Effective Computational Modeling for Early Arrhythmia Symptom Classification by Using Decision Tree Approach

Mohamad Sabri bin Sinal*, Eiji Kamioka
Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo 135-8548 Japan.

* Corresponding author. Email: nb16109@shibaura-it.ac.jp, kamioka@shibaura-it.ac.jp
Manuscript submitted December 10, 2018; accepted January 10, 2019.

Abstract: Heart disease has been the leading global cause of death for almost 15 years. One of the common causes lead to chronic heart disease and sudden death is Arrhythmia. However, the conventional or computational approach of Arrhythmia detection is not an easy task. It requires suitable method with a very specific timeline to detect the symptom. In addition, the symptom itself is very complex in behavior. Therefore, an automatic detection method with simple computational model to detect accurately Arrhythmia in ECG data is needed to deal with such critical issue. In this paper, a novel framework based on decision tree approach by utilizing five peaks taken from ECG segment is proposed to detect Arrhythmia from the first minute of the ECG data. The experimental results show that the proposed decision tree approach with the proposed five peaks is able to detect Arrhythmia with the accuracy of 98% outperforming the other data mining techniques. Moreover, the five proposed parameters to classify the disease show that these computational models have a strong level of sustainability in detecting Arrhythmia when it is compared to different numbers of parameters and methods.

Key words: Heart disease detection, decision tree for heart disease, arrhythmia detection, computational analysis for heart disease symptom.

1. Introduction

Heart disease has been the leading global cause of death around the world for the last 15 years. In 2016, over 15.6 million of 56.9 million deaths were observed as heart disease deaths [1]. The most common cause that leads to chronic heart disease and sudden death is contributed by Arrhythmia symptom. Arrhythmia is defined as irregularity of the heart activity that occurs in time. These phenomena can be divided into two categories. The first one is known as tachycardia where the heart beats too fast and the second is known as bradycardia where the heart beats too slow. In order to detect such anomaly phenomena, conventionally, electrocardiogram (ECG) signal is used. The ECG is well known in the medical field as the best reference to represent an individual's heart activity. It represents the electrical activity of the heart recorded from 12 leads of the body. Since abnormal electrical activity of the heart can be a life threatening, relentless efforts in biomedical engineering and computer science research domain have been conducted continuously to overcome such critical issue. Moreover, computational intelligence approach has made a significant impact on solving complicated problem including dealing with heart disease symptom.

With the complexity of biomedical data and advance computational intelligence technology available, the effectiveness of data analysis to interpret and classify such behavior can be achieved constructively. Conventionally, computational intelligence approaches to diagnose Arrhythmia symptoms or heart disease...
symptom from the ECG signal consist of 4 steps, which are 1) ECG signal processing, 2) heartbeat segmentation, 3) feature extraction and 4) classification. For feature extraction method, the most common methods used are the wavelet feature [2]-[4], the higher order statistical feature [5], [6] and the morphological feature [7], [8]. For classification, the most widely method used in these research areas are the artificial neural network [3], [9], the K-Nearest-Neighbor (KNN), the support vector machine [3]-[5], [10], the linear discrimination, the naïve Bayes [11], [12], the principal component analysis [13] and the Decision Tree approach [14], [15].

However, there are huge challenges in designing an efficient computational intelligence approach to detect heart disease symptom. One of the biggest challenges is to choose the right feature obtained from the data to represent the disease symptom. An efficiency method is usually relying on the right feature to detect heart disease accurately [16]. Even though a huge number of attributes of ECG signal is used, the complexity in the feature extraction may occur and over fitting in the classification stage may trigger. As a result, the accuracy of detection may decrease. Even with a small number of attribute is propose, the possibility to achieved high accuracy of detection is not 100% guarantee. Due to the lack of information to characterize the symptom accurately, under fitting may happen during training of the dataset. As a result, the accuracy of detection may also decrease. There is always a trade-off between the number of attributes, the complexity of the procedure and the accuracy of detection. Biomedical data in general and ECG data in particular are well known as data which consist of complex and diverse characteristics. Therefore, characterizing these types of data is a big challenge for this domain. Moreover, different computational models require different data mining approaches to deal with such complexity. For that, it is important to highlight that selecting the right data mining approach is a one step forward to achieve high detection rate after defining the right number of attributes. High suitability information describing the disease symptom quantitatively may help in computational diagnosis of heart symptom abnormalities more accurately and quickly. To conclude, there are two challenges to overcome in this study; 1) to clarify the most optimal and suitable attribute in ECG data to represent the two different symptoms quantitatively based on ECG peak and 2) to determine the method to classify the abnormalities from the normal cycle in the early minutes of ECG data.

This paper aims is to focus on proposing an efficient and accurate method to discriminate the Arrhythmia from the Normal Sinus at the early minute of ECG data using data mining approach. In this study, five parameters are introduced to characterize the symptoms. In addition, several data mining approach and parameters are tested to validate the effectiveness of the proposed works to detect the disease at the early minutes of the ECG data.

The remainder of this paper is organized as follow: Section 2 will introduce briefly about Normal Sinus and Arrhythmia Rhythm. Section 3 will discuss related works of this study to clarify issues that exist in this domain particularly on detecting Arrhythmia symptom using the ECG data. Section 4 will explain the detail of the framework to detect Arrhythmia based on the proposed parameters from one minutes data. For section 5, the evaluation results based on the proposed framework will be discussed. In section 6, conclusion will be made as a summary of this work.

2. Principle of Electrocardiogram

ECG is a periodic signal wave generated by the heart muscle. The signal itself represents heart activities and it can be used to analyse the health condition of the heart. Conventionally, cardiologists use this type of data to diagnose the heart condition in detail and find any potential abnormalities that may exist within the signal form, orientation and rhythm. In this paper, two types of symptoms are chosen to represent a healthy and an unhealthy heart condition which are Normal Sinus and Arrhythmia symptoms.

2.1. Normal Sinus Rhythm
Normal Sinus rhythm is a condition where the functioning conduction systems in the body behave normal. The electrical current from an electrocardiograph flows through the normal conduction pathway without any disturbance [17]. This normal heartbeat is originated from the sinoatrial node.

2.2. Arrhythmia Rhythm

Arrhythmia is a heart disorder that affects the heart activity randomly in time. The symptom can drive the heart beats too fast, too slow, too early or too erratic. An Arrhythmia happens when the electrical impulse does not function properly. Arrhythmia consists of a variety of sub symptoms where each symptom consists of its own unique identification characteristics.

2.3. P, Q, R, S, T Wave Morphology

P, Q, R, S and T waves morphology involve two main processes which are depolarization and repolarization of the atrial and the ventricular. It will generate a series of periodical waves which start with the P wave and then followed by the QRS complex and T wave. The P wave is a reflection of the heart’s depolarization activities at the right and left atria. It is followed by the Q, R and S waves which represent the activation of the right and left ventricles. The T wave reflects the heart’s repolarization which happens at the ventricles. A healthy heart condition normally behaves periodically and regularly in time contrastive to an unhealthy heart condition. Figure 1 and figure 2 represent Normal Sinus and one sample of the Arrhythmia waveforms, respectively. Both waveforms show a huge wave difference in patterns. In this research, the characteristic of both symptoms in each heartbeat is used to classify the symptoms by using data mining approach.

![Normal Sinus Rhythm](image1.png)

![Arrhythmia Rhythm](image2.png)

Fig. 1. Normal Sinus heartbeat segment.  
Fig. 2. Arrhythmia heartbeat segment.

3. Related Works

Ali Isin and Selen Ozdilili [17] proposed the utilization of deep learning approach to classify three different cardiac conditions in the ECG data. The proposed framework of this study consists of 2 important procedures. First, the feature extraction procedure is done by using a deep convolutional neural network. Second, the back propagation of the neural network is utilized to classify the disease symptom based on image pattern. This study used MIT-BIH Arrhythmia database to validate the effectiveness of the proposed method. The results of the experiments signify that the combination deep convolutional neural network with back propagation of the neural network is effective with 98.51% of correct recognition is achieved. However, there is a drawback in this study particularly in the implementation of deep convolutional neural network. These techniques only accept specific size of images as inputs for classification. To solve this issue, a conversion of the original photo into binary images is required. The process itself is complex and not
practical.

Likewise, S.Karimifard et al. [18] proposed a morphological heart Arrhythmia detection using Hermitian basis functions and KNN classifier. The Arrhythmia detection is based on the features of electrocardiography (ECG). This study utilized MIT-BIH Arrhythmia database to validate the effectiveness of the proposed method. The proposed methods of this study involve two steps; first, the ECG beats were first modelled using Hermitian basis function. Second, the feature vector which consists of parameter of the model is used as an input to k-Nearest neighbor to examine the efficiency of the model. In this study, seven different types of Arrhythmia are covered for Arrhythmia detection. The results show that the sensitivity of 99% and specificity of 99.84% are achieved. However, the proposed methods required some additional optimization procedures to minimize any possibility of error that may occur during modelling the parameter. In addition to that, this additional procedure is one of the key important factors to ensure sustainability of the proposed method to achieve high accuracy in disease detection. Therefore, it is assumed that without relying too much to such optimization procedure and minimization of the complexity in Arrhythmia detection may increase the practicality of using the mechanism.

Uday Maji et al. [19] proposed Variational Mode Decomposition (VMD) approach to detect abnormalities in the ECG signal. VMD is a non-recursive signal decomposition method where it has the capability to decompose the input signal into the desired number of mode N. The proposed method is utilized mainly to characterize the atrial and ventricular Arrhythmia simultaneously and independently from each heart cycle segment. The experimental results had revealed that the proposed method achieved the Arrhythmia detection accuracies of 99.1% for ventricular and 99.8% for atrial segment, respectively. However, there are two limitations to this study. First, the proposed methods of this study require separation of the ECG segment into two parts in order to accurately analyse the abnormalities. This has increased the complexity of the classification process. Second, it requires multi-stage classification to ensure the sustainability of high detection accuracy. As a result, it is not efficient and the computational cost is high. Therefore, it is important to find the best method and computational model to detect abnormalities of heart condition quantitatively and accurately. Most of the existing works have shown that there are huge complexities in characterizing the abnormalities of heart condition. Therefore, an autonomous and effective computational model is needed to achieve the detection of heart condition abnormalities with high accuracy. The detection of the Arrhythmia symptoms will be the main focus for this study.

4. Methodology

In this section, a computational intelligence approach to detect abnormalities of heart condition based on five parameters is proposed. The five proposed parameters taken from the ECG segment are P, Q, R, S and T peaks. Figure 3 represent the location of each parameter labeled in detail. In this study, there is no filtering process is implemented towards the ECG data before classification. This is purposely to measure the level of sensitivity of the proposed data mining approach to detect abnormalities of heart condition regardless the condition of the data and which segment of ECG data is used. One of the most important criteria to achieve high accuracy is not only rely on data mining approach and filtering method but also which segment in ECG segment is used to represent the disease symptoms. Therefore, the values of each parameter are taken directly from the ECG data to represent each symptom. Decision tree is mainly used to classify the abnormal ECG patient data from normal ECG patient’s ECG data by using suitable number of attribute for high accurate detection.

Decision tree is a predictive model which utilizes graphical representation like root, nodes and branching decision to represent specific decision situation precisely. Decision tree has strong advantage in dealing with different scale of data and it is less sensitive to the outlier. Due to that, it is expected that high accuracy
of disease classification can be achieved even though the noise signal is included inside the ECG data. Moreover, the decision tree is proven to be among the most common technique for disease classification in medical science [20]. Therefore, in this study, three important stages are introduce and described below:

- **Feature extraction**: extract feature from 1 minutes ECG signals. It is based on P, Q, R, S and T peaks.
- **Dataset preparation**: prepare a complete dataset in 1 minutes for classification. All the five peaks are put in order.
- **Disease classification**: In this phase, decision tree based on five parameters is perform. Five-fold cross validation is utilized to avoid over fitting and under fitting.

![Fig. 3. Five proposed parameters from a heart cycle segment.](image)

5. **Evaluation**

In this section, the performance of the proposed framework to detect Arrhythmia symptom is discussed. The accuracy of detecting Arrhythmia is used as a metric to evaluate the overall performance of the purpose framework. The classification accuracy is defined as the overall performance of the proposed framework to differentiate Normal Sinus from Arrhythmia [21]. The classification accuracy is described as follows:

\[
\text{Accuracy} = \frac{\text{True negatives} + \text{True positive}}{\text{True negatives} + \text{True positive} + \text{False negative} + \text{False positive}}
\]

where,

- **True positive**: The number of Arrhythmia peaks correctly identified as Arrhythmia peaks.
- **False negative**: The number of Arrhythmia peaks incorrectly identified as Normal Sinus peaks.
- **False positive**: The number of Normal Sinus peaks incorrectly identified as Arrhythmia peaks.
- **True negative**: The number of Normal Sinus peaks correctly identified as Normal Sinus peaks.

The proposed framework is tested with three different types of decision trees by using five parameters taken from the ECG data. 10 ECG data from MIT-BIH Normal Sinus and 10 ECG data from MIT-BIH Arrhythmia database are selected as the main database for the evaluation of this study. These two databases are provided by PhysioNet. PhysioNet is a free open source database which provided varies type of biomedical database for research and development. All ECG data are original data from real patients with Arrhythmia and Normal Sinus symptom. Therefore, the result of this experiment reflects the real disease behavior. In this experiment, 774 complete dataset of PQ,RS and T peaks for Normal Sinus and 626 PQ, R,S and T peaks for Arrhythmia taken from 1 minute ECG data are analysed. Each peak is investigated in detail to ensure the exact peak is correctly extracted from each periodical heart cycle segment of 1 minute without including the noise peak. As a result, the overall quality of the ECG data and the peaks are acceptable for evaluation and analysis. From the experimental result, the overall detection accuracy of Arrhythmia is 98%.
Table 1 shows the overall accuracy when it is tested with 3 different types of decision trees.

<table>
<thead>
<tr>
<th>Technique</th>
<th>P,Q,R,S,T Peak Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree (Complex Tree)</td>
<td>97.9</td>
</tr>
<tr>
<td>Tree (Medium Tree)</td>
<td>98</td>
</tr>
<tr>
<td>Tree (Simple Tree)</td>
<td>96.5</td>
</tr>
</tbody>
</table>

In Table 1, the overall detection accuracy with three different decision trees are consistently high. It verifies the effectiveness of the proposed framework to detect Arrhythmia with five peaks. In order to confirm that the decision tree is the best classifier for these five proposed parameters, 18 different classifiers are selected and evaluated. Moreover, five proposed parameters are compared to the other number of parameters to ensure that five proposed parameters are the best amounts of parameters used to detect Arrhythmia symptom with decision tree. This comparison is very important in order to identify the most sustainable computational modeling by using ECG peak to detect Arrhythmia at the early minutes. Table 2 represent the overall accuracy detection for 18 different classifiers using five proposed parameters. Table 3, table 4 and table 5 represent the overall accuracy for Arrhythmia detection by using 3 different decision trees with one parameter, two parameters and three parameters by using different combination of peaks which represent Normal Sinus and Arrhythmia symptom.

Table 2, Table 3, Table 4 and Table 5 provided strong evidence showing that decision tree has overcome the other data mining approaches quantitatively. Moreover, five proposed parameters perform much better in terms of accuracy when it is compared with different number of parameters. From the experiments, it is concluded that the variability of two different symptoms represented by five peaks is high, identical, and thus, significant. Therefore, it is confirmed that those five parameters are the influential factor towards high
accuracy detection in disease classification. The proposed computational model to detect Arrhythmia based on five parameter shows that this model are very sustainable compared to 18 different classifiers. The overall accuracy is very high and consistent from one classifier to another. The decision tree has shown a strong capability with this kind of dataset. Hence, the sensitivity towards outlier is less with decision tree. Therefore, a small difference in value in small size dataset may not affect the overall classification accuracy of performance. Moreover, decision tree requires less time compared to other parametric model if the model does not suffer from over fitting or under fitting. As a conclusion, this finding justifies the significance and the robustness of the proposed framework in classifying Arrhythmia from Normal Sinus with 1 minutes ECG data. This study can be a good reference as a pilot study to design a sustainable computational model for detecting Arrhythmia based on ECG peak.

Table 3. Detection Accuracy of Arrhythmia in Each Decision Tree by Using One Parameter

<table>
<thead>
<tr>
<th>Technique</th>
<th>P peak Accuracy (%)</th>
<th>Q peak Accuracy (%)</th>
<th>R peak Accuracy (%)</th>
<th>S peak Accuracy (%)</th>
<th>T peak Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree (Complex Tree)</td>
<td>87.8</td>
<td>70.7</td>
<td>75.2</td>
<td>71.2</td>
<td>86.8</td>
</tr>
<tr>
<td>Tree (Medium Tree)</td>
<td>85.2</td>
<td>65.8</td>
<td>74.3</td>
<td>68.2</td>
<td>85</td>
</tr>
<tr>
<td>Tree (Simple Tree)</td>
<td>85</td>
<td>62.6</td>
<td>74.9</td>
<td>63.9</td>
<td>84.5</td>
</tr>
</tbody>
</table>

Table 4. Detection Accuracy of Arrhythmia in Each Decision Tree by Using Two Parameters

<table>
<thead>
<tr>
<th>Technique</th>
<th>P, Q Peak Accuracy (%)</th>
<th>P, R Peak Accuracy (%)</th>
<th>P, S Peak Accuracy (%)</th>
<th>P, T Peak Accuracy (%)</th>
<th>Q, R Peak Accuracy (%)</th>
<th>Q, S Peak Accuracy (%)</th>
<th>Q, T Peak Accuracy (%)</th>
<th>R, S Peak Accuracy (%)</th>
<th>R, T Peak Accuracy (%)</th>
<th>S, T Peak Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree (Complex Tree)</td>
<td>90.2</td>
<td>87.6</td>
<td>86.1</td>
<td>94.6</td>
<td>80.9</td>
<td>83.8</td>
<td>91.5</td>
<td>82.8</td>
<td>89.8</td>
<td>98.3</td>
</tr>
<tr>
<td>Tree (Medium Tree)</td>
<td>90.6</td>
<td>88.6</td>
<td>86.8</td>
<td>94.7</td>
<td>79.3</td>
<td>83.8</td>
<td>88.2</td>
<td>80.7</td>
<td>88.8</td>
<td>97.7</td>
</tr>
<tr>
<td>Tree (Simple Tree)</td>
<td>88.8</td>
<td>87.7</td>
<td>84.6</td>
<td>94.9</td>
<td>75.8</td>
<td>75.7</td>
<td>83.9</td>
<td>75.7</td>
<td>83.6</td>
<td>92.5</td>
</tr>
</tbody>
</table>

Table 5. Detection Accuracy of Arrhythmia in Each Decision Tree by Using Three Parameters

<table>
<thead>
<tr>
<th>Technique</th>
<th>P, Q, R Peak Accuracy (%)</th>
<th>P, Q, S Peak Accuracy (%)</th>
<th>P, Q, T Peak Accuracy (%)</th>
<th>P, R, S Peak Accuracy (%)</th>
<th>P, S, T Peak Accuracy (%)</th>
<th>Q, R, S Peak Accuracy (%)</th>
<th>Q, R, T Peak Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree (Complex Tree)</td>
<td>92</td>
<td>91.9</td>
<td>96.1</td>
<td>96.9</td>
<td>93.5</td>
<td>91.6</td>
<td>90.9</td>
</tr>
<tr>
<td>Tree (Medium Tree)</td>
<td>92.1</td>
<td>91.4</td>
<td>96</td>
<td>96.7</td>
<td>93.1</td>
<td>95.8</td>
<td>91.1</td>
</tr>
<tr>
<td>Tree (Simple Tree)</td>
<td>91.6</td>
<td>88.9</td>
<td>94.3</td>
<td>96.7</td>
<td>91.4</td>
<td>94.1</td>
<td>79.4</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, Arrhythmia classification based on five peaks from periodic heart cycle is proposed. The five peaks are the P, Q, R, S and T peak. Decision tree is used as a main classifier to classify the two symptoms. From the experimental results, the evidence proves that detection accuracy of 98% in classifying Arrhythmia from Normal Sinus by using decision tree is achieved. Based on the computational model proposed in this study, it can be concluded that the five parameters show the effectiveness in term of accuracy when it is compared with different numbers of parameters and different combination of parameters to detect the same disease. It is confirmed based on this study that decision tree outperforms...
other machine learning approaches for 1 minute ECG data. In addition to that, the five peaks play such an important role to influence the overall accuracy in disease classification. The proposed computational model shows that these models are very sustainable when it is compared to 18 different classifiers. The overall accuracy between each classifier is high and consistent. As a conclusion, this research can contribute to the biomedical engineering field in designing computational model to detect Arrhythmia or abnormalities of heart condition.

References


Mohamad Sabri bin Sinal obtained his diploma in information technology (2008) from Island College of Technology, Penang, the bachelor of information technology majoring in software engineering (2012) from Universiti Utara Malaysia and the master of electrical engineering and computer science from Shibaura Institute of Technology (2016). Currently, he is in the final year of pursuing his PhD at Shibaura Institute of Technology and expected to graduate by March 2019. His research interests include health monitoring system, computational modeling for heart disease detection, data mining approach, cardiovascular disease, context-aware computing, computer algorithm, bioinformatics and Ubiquitous Computing.

Eiji Kamioka is a professor at Shibaura Institute of Technology. He received his B.S., M.S. and D.S degrees in physic from Aoyama Gakuin University. Before joining SIT, he was working for SHARP Communications Laboratory, Institute of Space and Astronautical Science (ISAS) as a JSPS research fellow and National Institute of Informatics (NII) as an assistant professor. His current research interests encompass mobile multimedia communications and ubiquitous computing.