Nested Ensemble Technique for Excellence Real Time Cardiac Health Monitoring

Mohamed Ezzeldin A. Bashir¹, Makki Akasha¹, Dong Gyu Lee¹, Gyeong Min Yi¹, Keun Ho Ryu¹, Eun-jong Cha², Jang-Whan Bae³, Myeong Chan Cho³, and Chae Woo Yoo⁴

¹Database/Bioinformatics Laboratory, Chungbuk National University, Korea
²Dept. of Bio. Engineering School of Medicine, Chungbuk National University, Korea
³Dept. of Internal medicine, College of Medicine, Chungbuk National University, Korea
⁴Dept. computer Engineering, Soongsil university
{mohamed, makki, dglee, min9709,khryu}@dlab.chungbuk.ac.kr
ejcha@chungbuk.ac.kr, drcorazon@hanmail.net, mccho@cbnu.ac.kr, cwyoo@ssu.ac.kr

Abstract - Electrocardiogram (ECG) is a series of waves and deflections recording the cardiac’s (heart) electrical activity sensed by several electrodes. ECG signal utilized to extract very useful information about the functional status of the heart. Of particular interest are the systems designed for monitoring people outdoor and detecting abnormalities (arrhythmia) on the real time. In this paper, we propose a nested ensemble technique for real time arrhythmia classification. Its main components include manipulating the training dataset for learning the classifier by up-to-data training data, and manipulating the ECG features to select the proper adequate set to enhance the accuracy and excellence classification performance. Our experiment works demonstrated the necessity of including all the ECG features in cardiac health evaluation. Moreover prove the outstanding quality achieved by our model.

Keywords: Arrhythmia, Classification, Feature Selection, and Ensemble Technique.

1 Introduction

Heart activity is synchronized by electrical impulses across the cellular membrane, due to charge differences between the inside and the outside of muscle cells. These impulses cause a discontinuous atria and ventricle contraction, through which the blood is pumped all over the body. The heart’s electrical activity reflected and spread on the surface of the skin in the form of varying voltage waves that can be recorded and analyzed. Electrocardiogram (ECG) is a series of waves and deflections recording the cardiac’s (heart) electrical activity sensed by several electrodes, known as leads, monitor voltage changes between electrodes placed in different positions on the body. ECG signals, generated by sensing the current wave sequence of P wave to represent the Atrial depolarization, QRS complex to ventricular depolarization and T wave for ventricular repolarization related to each heart beat. Figure 1 depicts the basics shape of a healthy ECG heartbeat signal.

![ECG Figure](image-url)

Fig. 1 ECG for one normal heartbeat showing the amplitudes and time durations of P, QRS, and T waves.

ECG signals are very important medical instrument. That can be utilized by Clinicians to extract very useful information about the functional status of the heart. So as to detect heart arrhythmia which is the anomalous heart beat, mapped with different shape in ECG signal noticed by deflection on the P, QRS, and T waves, which acquired by some parameters. That judge against reference ones obtained through the average of normal ECG wave forms sampled from healthy people classified by age, sex, constitution and lifestyle. And then an enormous finding produced. ECG is essential to diagnose arrhythmias, as an indicator for cardiovascular disease, and straightforwardly monitors the patient cardiac health [1].

Cardiovascular disease is the leading cause of death along the world. It refers to various medical conditions that affect the heart. These conditions include coronary artery disease, heart attack, angina, heart failure, and sudden cardiac death. Which all of them cause’s death, or damage to a part of the heart muscle, because blood supply is severely reduced or
stopped [2]. Many studies have been conducted especially on the ECG detection and analysis in view of the fact that, it is the focal tool for diagnosing heart arrhythmias [3] [4].

Wire ECG monitoring in hospital such as in surgery theatres and intensive care are very crucial for saving people’s life. But this kind of monitoring is inadequacy for the coronary cardiac disease’s patients, who need following up in home and in the open air, those who need continues monitoring system to save their life. There for, there has been a great deal of interest in the systems used to provide real time ECG classification through intermediary local computer between the sensor and the control center [5].

It’s vital for the automated system to accurately detect and classify ECG signals very fast, to provide a useful means for tracing the heart health in the right time. The accuracy of the arrhythmias classification software is significance for the precise cardiac dysfunctions diagnosis based on all types of systems especially in the real time. The effectiveness of such systems is affected by several factors, including the ECG signals, the estimated ECG’s features and descriptors, the dataset used for learning purpose and the classification rule which applied. In this paper we propose a nested ensemble technique for real time arrhythmia classification. Its main components include manipulating the training dataset for learning the classifier by up-to-data training data, and manipulating the ECG features to select the proper adequate one to enhance the accuracy and excellence classification performance.

In the rest of this paper, we give a brief background of related work, the heart health monitoring problems and motivation, description of nested ensemble technique, experimental works, and finally the conclusion.

2 Related work

In the following section we will introduce some related work, including the most famous philosophies used to select the training dataset, incremental learning algorithms utilized to classify the arrhythmias on real time, and the techniques for solving the concept drift problem.

2.1 Training dataset

considering the two main types of training dataset, the local learning set which is a customized set to specific patient, and global which is built from large database [6] [7]. The later one is preferable to build a classifier model, while it is static which means the learning process takes place through specific set of data under specific circumstances. Therefore, the model generated will be very accurate in similar situation while it doesn’t in different cases. Which means it can be used to monitor the ECG in hospital such as in surgery theatres and intensive care. But this kind of monitoring is inadequacy for keeping an eye on the coronary cardiac disease patients, who need following up in home and in the open air, those who need continues monitoring system to save their life.

In contrast, in the case of first philosophy there will be a patient adaptable local learning set. In this sense, specific strategies are adopted for local learning in some arrhythmia monitoring. Obviously, the size of the training data is very big and it is satisfy the need of monitoring just specific arrhythmia of specific patient.

2.2 The classification models

A supervised training technique was used to build a model for classifying the ECG data. The classifier model maps the input features to the required output classes, using adjustable parameters specified during the training process. Several data mining techniques were used for this intention such like decision tree [8], support vector machine [9], neural networks [10], nearest neighbor [11], rule base classifier [12], and fuzzy adaptive [13]. These methods are general purpose and can be applied in any classification task. The judgment upon such techniques bases on the accuracy: the right description of the arrhythmia, effectiveness: the sensitivity to detect the abnormalities on the same time when it take place, efficiency: the speed by which the class of the arrhythmia is going to be specified, and the reliability of the classifier: how far doctors can trust on that model to judge future unseen ECG data. These factors are fluctuating from one to another method.

2.3 Concept drifts

The solutions to the problem of concept drift can be generally categorized into three groups: 1) Window-based 2) weight based and 3) ensemble classifiers. The window based approaches [14], [15] are detecting the training dataset within fixed or dynamic window sizes to make the classification model.

In the weight-based approaches there are distinctive weigh assigned to each training set based on certain factors such like the time for storing the data [16]. To some extend it looks similar to the widow based approach.

In the ensemble classifier, the classifier model is constructed after voting multiple classifiers [17]. The simple classifiers are builds for each block of data. According to the accuracy and other factors like time consumed to build the classifier, the weights are distributed and then voting is take place so as to construct the base one.

Ensemble classifier can be constructed by different four methods: 1) by manipulating the training set 2) manipulating the input features 3) manipulating the class labels and 4) manipulating learning algorithm [18]. In the first approach there is multiple training set will be revise. In the second one using sub set of features as an input for each training set,
while the third one is preferable when the number of classes is very high, and in the last one the classifier algorithm applied many times on training data.

3 Research problems and motivations

The following sections describe part of real time cardiac health monitoring issues related to the features set involved in classifying arrhythmias and the training dataset introduced to learn the classifier model.

3.1 ECG features

The literatures covered many works to detects and classify arrhythmias depending on features related to QRS complex mainly the R wave without introducing any other parameters like P, T waves [19] [20] [21]. QRS complex facilitates in detecting the RR interval and diagnosing many arrhythmias, such as normal heart beat, premature ventricular contractions, left and right bundled branch blocks, and paced beats. In contrast QRS complex alone is not capable to offer complete cardiac health monitoring which means, there are so many arrhythmias which couldn’t be detected without considering the P and T waves [22].

Employing other elements of ECG signal like the P wave in the assessment process, can give more indications and allow supplementary truthful judgment about the cardiac health. The other elements of the ECG, together with the QRS complex, for sure provide useful information such as to what type of arrhythmia the patient has. Some arrhythmias, though they may have a different cause, apparent themselves in similar ways on the ECG. Furthermore, taking into account the main two grouping of arrhythmias 1) the Ventricular, that occurs in the ventricles are recognized because of the abnormal QRS-morphology. And 2) the Supraventricular arrhythmias, which occur in the atrium however, can only be predetermined because they have an effect on the ventricular rhythm. For example, prematurity is used as a parameter to detect non-sinus beats, sudden pauses as indicators of atrioventricular conduction disturbances or sinus pauses, and sometimes irregularity as a measure for the presence of atrial fibrillation or flutter. Accordingly, Supraventricular’s abnormalities causing no or only gradual changes in ventricular rhythm are not observed by the current analysis systems, those who are referring only to the QRS complex for tracing the cardiac activity [23].

For that explanation, the way out of this problem would be, of course to analyze the P and T waves and other ECG’s elements. Not only but also, measuring the time interval between these elements. Nevertheless, this is technically not feasible in the current remote cardiac health monitoring systems because of computational cost considerations.

3.2 Organizing the training dataset

The organization of the classifiers’ training dataset plays a major role in improving the performance of detecting the arrhythmia classes. The two techniques – local and global training dataset – for selecting the right dataset are not satisfying the need to detect the arrhythmias accurately in different situations. The nature of ECG data in real time monitoring applications involves many changes through the time. For example, the different situations and activities of the monitored person are varying along the day. Sometimes, sleeping and sometimes running or working very hard. That beside the sex, age, environment factors, temperature, different kind of foods and drinks, medicines, and the mode which changes through the hour. All of these factors and others, affecting the heart activities and in usual cases it is not a bad effects but it will be detected by the ECG leads and transferred to the classifier model which is constructed by old training dataset. In case of worse condition when arrhythmias occur, the current classifier model may not be able to detect the abnormalities or it may detect them but afterward.

As a result, the model constructed using the old training data no longer need to be adjusted in order to identify with the new concepts. In view of that, developing one classifier model to satisfy all patients in different situation using static training datasets is unsuccessful.

4 The nested ensemble technique

The process of monitoring the cardiac health in a real time is a very complicate process. Although it is very useful to detect the type of arrhythmias in the right time with high accuracy, which is offer a good chance to save many people life. The literature shows many research progress and technologies introduced by data mining tools to find solutions for such problem. Our suggested technique is composed of two parts, the outer ensample for manipulating the training data set and the inner one which utilized to manipulate the ECG features, mainly the QRS, P, and the T waves. Each outer ensemble is organized by inner ensemble to ensure high quality of cardiac arrhythmia classification in real time with quite relief to the computational cost. Figure 2 illustrates the nested ensemble technique components, and the following steps show the sequences:

1. Building a model with different dataset.
2. In each data set there is evaluation with the P, T, and QRS complex.
3. Majority voting for the inner ensemble to select the best features.
4. Majority voting for the different training set to select the best one.
5. Building the classifier with suitable features through proper training data set.

![Nested ensemble technique components](image1)

**Fig. 2** nested ensemble technique components

4.1 Multiple Training Datasets

The ECG outer ensemble simply introduced to learn the classifier model by up-to-date training data. Figure 3 illustrates our idea to build an outer ensemble classifier model.

![ECG ensemble training data set](image2)

**Fig. 3** ECG ensemble training data set

The ECG data can be received from the person in different situations or doing dissimilar activities. Accordingly the learning processes continue through the day. Consequently, the classifier model will be built for each training set and by enhancing majority voting technique we can select the best model that can detect arrhythmias in any situation. Finally we can accomplish the predicted class which descripts the heart physical condition at that period of time in very sensitive manner.

We are applying the boosting method for this purpose. Therefore, the sampling is take place by replacement. First of all we organize the training data set with equal weights $1/n$. Then the classifiers are induced by the training set and used to classify the tested data, secondly updates the weight of the Boosting method at the end of each round either increasing or decreasing depending on the success or failure of base classifier respectively, and finally the ultimate model obtained by aggregating the base classifier generated by each round [18]. The reason behind select Boosting algorithm is that, the activities of the monitored persons to some extend can be specified and usually it takes long period of time, while shifting from one to anther activity takes place very fast. Therefore, ensemble the training dataset technique could relief the problem of improbability of the classifier which generated from learning the classifier model by old dataset.

4.2 Subset features design

In favor of detecting different types of arrhythmias, the R to R interval must be measured to perceive the heart rate variability HRV, the appearance of all waves with specific shapes, and the normal duration between similar waves must be checked in each heart cycle (bet to bet) [1] [24] [25]. Such computation is very complex to carry out by a computer with limited resources, although there are great interests in high accuracy to specify the arrhythmias, with high speed to save people life and avoid risks. This is the reasons to select a subset of these features in order to detect an arrhythmia accurately.

The inner model mainly designed to select the features related to the QRS complex plus those which related to P or T waves. We utilize the random forest method for this purpose which randomly selects N samples with replacement [18]. The subset feature selection models work inside each group of training dataset. So we design a unique feature set could be employed to describe specific arrhythmia in very sensitive manner.

5 The experimental works

5.1 Environment

We used a database generated at the University of California, Irvine [26]. It was obtained from Waikato Environment for Knowledge Analysis (WEKA), containing 279 attributes and 452 instances [27]. The classes from 01 to 15 were distributed to describe normal rhythm, Ischemic changes (Coronary Artery Disease), Old Anterior Myocardial Infarction, Old Inferior Myocardial Infarction, Sinus tachycardy, Sinus bradycardy, Ventricular Premature Contraction (PVC), Supraventricular Premature Contraction, Left bundle branch block, Right bundle branch block, degree
AtriOventricular block, degree AV block, degree AV block, Left ventricle hypertrophy, Atrial Fibrillation or Flutter, and Others types of arrhythmias Respectively.

The experiments were conducted in WEKA 3.6.1 environment. The data set was divided into two mutually disjoint sets, training and testing sets. The training one used to train the classifier model through the boosting method. The ECG features related to the QRS complex, P, and T waves are organized by the random forest method.

Our experiment were carried out by PC with processor intel core (T M) 2 DUO, speed to 2.40 GHz. And RAM 2.00 GB.

5.2 Results

We implemented two types of experimental works, the first one to prove the necessity of including the P and T waves in conjunction with the QRS complex to evaluate arrhythmias in the right way. And the second work to provide evidence about the value added by our technique regarding the improvement of the classifier’s accuracy. The performance was assessed in term of accuracy which means the correctness for specifying the arrhythmias type.

Referring to the first experiment we measure the performance of three different algorithms the OneR, J48 and Naïve Bayes according to the feature or features used to classify the arrhythmias. Table 1 summarizes the results for each algorithm.

<table>
<thead>
<tr>
<th>features</th>
<th>OneR</th>
<th>J48</th>
<th>Naïve Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRS only</td>
<td>60.4</td>
<td>91.2</td>
<td>76.5</td>
</tr>
<tr>
<td>QRS + P</td>
<td>60.4</td>
<td>91.4</td>
<td>77</td>
</tr>
<tr>
<td>QRS + T</td>
<td>61.3</td>
<td>91.2</td>
<td>76.7</td>
</tr>
<tr>
<td>QRS + P + T</td>
<td>61.1</td>
<td>92.3</td>
<td>77.7</td>
</tr>
</tbody>
</table>

Figure 4 illustrates the accuracy achieved by each algorithms with different features set. All of them – except the OneR – their accuracy increase when we include features related to QRS complex, P and T, while there was light improvement when QRS complex included with only P or T waves. The result proves that, there is a great need of using all types of features to detect all types of cardiac arrhythmias. Because the QRS complex alone, can perceive only some arrhythmias and the monitoring process can take place to merely a minority of the heart activity. On the other hand, we can attain the majority when we utilize the QRS with P and T waves.

In the second experiment works we introduced our nested ensemble technique in comparison with the J48, since it scored a high performance when compared with anther methods according to the previous experiments.

The experiment was conducted using the features related to the QRS complex, P and T waves. Figure 5 shows the superior of the nested ensemble technique in the process of detecting the arrhythmias with high accuracy when compared with the J48. It is a clear prove that the nested ensemble can enhance the process of detecting the different types of arrhythmia. Since the technique selects the right features within the suitable learning dataset. Thus, there could be very efficient cardiac health monitoring to specify the type of the arrhythmia in very accurate mode.
6 Conclusion

Cardiovascular diseases are responsible of huge amount of death along the world. Detecting the heart arrhythmias through ECG monitoring is mature research achievement. But it’s accessible to patient in hospitals only, with lack to handle others outside continually, since identifying arrhythmia in the right time is indispensable to protect patient and to minimize the risk as much as possible, a relief solution to provide 24 hours monitoring for the cardiac patient is one of demanding and promising area of research.

There has been much work in the area of classification arrhythmias on real time. The dataset used to train the classifier are often small. Moreover, the verity of the conditions and the diverse activities of the patients, make such learning dataset is of no use for the reason that there are rapely change during the time. As a result, the model constructed using the old training data no longer need to be adjusted in order to identify with the new concepts. In view of that, developing one classifier model to satisfy all patients in different situation using static training datasets is unsuccessful.

The ECG nested ensemble technique suggested for providing the classifier by two main resources, modern training data. Accordingly the learning processes continue constantly. Therefore, the classifier model will be selected to satisfy the needs at that moment and then carry out the predicted class which describes the heart physical condition at that period of time in very sensitive way. And the required features in addition to the QRS complex to satisfy the need to detect all types of arrhythmias.

As a result, and according to our experiments nested ensemble technique can relief the problem of unlikeliness of the classifier that generated when learning the classifier by old dataset and limited input features. In our future work we are going to improve the performance of the nested ensemble by utilizing different ensemble methods.

7 Acknowledgement

This work was supported by the grant of the Korean Ministry of Education, Science and Technology(The Regional Core Research Program / Chungbuk BIT Research-Oriented University Consortium), and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 2010-0001732).

This works was also funded by University of Medical Science and Technology UMSST. It is a private, non-profit making University, located in Khartoum, Sudan.

8 References


