International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 1 (2018) pp. 350-357

Hypertension Heart Disease Heart Rate Time Series Classification using Reshape Detrended Fluctuation Analysis

Rifki Wijaya
Student, Department of Electrical and Informatics Engineering,
Institut Teknologi Bandung, Bandung, Indonesia.
Orcid Id: 0000-0002-8247-6584

Tati Latifah Erawati Rajab
Professor, Department of Electrical and Informatics Engineering,
Institut Teknologi Bandung, Bandung, Indonesia.
Orcid Id: 0000-0003-1321-2072

Ary Setijadi Prihatmanto
Assistant Professor, Department of Electrical and Informatics Engineering,
Institut Teknologi Bandung, Bandung, Indonesia.
Orcid Id: 0000-0002-9471-0227

Richard Karel Willem Mengko
Assistant Professor, Department of Electrical and Informatics Engineering,
Institut Teknologi Bandung, Bandung, Indonesia.
Orcid Id: 0000-0002-5747-6742

Abstract

Heart Rate is generally used by health workers to check heart condition. Heart rate has a non-stationary nature, so it cannot be predicted or processed before it is processed first. The non-stationary process used is Detrended Fluctuation Analysis (DFA). Peng can use heart disease by inserting heart rate into the DFA. The heart rate that changes over time convert into a scaling index that can show the difference between healthy and sick. Toru Yazawa perfects Peng's Scaling index by creating a new data retrieval with new time box. Toru Yazawa have successfully differentiated people with specific heart disease. R-DFA is new algorithm proposed by this research as a refinement of Peng's DFA and Toru's DFA. Peng and Toru do not create algorithms for a very long data (100 data retrieval for one subject) and do not have the ability to determine sick or healthy trend. A case study of 9 patients was sampled every 15 minutes for 100 times at different times. There are 900 data process R-DFA algorithm with little inaccurate data, unlike ECG, but it took with Polar H7 device. The precision of the device is one second per data. The data process to form the time series of the DFA's Peng scaling index. The time series process with subtraction of the scaling index. Subtracted time series drawn a regression line. R-DFA can classify the three classes with 92% accuracy with significant ("p<0,05"). This method produces a threshold used to determine the three classes.

Keywords: Heart Rate, Time Series, DFA, ECG, regression, hypertension heart disease, Mild Left Ventrikel Diastolic Dysfunction.

INTRODUCTION

Currently heart disease is the leading cause of death in the world. The World Health Organization (WHO) has estimated 12 million deaths worldwide appear every year from heart disease. In 2008, 17.3 million people died of heart disease. More than 80% of deaths in the world caused by heart disease. Miss. Chaitrali S. Dangare [1] in his book says that The World Health Organization (WHO) estimates that by 2030, 23.6 million people worldwide will die of heart disease. Disorders or diseases that often occur in the heart is coronary heart disease, namely obstruction of blood flow in the coronary arteries that supply oxygen and nutrients to move the heart. Examples of other clinical categories of heart disease may include myocardial infarction, angina pectoris, and coronary insufficiency [2].

Statistically a person with a risk factor for heart disease will have a higher tendency to suffer from coronary disorders when compared with those without risk factors. The more risk factors that owned, the more doubled the possibility of heart disease. Risk factors include age, male gender, heredity, hypercholesterolemia, high blood pressure, diabetes, and smoking.

This heart disease is deadly, so technology is needed that can detect this disease early on. Early prevention to avoid this deadly heart disease. One variable that can measure or differentiate heart disease is heart rate (HR). The nature of the non-stationary heart rate signal makes the classification process cannot perform directly.

Research on heart rate has been around since 1926. Anrep (1926) writes a journal entitled The Central and reflex regulation of the heart rate [3]. Anrep tells of two mechanisms; adaptation of central nervous system that regulates the speed and strength of heart contraction; as well as heart muscle adaptation. Anrep research used two dogs to prove the mechanism.

Research continues to evolve from the medical side of how the heart works and the introduction of heart disease and how to cure it. So also with research from the engineering side of how to prevent it. Prevention is also done by research from the medical side by changing the pattern of life [4].

Research evolves so that there are pros and cons of heart rate as a variable to predict sudden death of humans. Sudden death is one of the most feared by humans. Various ways are done so that death can be predicted. One of them is Mäkikallio, (2001), in his journal entitled "Prediction of Sudden Cardiac Death by Fractal Analysis of Heart Rate Variability in Elderly Subjects" [5]. This study also has a lot of pros cons but it is important to see is how to predict using heart rate variability. Research on heart rate variability has also been done since 1949 by Malmo [6]. Malmo detects the mechanism of symptoms of stress patients using heart rate variability.

Heart disease began to be the spotlight, this is because heart disease is the biggest contributor to human death [1]. Greiser (2005) in her journal entitled "CARDA", CARDA Study used heart rate variability in determining cardiovascular disease with 1750 subjects [7].

Hjalmarson (2007) began to believe that heart rate was an independent factor against the risk of heart disease [8]. Hjalmarson described it in a journal entitled "Heart rate: an independent risk factor in cardiovascular disease". Hjalmarson looked at long-term epidemiological data showing an independent relationship between increased heart rate, cardiovascular mortality and morbidity.

Palaniappan (2008) started heart disease prediction using data mining. Palaniappan uses Naïve Bayes followed by Neural Network and Decision tree. Palaniappan's research is similar to Dangare [1] and Patil [9], using 13 variables to predict heart disease. In 2011 Vanisree, has added to 36 variables using the Decision Support System [10]. Data mining is one way of predicting heart disease in today's technological advancement era. Data obtained from the Cleveland Heart Disease database. This method is very good in determining heart disease, accuracy is very high. This is because the variable that is used very much ie 13 pieces, Dangare even try to use additional 2 other variables so that the accuracy increases to 100%. This way has a weakness, the patient must perform a series of tests to get the 13 variables so as to get good accuracy. A series of tests are now replaced by Echo or cardiac ultrasound. Echo can see the anatomical structure, function and blood flow of the heart. Heart echo takes 45 minutes to 1 hour in the echocardiography lab.

Sundar (2012) and Sellappan (2008) have reduced the variables originally used by Palaniappan into 4 variables namely age, sex, blood pressure and blood sugar [11]. Sundar uses data mining techniques to classify heart disease. These advances make the hope of predicting heart disease easy to do.

Tverdal (2008) in his journal states that the relationship between heart rate and Cardiovascular Disease (CVD) is weak to be a major risk factor for heart disease [12]. Tverdal saw an increase in heart rate as a sign of heart disease risk but not an independent factor. This is obvious because the increased heart rate can be due to prolonged stress, either due to uncontrolled emotions, or other thoughts. Increased heart rate because these thoughts can interfere with the increase of Tverdal's heart rate. In a state of sitting quietly asked not to think anything, patients still think what ordinary people think in the brain so that affect the performance of the heart.

Research on heart rate using time domain and frequency domain is done by Pitzalis (1996) [13]. Time domain and frequency domain can recognize heart disease with certain pattern. These values are also widely used by the medical world at this time. The study is continuing

In journal paper research, Peng C K Peng et al. (1995) [3] mentioned that Peng can distinguish heart disease with input time series heart rate for 24 hours using Electrocardiogram (ECG). As for Toru [3] can distinguish heart disease by input time series heart rate for 30-40 minutes using Electrocardiogram (EKG). What about continuous data? Neither Toru nor Peng can accommodate this. Patients want to see the growth over time with the direction of the trend improved or worsened.

METHODS

There is a lot of research on heart rate until 2017. However, not much discussed with long time series data. Most of the researchers recognize heart disease only from one data retrieval. In fact, the patient can feel how his body feels healthy or feel sick or will be sick. A patient is not aware of it. This will lead to the possibility of a bad patient is not known when the patient is gradually deteriorating condition of his body. With the analysis of the Reshape-Detrended Fluctuation Analysis (R-DFA) method used in timesheets over a long period, a threshold can be detected that can classify 3 classes of people. The collection of time series data in different times as much as possible is an important thing to do. Considering that humans are not always in the hospital or meeting a doctor at all times, these methods and tools used as the initial grip of the patient to see his doctor immediately when it feels sick. Polar database fill using Android phone connected with Polar H7 via Bluetooth. All data is taken using 2 Polar H7. Remove unreasonable peak in this case is eliminating the error when using the tool. Errors like this happen when the tool does not stick on the chest. So when polar move even just a little, it make a down peak as shown below.

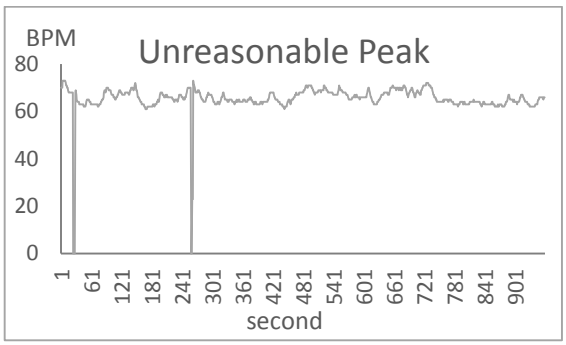

Figure 1. Remove unreasonable peak is eliminating errors when using tools. The example in the drawing is when the tool loses the connection (Lost contact) with the device.

Pre-Processing

Remove irrational data interpreted as eliminating the data taken is not possible. Usually the data has a very low variance. Even if sitting, variance should still exist. Such data is common because polar not actually attached to the chest. This is common in women. Here is one example image with irrational data.

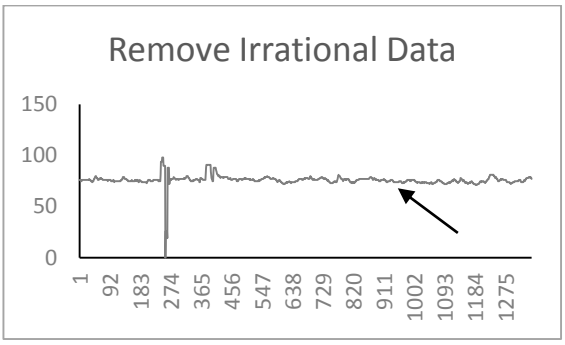

Figure 2. Remove irrational data is to eliminate impossible data. In the picture the variance is too small (variation less than 5BPM), so it is not possible.

When subject sit in the chair, variation of heart rate value around 12BPM. When subject run with 2km/hour and 4km/hour variation is around 38,9BPM. It is impossible if the variation less than 5BPM in sit position. In Figure 2, variation is less than 5BPM, so the data cannot be used.

Learning

Clustering data done by dividing every 15 minutes of data retrieval into 4 parts. Tagging data done by labeling each data. Checking the data means to verify the entered data is correct or wrong. The plotting process is re-done to ensure that data entered correctly by looking through the graph. Threshold performed repeatedly to get the most optimal value to determine the 3-class recognition process.

Detection

In the detection section, testing data is taken from RRI data that took using android phone and processed like Preprocessing Process, then begins detection process by inserting threshold, which have been obtained during learning process. The classification process generates class data then described in the form of data plots.

Data Specification

Data heart rate taken in similar environment. Data taken using Polar H7 in sitting position. There are 6 sample, 3 people has hypertension heart disease (HHD), 3 people has Mild Left Ventricle Diastolic Dysfunction (MLVDD) and 3 is healthy. All sample tested by echocardiogram by professional or doctor.

Each data taken in 15 minutes of retrieval. Each people taken 100 times retrieval. In total, there are 900 data. Data taken using Polar H7 and android phone. The output in BPM format.

Terms of taking data on each subject is the same that is,

1) Subjects in a state of sitting still
2) Subjects may not speak during recording
3) Subjects are not in a state of discharged activities such as sports, climbing stairs or other activities that are significant heart rate increase
4) The subject must be silent for 15 minutes
5) Subject uses polar on the left chest.

Steps of data retrieval

1) Polar mounted on the subject's chest. The installation method seen in Chapter II.3.
2) Open the Smart Health System As An Open Platform (SHESOP) application on your android phone, enter the name in the User name field
3) Connect apps with polar using Bluetooth
4) Start the app by pressing play button
5) When finished click the stop button and upload button to upload the data you just fetched to the dropbox.

Data collected using two polar. Polar has a battery power source with code C2025. This battery will be discharged in 5 or more usage. When the battery runs out, the application will be difficult to connect with polar, or heart rate value zero.

Research data is filtered using the Pre-processing process. Remove unreasonable peak, Remove irrational data due to errors in the installation of the tool. Remove irrational data is an event that could occur due to equipment breakdown or battery depletion.

In contrast to accurate ECG data, ECG data should be properly addressed. This can lead to calculation errors in the next process. In the paper Xiaopei [1], which uses wearable sensors for ambulance purposes taking data using android. Xiaopei reduced data transmission from 1.5 seconds to 0.75 seconds for new data. Intake of data made Xiaopei in addition must be fast must also be accurate, otherwise there will be error reading and fatal to the patient. Likewise with R-DFA, if the initial reading is wrong, then the result will also be wrong.

This is sample result of Polar H7 device. Figure 3 shows that BPM versus time of healthy data. Figure 3 is one of 300 retrieval data. Figure 4 show the Hypertension Heart Disease is one of 300 retrieval data, and so does Mild Left Ventricle Diastolic Dysfunction.

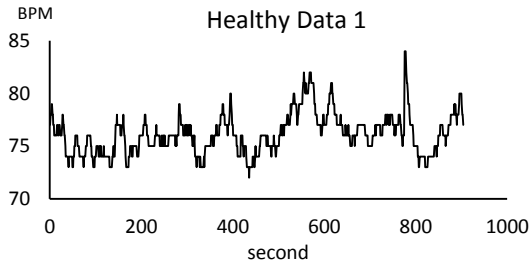

Figure 3. One of 300 Healthy Data

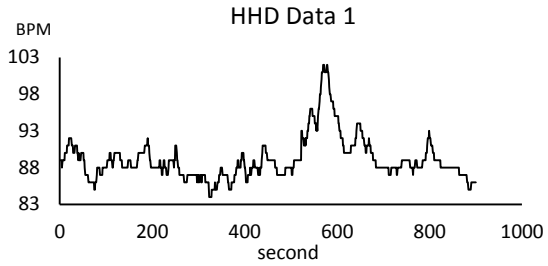

Figure 4. One of 300 HHD Data

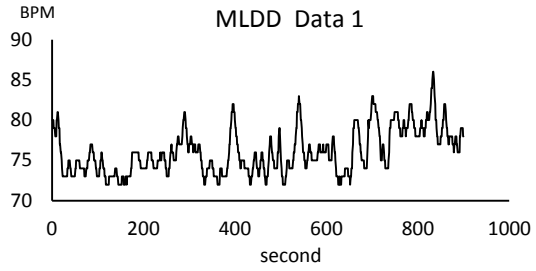

Figure 5. One of 300 MLVDD Data

Polar H7 take heart rate data every 1 second. We already take data using Fitbit on the same people, but it has not analyzed yet. It will be analyze in the next paper.

Table 1: Subject Database

Database	Specification
Healthy	3 women 30-35 years
Hypertension Heart Disease	3 women 55-60 years
Mild Left Ventrikel Diastolic Dysfunction	3 men and women 55-60 years

Detrended Fluctuation Analysis

Peng's DFA

Peng's DFA present a possibility for characterizing this as a function of correlation distance [2]. First integrate all the time series $x(i)$, where $i = 1 \ldots N$.

$$y(k) = \sum_{i=1}^{k} (x(i) - \bar{x}),$$

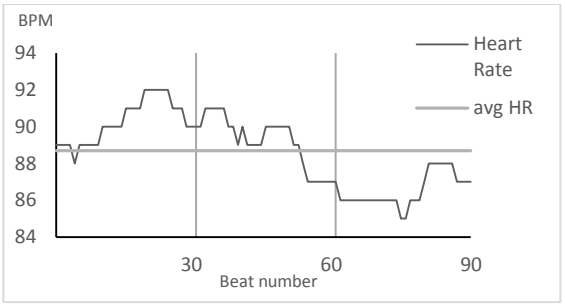

Figure 6. Sample Heart rate with 90 beat and average of Heart rate.

In the Figure 6, heart rate is fluctuate between 85 and 92. There is line $\bar{x} = 89$ mean of time series $x(i)$. N is total length of the graph $N = 90$. Next step is to count a new random walk $x_k = x(i) - \bar{x}$ where $k = 1..N$. Count q_i, so that :

$$q_i = \sum_{k=1}^{i} (x_k - \bar{x})$$

Next step is detrended phase, counting new random walk S_i so that $S_i = q_i - q_j$. q_j is the value of regression line.

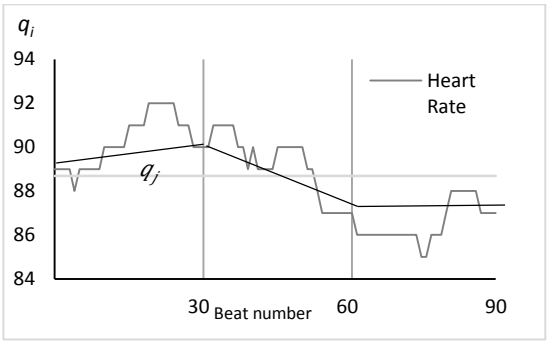

Figure 7. Detrended phase $S_i = q_i - q_j$

Detrended phase is to have regression q_j became one straight line.

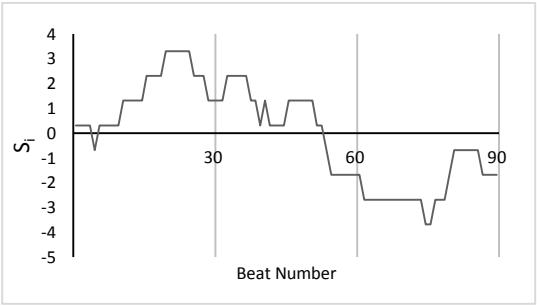

Figure 8. Random walk S_i

If $y_k = q_i$ with $i = k$ and $y\acute{}_k = q_j$. Next, divide the time series y_k into segment with length n. Calculate local trend by fitting regression line $y\acute{}_k$ for every segment. DFA count by using this formula:

$$DFA(n) = \sqrt{\frac{1}{N}\sum_{i=1}^{k}[y_k - \acute{y}_k]^2}$$

DFA often presented as log-log plot of $DFA(n)$ and n. DFA value often use $\propto_1$ and $\propto_2$ to compare to other data. $\propto_1$ and $\propto_2$ can be define as

$$\propto_1 (n) = \log(DFA(n))$$

$$\propto_2 (n) = \log(n)$$

R-DFA

R-DFA is continuation of Peng's DFA. After DFA finishing log-log plot, R-DFA is subtracting $\propto_1$ and $\propto_2$ in every 3.75 minutes and average it. The window of one time recording is 15 minutes. There is 4 region in 15 minutes data it means 3.75 minute per part. This is the R-DFA flow until it gets the classification of three class.

Research stages with the R-DFA method that produces the threshold to determine the 3 classes (pain, illness and health) can be seen in the Figure 9. Block diagram of R-DFA can be seen in the Figure 10.

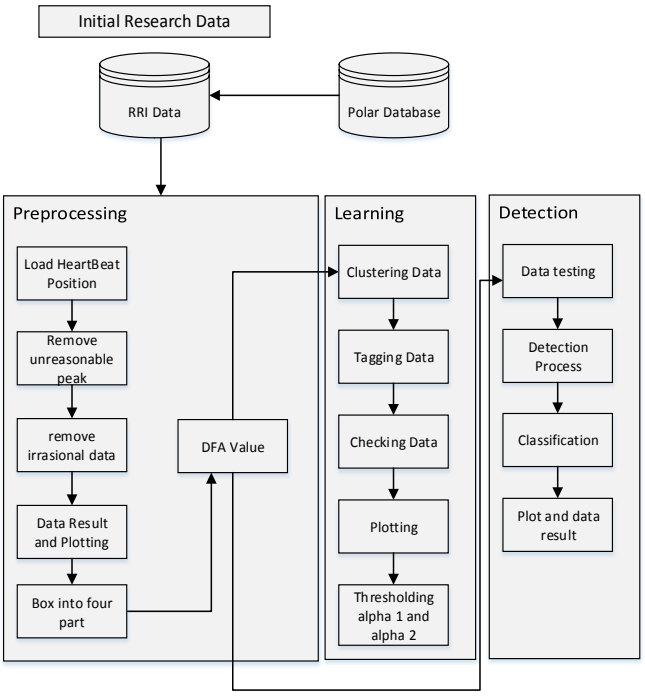

Figure 9. Research stages with R-DFA method starts with polar database, RRI data, Pre-processing, Learning and Detection.

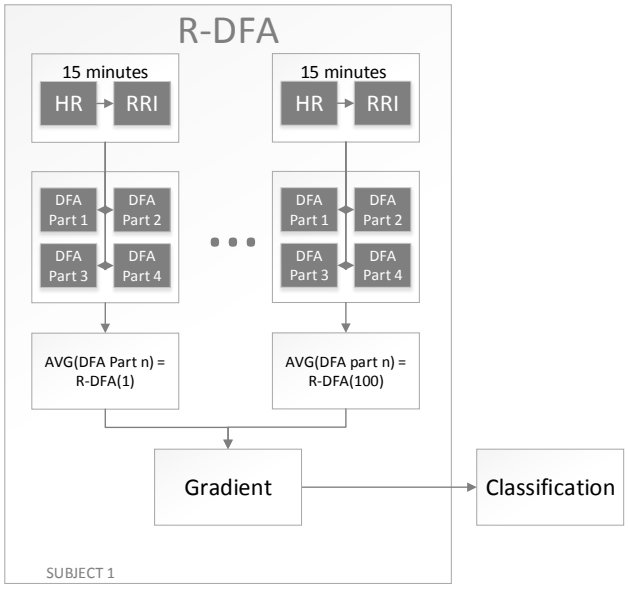

Figure 10: The block diagram of R-DFA uses a gradient from R-DFA 1 to n. Gradients can show trends or tendencies of the subject.

The idea of R-DFA is to see differences between 3 classes through a regression of DFA value that can't be seen in one time retrieval DFA. R-DFA is a way of using in accurate data heart rate to have a significant result.

Different from R-DFA, Chandrakar Kammath with ECG data in 2015 in one time retrieval can distinguish Congestif Heart Failure and healthy person with 98.2% with $\propto_1$ as feature [17]. ECG data is the most used data in the research.

RESULT

After 100 times retrieval of each patient (9 patient), the result is quite significant with $p<0.05$. Here it is the graph of three gradient classification with R-DFA.

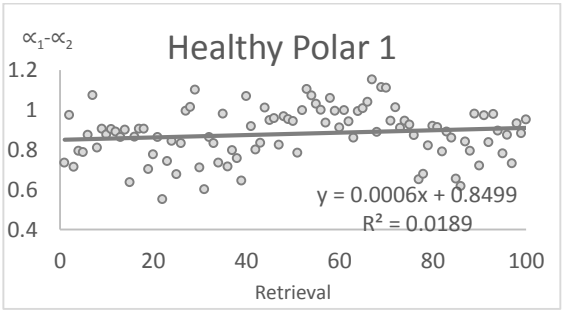

Figure 11. Healthy patient 1 with trend line y = 0.0006x + 0.8499

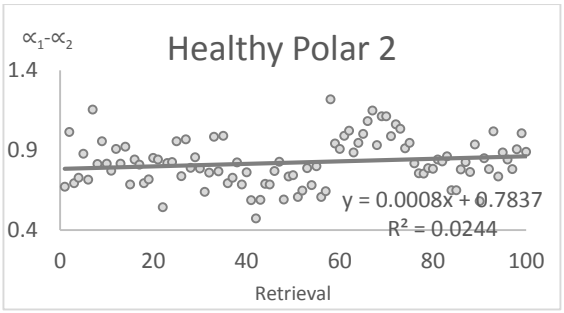

Figure 12. Healthy patient 2 with trend line y = 0.0008x + 0.7837

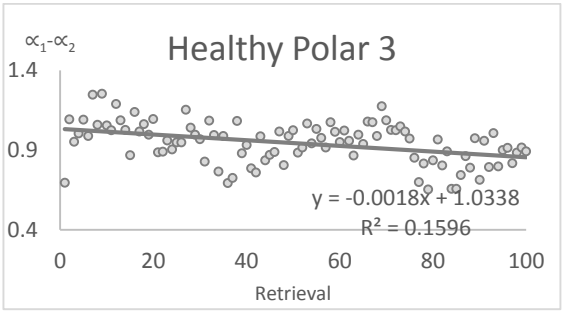

Figure 13. Healthy patient 3 with trend line y = -0.0018x + 1.0338

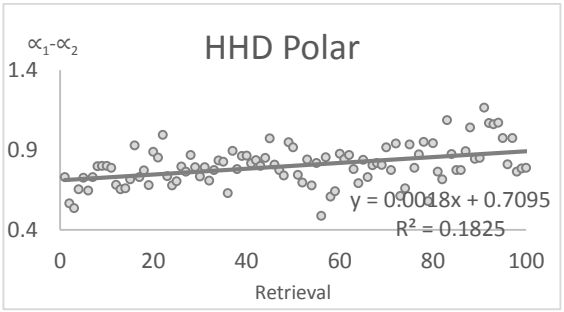

Figure 14: Hypertension Heart Disease patient 1 with trend line y = 0.0018x + 0.7095

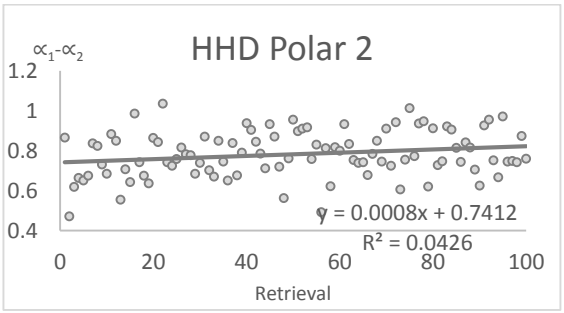

Figure 15: Hypertension Heart Disease patient 2 with trend line y = 0.0008x + 0.7412

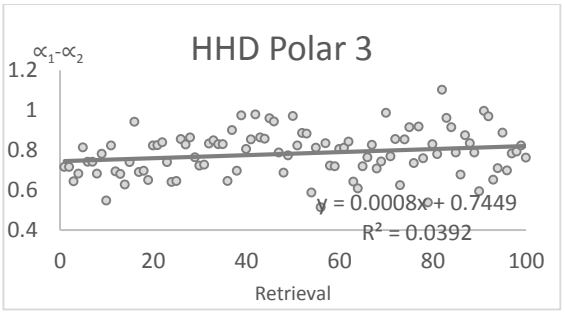

Figure 16: Hypertension Heart Disease patient 3 with trend line y = 0.0008x + 0.7449

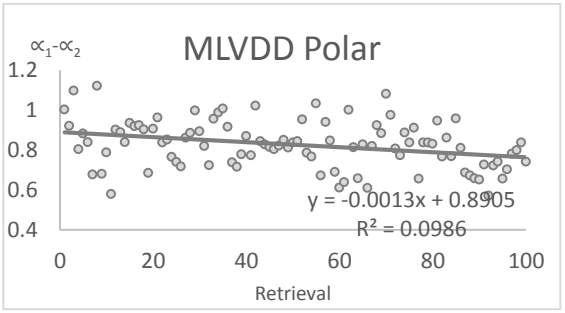

Figure 17: Mild Left Ventricle Diastolic Dysfunction patient 1 with trend line y = -0.0013x + 0.8905

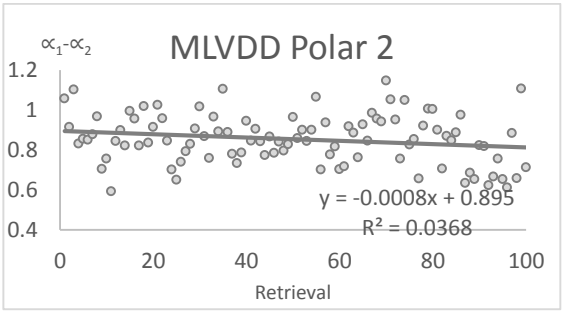

Figure 18: Mild Left Ventricle Diastolic Dysfunction patient 2 with trend line y = -0.0008x + 0.895

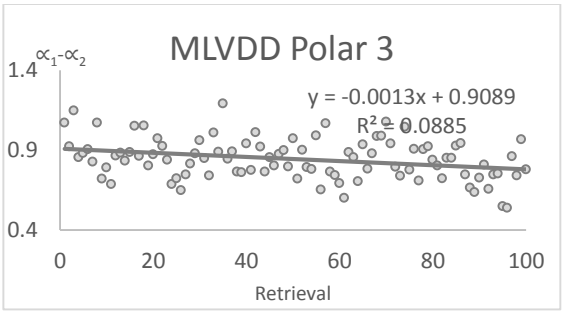

Figure 19: Mild Left Ventricle Diastolic Dysfunction patient 3 with trend line y = -0.0013x + 0.9089

9 patient with the trend line can remove noise that occur during the retrieval. Sometimes when retrieval patient accidentally thinking or up and down temperature, or wind, or something that might have ruin the data. It makes heart rate fluctuate often.

When patient are thinking something unhappy, it will increase the heart rate. Here it is table of gradient lines or trend of three classes.

Table 2: Table of gradient lines healthy, MLVDD, and HHD classes.

	Formula regresi	m(R-DFA) *10000
Healthy 1	y = 0.0006x + 0.8499	6
Healthy 2	y = 0.0008x + 0.7837	8
Healthy 3	y = -0.0018x + 1.0338	-18
HHD 1	y = 0.0018x + 0.7095	18
HHD 2	y = 0.0008x + 0.7412	8
HHD 3	y = 0.0008x + 0.7452	8
Mild LVDD	y = -0.0013x + 0.8905	-13
Mild LVDD 2	y = -0.0008x + 0.895	-8
Mild LVDD 3	y = -0.0013x + 0.9089	-13

$m(R-DFA)$ is the threshold criteria of three classes. This threshold has been set depends on trend line data. This threshold can be update along with patient number growth.

Table 3: Criteria of threshold of three classes

	Kriteria		
	-8<healthy<8 and healthy ≥-18	HHD≥ 8	-18< MLVDD < -8
Healthy 1	True	False	False
Healthy 2	False	True	False
Healthy 3	True	False	False
HHD 1	False	True	False
HHD 2	False	True	False
HHD 3	False	True	False
Mild LVDD	False	False	True
Mild LVDD 2	False	False	True
Mild LVDD 3	False	False	True

There are 27 data with false and true data based on threshold criteria. There are two false data, healthy criteria on healthy 2 and HHD criteria on healthy 2. This means the accuracy of 92% with three classes (HHD, MLVDD, and healthy). Out of 900 data it means 92% x 900 = 828 data is correct and 72 data classification errors.

Testing the classification by t-test between healthy subjects with HHD subjects, healthy subjects with MLVDD.

Table 4: T-test of healthy patient with HHD

	Healthy	*HHD*
Mean	0.880400442	0.8011633
Variance	0.016337526	0.015206711
Observations	100	100
Hypothesized Mean Difference	0	
df	198	
t Stat	4.461375546	
P(T<=t) one-tail	6.82162E-06	
t Critical one-tail	1.652585784	
P(T<=t) two-tail	1.36432E-05	p<0.05 (SIGNIFICANT)
t Critical two-tail	1.972017478	

The result of T-test of healthy patient with HHD is significant with p<0.05. Next is to test the healthy patient with the MLVDD.

Table 5: T-test of healthy patient with MLVDD

	Healthy	*MLVDD*
Mean	0.880400442	0.826294
Variance	0.016337526	0.013791
Observations	100	100
Hypothesized Mean Difference	0	
df	197	
t Stat	3.117140896	
P(T<=t) one-tail	0.001049663	
t Critical one-tail	1.652625219	
P(T<=t) two-tail	0.002099327	P<0.05 (SIGNIFICANT)
t Critical two-tail	1.972079034	

The result of T-test of healthy patient with MLVDD is significant with $p<0.05$. To make sure the data, Peng's DFA applied to R-DFA data. and this is the result.

Table 6: Comparison of R-DFA data applied to DFA Peng

HHD Patient		DFA Peng
Alpha 1	Alpha 2	
1.4946	0.71556	Close to healthy
1.3966	0.61797	Close to healthy
1.2702	0.67101	Close to healthy
1.5533	0.56288	Close to healthy

CONCLUSION

R-DFA data that has been taken using polar data has a less accuracy compared to ECG data. So when R-DFA data applied to DFA Peng, HHD patient judge close to healthy.

The result is Threshold Line regression of gradient HHD≥ 8, -8<healthy<8 and healthy≥-18, -18<MLVDD< -8. This threshold will classified HHD, MLVDD, and healthy person.

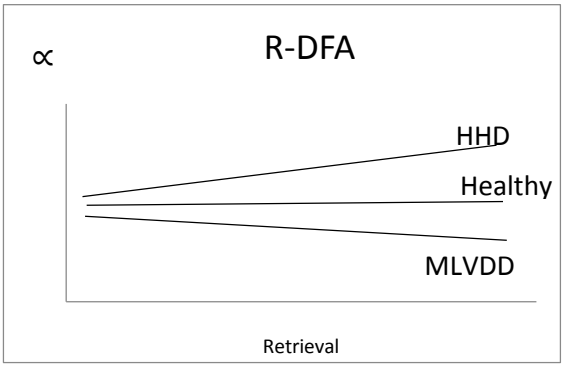

Figure 20: Threshold Regression

REFERENCES

[1] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "A Data Mining Approach For Prediction of Heart Disease Using Neural Networks," International Journal of Computer Engineering & Technology (IJCET), pp. 30-40, 2012.

[2] T. C. Timmreck, Epidemiologi: Suatu Pengantar (Terjemahan), Jakarta: Penerbit Buku Kedokteran EGC, 2004.

[3] G. V. a. S. H. N. Anrep, "The central and reflex regulation of the heart rate," The Journal of physiology, Vols. 215-231, p. 61(2), 1926.

[4] J. Draft, Cardiovascular Disease, nowsley JSNA Report, 1947.

[5] Timo H. Ma"kikallio, MD, Heikki V. Huikuri, MD, FACC, Anne Ma¨kikallio, MD et. al., "Prediction of Sudden Cardiac Death by Fractal Analysis of Heart Rate Variability in Elderly Subjects," Journal of the American College of Cardiology, vol. 37, pp. 1395-1402, 2001.

[6] R. B. a. S. C. Malmo, "Physiologic study of symptom mechanisms in psychiatric patients under stress," Psychosomatic Medicine, vol. 11(1), pp. 25-29, 1949.

[7] Greiser, K. H., Kluttig, A., Schumann, B., Kors, J. A., Swenne, C. A., Kuss, O., Werdan, K., and Haerting, J., "Cardiovascular disease, risk factors and heart rate variability in the elderly general population: Design and objectives of the CARdiovascular disease, Living and Ageing in Halle (CARLA) Study," BMC Cardiovascular Disorders, vol. 5, p. 33, 2005.

[8] A. Hjalmarson, "Heart rate: an independent risk factor in cardiovascular diseas," European Heart Journal Supplements, vol. 9(suppl_F), pp. F3-F7, 2007.

[9] Patil, S. B., and Kumaraswamy, Y. S., "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network," European Journal of Scientific Research, vol. 31(4), pp. 642-656, 2009.

[10] K. &. S. J. Vanisree, "Decision Support System for Congenital Heart," International Journal of Computer Applications, vol. 19, p. 0975–8887, 2011.

[11] Sundar, N. A., Latha, P. P., and Chandra, M. R., "Performance Analysis Of Classification Data Mining Techniques Over Heart Disease Data Base," nternational Journal Of Engineering Science & Advanced Technology, pp. 2250-3676, 2012.

[12] Tverdal, A., Hjellvik, V., and Selmer, R., "Heart rate and mortality from cardiovascular causes: a 12 year follow-up study of 379 843 men and women aged 40–45 years," European Heart Journal, vol. 29(22), pp. 2772-2781, 2008.

[13] Pitzalis, M. V., Mastropasqua, F., Massari, F., Forleo, C., Di Maggio, M., Passantino, A., Colombo, R., Biasse, M. D., and Rizzon, P, "Short-and long-term reproducibility of time and frequency domain heart rate variability measurements in normal subjects," Cardiovascular research, vol. 32(2), pp. 226-233, 1996.

[14] C K Peng, Shlomo Havlin, H. Eugene Stanley, Ary. L Goldberger, "Quantification of scaling exponent and crossover phenomena in nonstationary heartbeat timeseries," CHAOS, vol. 5, 1995.

[15] T. Yazawa, "Modified Detrended Fluctuation Analysis (mDFA)," in Modified Detrended Fluctuation Analysis (mDFA), The American Society of Mechanical Engineers, 2015.

[16] Xiaopei Wu, Robert Dunne, Zhifeng Yu, Weisong Shi, "STREMS: A Smart Real-Time Solution toward Enhancing EMS Prehospital Quality," in International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Philadelphia, 2017.

[17] Kamath, M. V., Watanabe, M., and Upton, A. (2013): Heart rate variability (HRV) signal analysis: clinical applications. Boca Raton: CRC Press.