HEART DISORDER DETECTION USING R TO R INTERVAL SIGNAL CLASSIFIER

Subairi[1], Delila Cahya Permatasari[1], Yandhika Surya Akbar Gumilang[1,*] and Wahyu Dirgantara[1]
[1]Electrical Engineering Department, University of Merdeka Malang, Indonesia
* Corresponding Author: Subairi. E-mail: subairi@unmer.ac.id

ABSTRACT
The heart is the most important part of our body. So, the condition of the heart is affected by all of our bodies. If our heart has trouble, our body feels it too. So, a routine heart health check is important. One of the heart disorders is arrhythmia. Arrhythmia is a disorder of the heart rate or heart rhythm which is characterized by an irregular heartbeat, which can be too fast or too slow from the normal heartbeat. One method to detect arrhythmia is to determine the value of the R to R interval signal on the heart rhythm signal (ECG Signal). ECG signal has a peak wave value, which is called an R signal. From the R signal, the researcher can detect the condition of the heart. Purposed research is an interval R to R signal classifier using the Decision Tree Algorithm. The decision tree can find solutions to problems by making criteria as nodes that are interconnected to form a tree-like structure. The process of the purposed method first is the detection of R to R using filters signal is called preprocessing, second is the extraction feature, from ECG signal extracted to 7 features, and the last process is the decision tree using Gini index. The result of the purposed method shows that the accuracy of heart disorder detection using a decision tree is 98,27% accuracy. From 3140 epochs, it detects 2105 normal (N), 972 arrhythmia (A), and 63 misclassifications. It is the purposed method that has a 98% success rate. From these results, the purposed method is valid for heart disorders.

Keywords: Heart Disorder, ECG signal, Interval R to R, Decision Tree.

1. INTRODUCTION

The heart is a complex structure consisting of fibrous tissue, cardiac muscle, and an electrical conduction heart. The heart has the main function of pumping blood. This can be done well if the ability of the heart muscle to pump is good enough, and the pumping rhythm is good (Brunner et al., 2002). If the heart is not pumping well enough, it will be very risky for heart disease or heart failure.

To do the right treatment in patients with heart disorders, it is necessary to know the cause of the abnormalities suffered. One method that can be used is to observe the Electrocardiogram (ECG) signal from the patient with the disorder. An electrocardiogram (ECG) is a signal that represents the electrical activity of the heart. In the ECG signal, there are special signals, namely the P signal, the T signal, and the QRS complex. By using these signal components, abnormalities caused by the abnormal electrical activity of the heart can be detected. One of the most detected components is the R signal which is used to calculate the heart rate value (beats per minute).

Under normal circumstances, the adult human heart rate is 60-100 beats/minute. A heart rate that is less or more than this range can cause problems with blood circulation in the body known as arrhythmias. Based on the number of heartbeats, arrhythmias are divided into two, namely bradycardia (heart rate less than 60 beats/minutes) and tachycardia (heart rate more than 100 beats/minutes).

Various kinds of detection systems with different detection techniques were 2018(Stefanie & Bangsa, 2018) designed for the detection of Interval R as an indicator of cardiovascular disease. In 2015, (Akhter et al., 2015) designed a system that is more effective in detecting R-R interval irregularities by removing ectopic data from the ECG. Based on the description above, in this article, a design for detecting heart problems using the R interval signal classification will be carried out. This research is aim to provide a more effective and efficient solution for the detection of heart problems. The goal to be achieved is to detect and determine disorders of the heart through the R to R interval classifier on the Electrocardiogram (ECG) signal.

2. METHOD

2.1. ECG Signal

In the medical world, heart health is a crucial and major problem. Statistically, heart disorder is the number one cause of death in the global country. Heart disease/disorder cause 18 million deaths in 2016. Because of it, the development of technology made we can monitor heart conditions and performance by checking for fluctuations in ECG signals. Several electrodes are placed on the patient's body, especially the chest area, it easier for electrodes to detect the heart's electricity. The heart's electricity measured and recorded is called an electrocardiogram (ECG) (Wasimuddin et al., 2020). The ECG signal is a periodic signal that can be studied cycle by cycle. Figure 1 shows a cycle of the ECG signal.

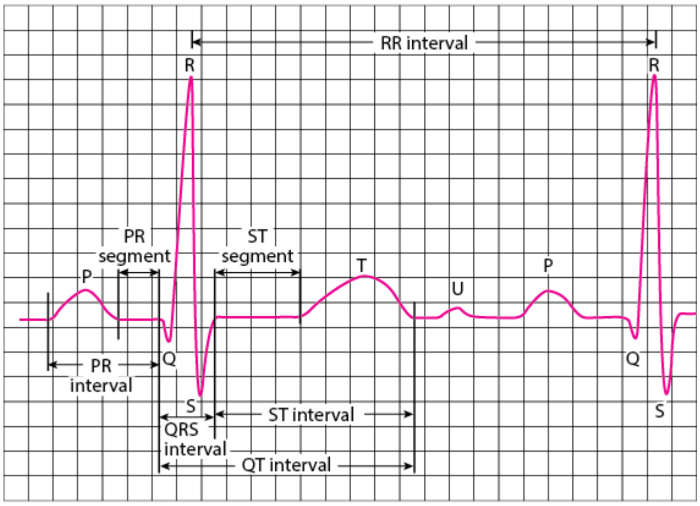

Figure 1: ECG Signal Wave

From Figure 1, it can be observed in the ECG signal has many wave points. P, R, and T wave points are the maximum extreme wave points. Q and S wave points are the minimum extreme wave points. One cycle of the ECG signal contains extreme wave points, namely the PQRST wave points. In this study, a preprocessing method was used to find extreme wave points, the locations of these important wave points could be identified for each cycle of the ECG signal (Guyton & Hall, 2000). Detection of extreme wave points can be determined by tracking the slope. The slope of a curve f(x) can be found from its first derivative f'(x). These criteria can be seen in the following figure (Afroni & Basuki, 2020):

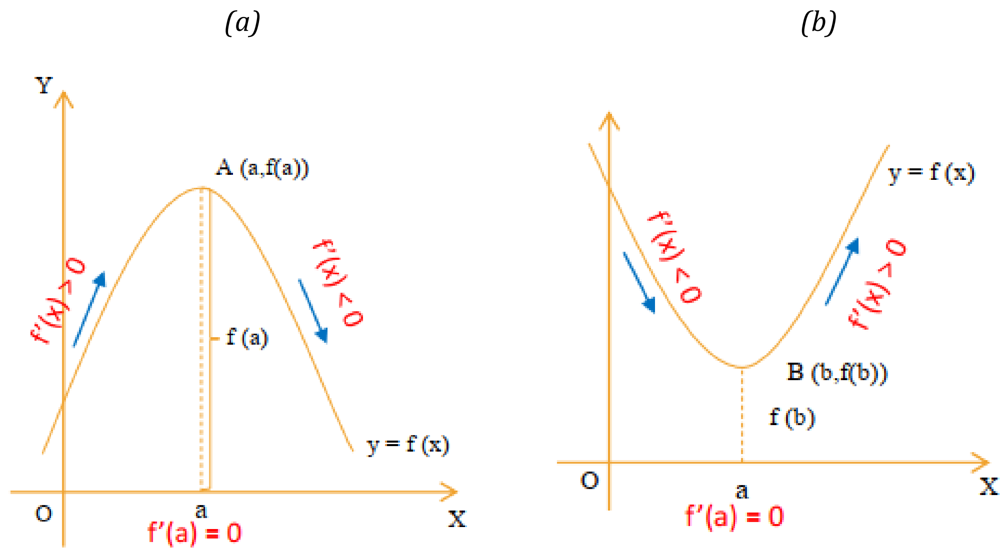

Figure 2: Extreme Wave Point (a) Maximum (b) Minimum

Source: (Afroni & Basuki, 2020)

Figure 2, showing the slope properties of the extreme points is obtained from the first derivative of the function, namely f'(x). f'(x) can be calculated according to equation (Guyton & Hall, 2000):

$$f'(x) = \frac{f(x+1) - f(x)}{(x+1) - (x)} \tag{1}$$

ECG signal recording data is discrete data because it is obtained from a real signal sampling process, so it is very difficult to obtain an extreme point that meets the criteria f'(x) = 0. To overcome this, the following criteria adjustments are used (Afroni & Basuki, 2020).

a) Let the slope of the curve at point x be f'(x).

b) If f'(x) > 0 and f'(x+1) < 0 then x is the maximum extreme wave point.

c) If f'(x) < 0 and f'(x+1) > 0 then x is the minimum extreme wave point.

To detect point R which is the highest peak of each cycle of the ECG signal, a search for the highest maximum extreme point is used from a collection of maximum extreme points that have been detected. Point R is used as a reference to detect the maximum point in another cycle, namely, P and T. Detection of point S, which is the lowest minimum extreme point, is used to find the minimum extreme point of all detected minimum extreme points. After point S is found, the minimum extreme point before S is point Q in the same cycle (Gupta et al., 2020). With the discovery of point R, the next maximum extreme points will be recognized as points T and P. Point R in the first cycle of data recording can be used as the starting point for the ECG signal and a reference to identify the start of the cycle. In the same way, the minimal extreme points of S and Q can be identified (Afroni & Basuki, 2020).

2.2. Dataset

Datasets of arrhythmia are collected from the physionet site. Datasets were researched from the MIT-BIH Arrhythmia database (Moody & Mark, 2001). MIT-BIH Arrhythmia database is recorded to take place in Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center). It recorded 47 subjects between 1975 and 1979. It was recorded for 3-4 hours, every subject has 650.000 ECG data. MIT BIH Arrhythmia dataset has 5 classifications of arrhythmia. This research has a modification in that data combine five classifications into one label, which is arrhythmia. So there are 2 labels, normal (N) and arrhythmia (A).

2.3. Preprocessing Method

Pre-processing processes for signal demonizing and signal compression. The preprocessing process is to obtain the R to R Interval value. The stages of preprocessing are shown in figure 3 below.

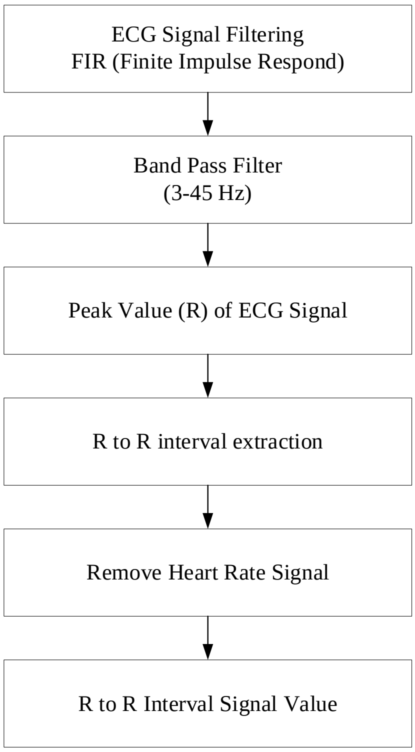

Figure 3: ECG Signal Preprocessing Process

2.4. Extraction Feature

The extraction feature is the process of taking the features contained in the signal. The extraction feature is a step in classifying and interpreting (Hardani, 2015). In this research, the statistical feature extraction method was used. ECG signal extract into 7 features. The features are max, min, median, mean, std (standard deviation), kts (Kurtosis), and skw (Skewness).

2.5. Decision Tree Classifier

A decision tree algorithm is used for decision-making. Decision trees can find solutions to problems by making criteria as nodes that are interconnected to form a tree-like structure (Pramadhani & Setiadi, 2014). The decision tree is a predictive model of a decision using a tree-like structure or hierarchy. The decision tree has branches, the branch has an attribute that must be met, if it is fulfilled it will go to the next branch until it ends in a leaf which means there are no more branches. The following is a diagram of the decision tree algorithm.

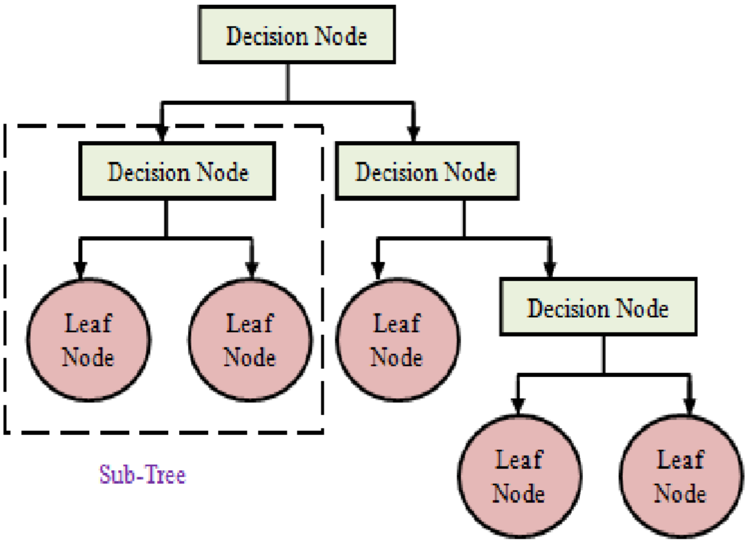

Figure 4: Decision Tree Diagram

Source: (Rohan & Kumari, 2021)

In the decision tree algorithm, there are two nodes, the decision node, and the leaf node. A decision node is used to complete any decision and it has different branches. The leaf node is the result of the decision and contains no further branches. The decision Tree has two methods, Entropy, and the Gini index. The Gini index and entropy are the criteria for calculating information gain. Decision tree algorithms use information gained to split nodes. Gini and entropy are both measures of node impurity. Multi-class nodes are impure, while single-class nodes are pure. Entropy in statistics is like entropy in thermodynamics, meaning disorder. If a node has multiple classes, that node is faulty. The purposed method is using the Gini index. Percentage of whiteness or impurity in settling on the CART decision tree (classification regression tree). If it differs from high GI, traits with a low Gini index (GI) should be preferred.CART estimation uses GI to create binary parts by only creating combined parts. The Gini index can be determined using the formula (Rohan & Kumari, 2021):

$$GI = 1 - \sum_{i=1}^{n} P^2(c_i) \quad (2)$$

P is the probability or percentage of class c_i in a node.

2.6. Diagram Block System

The Diagram of the research was shown in Figure 3. Including input from the arrhythmia dataset then processed to determine the R to R interval then using a decision tree algorithm to detect arrhythmia. The following figure 4 is a block diagram process of this research system.

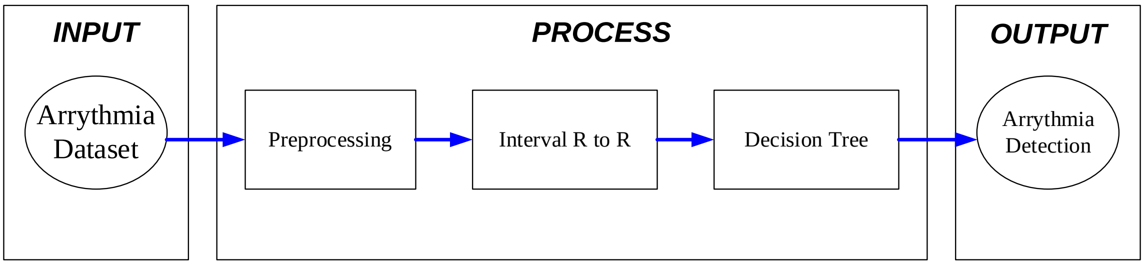

Figure 5: Purposed Method Diagram

3. RESULT AND DISCUSSION

Preprocessing process is useful to obtain the R to R interval. R to R interval can show the condition of the heart. By knowing the R to R interval, it can be shown subject has normal breathing (normal heart) or arrhythmia. The following is the result of the preprocessing to obtain the range R to R. In figure 6, it is the process to obtain the R to R interval value.

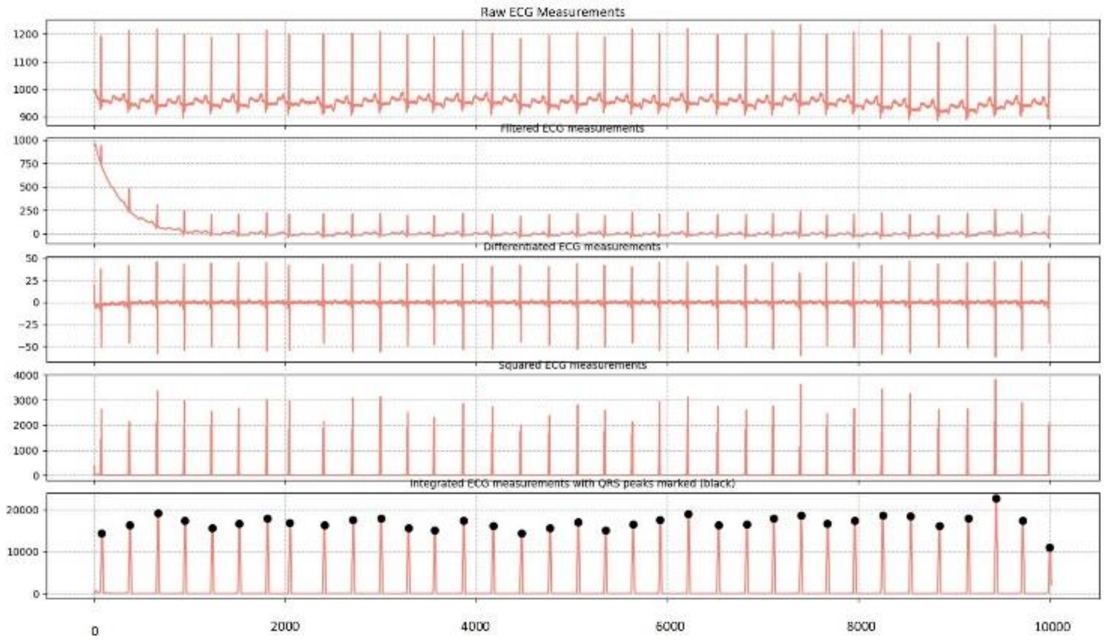

Figure 6: R to R interval Process

After the R to R interval value is obtained, the next step is the extraction feature. This is a sample of extraction feature value results shown in table 1.

Table 1: Extraction Feature Result

	max	min	median	mean	std	kts	skw	target
0	1218.0	962.0	2.855	2.655	0.493575	0.790806	-1.555976	0.0

1	1234.0	945.0	2.86	2.55087	0.595716	-0.826523	-1.051704	0.0
2	1234.0	945.0	2.84	2.4775	0.612809	-1.306062	-0.76141	0.0
3	1234.0	947.0	2.85	2.52913	0.591335	-0.866776	-1.009518	0.0
4	1228.0	947.0	2.9	2.673182	0.531348	0.697875	-1.611426	1
...	...	...	...	...	...	...	...	...
210	1276.0	1193.0	2.7	2.699091	0.096668	-0.806914	0.382562	0.0
211	1281.0	972.0	2.81	2.622273	0.452377	1.932533	-1.88418	1
212	1281.0	954.0	2.85	2.465417	0.617124	-1.291028	-0.784397	0.0
213	1276.0	954.0	2.77	2.545652	0.514467	-0.048909	-1.308767	0.0
214	1278.0	1198.0	2.64	2.667727	0.108248	-1.051014	0.027124	0.0

Feature data is split into the training dataset (60%) and the testing dataset (40%). Decision trees obtain an accuracy of 98,27%. It is shown in table 1 below.

Table 2: Decision Tree Accuracy Result

Arrhythmia Detection using Decision Tree	
Accuracy	98,27%
Training Value	96,75%
Test Value	94,56%

The confusion matrix of the decision tree is shown in figure 7. From figure 7, 2105 Normal (N) epochs, 972 arrhythmias (A) epochs, and 63 misclassifications epochs. The success rate of this research is 98%.

Table 3: Confusion Matrix Result

	Positive	**Negative**
Positive	2105 (TP)	25 (FP)
Negative	38 (FN)	972 (TN)

TP: True Positive *FP: False Positive*
FN: False Negative *TN: True Negative*

4. CONCLUSION

The following conclusions can be drawn:

1. This research aims to detect arrhythmia. The purposed method is a decision tree classifier using R to R interval value. The R to R interval value uses the ECG signal for the filtering method, bandpass filter, and signal extraction.
2. Arrhythmia detection by decision tree has 98,27% accuracy. From 3140 epochs, it detects 2105 normal (N), 972 arrhythmia (A), and 63 misclassifications. It's the purposed method that has a 98% success rate. From these results, the purposed method is valid for arrhythmia detection

REFERENCES

Afroni, M. J., & Basuki, B. M. (2020). Algoritma Pendeteksi Titik Ekstrim Pada Sinyal ECG Untuk Analisis Awal Gejala Aritmia. *JE-Unisla*, *5*(2), 400–404.

Akhter, N., Gite, H., Tharewal, S., & Kale, K. v. (2015). Computer-based RR-interval detection system with ectopy correction in HRV data. *2015 International Conference on Advances in Computing, Communications, and Informatics (ICACCI)*, 1613–1618.

Brunner, E., Bathke, A. C., & Konietschke, F. (2002). *Nichtparametrische datenanalyse*. Springer.

Gupta, V., Mittal, M., & Mittal, V. (2020). R-peak detection based chaos analysis of ECG signal. *Analog Integrated Circuits and Signal Processing*, *102*(3), 479–490.

Guyton, A. C., & Hall, J. E. (2000). Textbook of medical physiology, 2006. *WB Saunder's Co, Philadelphia*.

Hardani, D. N. K. (2015). Ekstraksi Fitur Sinyal Elektrokardiogram Berbasis Independent Component Analysis. *Techno (Jurnal Fakultas Teknik, Universitas Muhammadiyah Purwokerto)*, *16*(1), 10–15.

Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, *20*(3), 45–50.

Pramadhani, A. E., & Setiadi, T. (2014). Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit ISPA (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (ID3). *Jurnal Sarjana Teknik Informatika*, *2*(1), 831–839.

Rohan, R., & Kumari, L. V. R. (2021). Classification of sleep apneas using decision tree classifier. *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1310–1316.

Stefanie, A., & Bangsa, I. A. (2018). Perancangan Sistem Pendeteksi Iregularitas Interval RR Pada Sinyal Elektrokardiogram (Ekg) Sebagai Indikator Penyakit Kardiovaskular. *EEICT (Electric, Electronic, Instrumentation, Control, Telecommunication)*, *1*(2).

Wasimuddin, M., Elleithy, K., Abuzneid, A.-S., Faezipour, M., & Abuzaghleh, O. (2020). Stages-based ECG signal analysis from traditional signal processing to machine learning approaches A survey. *IEEE Access*, *8*, 177782–177803.