past few decades to attempt simplify the monitoring task. Notably Palreddy et al. (1997) employed a multiple-classifier architecture composed of Self Organizing Maps (SOM) and Learning Vector Quantization (LVQ) to classify premature ventricular contraction (PVC) beats and the non-PVC beats. Acharya et al. (2004) used a fuzzy equivalence classifier to classify four cardiac arrhythmias using certain parameters extracted from heart rate signals. Babak Mohammadzadeh-Asl et al. (2006) used both linear and non-linear parameter extracted from heart rate signals with multilayer feed forward neural networks to classify five types of arrhythmias. Chazal et al. (2006) proposed a method for automatic classification of heartbeats using ECG morphology, heartbeat interval features and RR intervals to discriminate five different beat types. Dingfie et al. (2002) classified six arrhythmias using autoregressive modeling and Generalized Linear Model (GLM). Linh et al. (2006) selected the Hermite Function Expansion as the feature extraction method to represent the QRS complex. They proposed a fuzzy neural network where Hermite coefficients served as the features to classify seven different types of arrhythmias. Kannathal et al. (2005) used three non-linear parameters as inputs to the proposed ANF classifier for classification of ten different types of arrhythmias. Kadbi et al. (2006) used wavelet parameters along with RR interval and Form Factor as inputs to an ANN classifier to discriminate ten different arrhythmias. There is always a growing need to develop new detection schemes with a high level of accuracy, or equivalently, with low false-positive and false-negative results, for them to be useful in practical applications. In this direction a new approach based on Random Forests classifier is presented in this paper. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. A random forest is a multi-way classifier consisting of a number of trees, with each tree grown using some form of randomization. The leaf nodes of each tree are labeled by estimates of the posterior distribution over the arrhythmia classes. Each internal node contains a test that best splits the space of data to be classified. An arrhythmia is classified by sending it down every tree and aggregating the leaf distributions reached.

MATERIALS

ECG data for the analysis and classification were obtained from the MIT-BIH arrhythmia database and signals from ML-II leads are used. This digital data has been sampled at a frequency of 360Hz and preprocessed to remove noise due to power-line interference, muscle tremors, spikes etc. The work presented in this paper focuses on several important arrhythmia types such as Paced beat (P), Atrial premature beat (A), Right bundle branch block (R), Left bundle branch block (L), Ventricular escape beat (E), Ventricular flutter wave (!), premature ventricular contraction (V), Fusion of ventricular and normal beat (F), Fusion of paced (f), Blocked Atrial Premature Beat (x) and the Normal beat segment (Normal). One minute segments of each beat type are extracted from these records for analysis. The number of segments of each type extracted from the database records is given in Table 1.

METHODOLOGY

Feature Extraction

Arrhythmia identification involves determination of several characteristic features of the ECG signal. The features extracted from the ECG signal vary from simple ones, like the duration and amplitude of ECG waves, to more complex ones like inter-