JOURNAL OF DATA SCIENCE AND ITS APPLICATIONS

Classification of Electrocardiogram Signals using Principal Component Analysis and Levenberg Marquardt Backpropagation for Detection Ventricular Tachyarrhythmia

Astrima Manik¹, Adiwijaya², Dody Qori Utama³

School of Computing, Telkom University

Jalan Telekomunikasi No.1 Bandung, Indonesia

¹ astrimamanik@students.telkomuniversity.ac.id

² adiwijaya@telkomuniversity.ac.id

³ dodygori@telkomuniversity.ac.id

Received on 22-05-2018, revised on 08-02-2019, accepted on 09-04-2019

Abstract

Ventricular Tachyarrhythmia (VT) are the primary arrhythmias which are cause of sudden death. For someone who already has symptoms of VT should immediately perform an examination of one of them by using an electrocardiogram (ECG). An electrocardiogram is a recording of the heart's electrical results in a waveform. However, limited ability in analysis and diagnosis of ECG reading is still difficult to do. Therefore, the classification of ECG signals is needed to detect a person, especially those with VT or not. This research focuses on the classification of VT heartbeats from ECG signals by using median filter method in preprocessing, Principal Component Analysis (PCA) technique in the characteristic extraction and modified Backpropagation (MBP) as classification. Moreover, the performance effects of several dimension reduction approach tested with PCA. This research used machine learning method that is a neural network with backpropagation modification that is Levenberg Marquardt to speed up network training process. The best VT detection performance results were based on the average accuracy of the overall scheme of 91.67% with the best parameters that principal component=10 and 20, hidden neuron=4, and μ value=0.001 as well training time 1 seconds with a comparison of train data and test data that is 80:20 percent.

Keywords: Electrocardiogram, Levenberg Marquardt Backpropagation, Median filter, Principal Component Analysis, and Ventricular Tachyarrhythmia

I. INTRODUCTION

The heart is one of the human organs that have the role of the most important function is to pump blood throughout the body. When the heart pumps blood throughout the body, the heart muscles contract and produce heartbeats. Heartbeats have a frequency of rhythm due to the action potential generated by the heart.

The heart rhythm has the meaning that the heart beats normally or not. A normal heartbeat indicates a heart condition that can pump blood optimally throughout the body. While abnormal heart rate there is a symptom in the heart that can lead to death. In Indonesia, the death rate caused by heart disease reaches 26 to

30 percent. Based on data presented World Heart Federation (WHF), heart disease reached 29.1 percent or as many as 17.1 million patients each year died worldwide. Heart disease that causes death is one of them is an arrhythmia.

The incidence of arrhythmias will increase with age. It is estimated that the geriatric population (elderly) will reach 11.39% in Indonesia or 28 million people in Indonesia by 2020. Increasingly aging, the percentage of arrhythmia incidence is increasing, is 70% at age 65-85 years and 84% over 85 years [7,12]

Arrhythmias are heart problems that occur when the organ is beating too fast, too slow, or irregularly. Rapid heartbeats are grouped into the type of tachycardia, while slow heartbeats fall into the bradycardia type. Arrhythmias are also grouped into two major groups of atrial and ventricular types. Grouped in atrial type when abnormalities occur in the atrium. In contrast, ventricular types occur when abnormalities occur in the ventricular part. Many types of arrhythmias are included in the type, but this study is limited to ventricular type arrhythmias with excessive heart rhythms called Ventricular Tachyarrhythmia (VT). VT are dangerous arrhythmic events leading to inevitable death if no defibrillation shock is applied to the subject within a few minutes [6].

In recent years, many experiments have been conducted to detect ventricular tachyarrhythmia disease. One of them using the digital medical image. Digital medical images and medical records of patients have been widely practiced medicine through computer networks for use in the practice of a person's health examination [5]. Digital medical images can be obtained in several ways such as research [6] providing digital image compression using the quantization graph coloring adopted from the VQ principle. One of the digital medical images used in these cardiac signal checks is the electrocardiogram (ECG). ECG device can be used to detect heart disease likes VT symptoms.

In this paper, an effective and comparative approach was developed for the classification of ventricular tachyarrhythmias. The ultimate goal is to detect whether a person has normal VT or cardiovascular disease and improve the accuracy of the VT classification. This study used PCA to reduce the sample feature size taken from QRS waves. Neural networks are used in the classification with backpropagation modifications of Levenberg Marquardt to speed up the training process on network systems. The results showed that the method used obtained a high degree of accuracy 91.67% with the selected parameters.

II. LITERATURE REVIEW

A. Research on ECG

Recently, much research has been done for the classification of heart disease on ECG signals. Zumray and Tamer (2011) present a hybrid neural network for ECG signal classification, in which the authors use two extraction methods of the fourier and wavelet features. Elif (2010) uses backpropagation neural network training with Levenberg-Marquardt algorithm for classification of ECG signal. Arrhythmia classification using PCA as a selection feature and Elman Neural Network as a classification method produces accuracy of 95% [2], to improve computational efficiency, the reduction feature uses PCA in order to eliminate both low significance and redundancy features [3]. Based on the above research sources, the authors chose to use PCA method in extracting ECG signal characteristics and the extraction results of these characteristics will be identified by the ANN Backpropagation algorithm. In the training process using levenberg marquardt backpropagation [11]. The identification results will reveal the ECG signal condition is classified as VT or normal disease.

B. Ventricular Tachyarrhythmia (VT)

Ventricular Tachyarrhythmia (VT) is an arrhythmia disease that has a rapid heart rate located at the bottom of the heart (ventricular). There are three types of VT disease that Ventricular Tachycardia, Ventricular Flutters, and Ventricular Fibrillation [1].

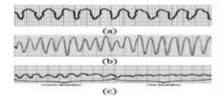


Fig 1. (a) Ventricular Tachycardia, (b) Ventricular Flutter, (c) Ventricular Fibrillation [1]

Based on Figure 1, there is the main characteristic that stands out from this disease that is the wave of the QRS complex widens and rises with the rhythm that sometimes regularly but also irregular[1].

C. Electrocardiogram (ECG)

Electrocardiogram (EKG) is an illustration of the electrical potential generated by the electrical activity of the heart muscle. The electrical activity's basic pattern consists of three waves: P, QRS (complex wave), and T [9]. Wave P is a result of atrial contraction that pumps blood into the ventricle. QRS wave is a result of ventricle contraction that pumps blood into the rest of the body. Figure 2 shows that QRS has the highest amplitude due to the immense amount of energy produced. Meanwhile, T wave is a result of ventricle relaxation process when the contraction ends and blood begins to be pumped from atrium into the ventricle [8].

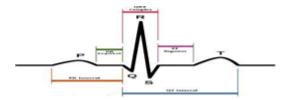


Fig 2. Waveforms of ECG Signals [8]

III. RESEARCH METHOD

Research is conducted through implementing ANN in Matlab software. VT detection design consists of three main phases shown in figure 3.

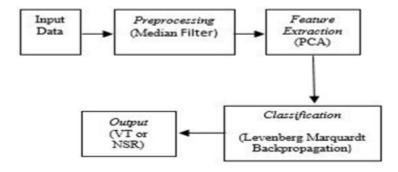


Fig 3. System Design Diagram

A. Data Input

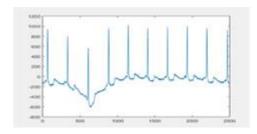
This research uses input ECG record data obtained from data ready from a Physionet web [4], namely CU Ventricular Tachyarrhythmia Database (cubd) and MIT-BIH Normal Sinus Rhythm Database (NSRB) in the .mat extensionfile.

TABLE I
DATA INPUT

	Jumlah Record	Jumlah Sampel
Data Penyakit VT	34 records	@2500 sampel
Data NSR	18 records	@1280 sampel

B. Pre-processing

The data used in this research is taken from CU Ventricular Tachyarrhythmia Database and MIT-BIH Normal Sinus Rhythm Database [4]. The first stage performed on the signal data is preprocessing. Preprocessing is a process of processing the original / raw data into more quality data. The median filter is one method of preprocessing that can eliminate noise such as baseline wander. Baseline wander a noise that affects ECG signals, as typically descending and rising and not consistently in the line of isoline or zero lines causing difficulty algorithms to detect peak R [3]. The advantage of using this median filter is to eliminate and reduce noise, it also has advantages during the filter process without reducing even retaining important information from the ECG signal. This can help for the next process in carrying out feature extraction and detecting heart disease. The median filter used as many as 15 data filters that will produce the form of the signal looks smoother.



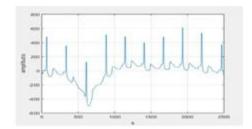


Fig 4. Before pre-processing

Fig 5. After pre-processing

C. Feature Extraction

PCA is one of the feature extraction methods used to convert huge correlated native variables into a single set of new, smaller, independent variables that principal component without significantly reducing data characteristics [2]. In ECG signals there is a repetitive signal pattern so that it is difficult to detect signal pattern characteristics from a heart disease, therefore PCA is very necessary to use because it can detect signal characteristics in the disease by reducing samples on ECG signals. So that by using PCA, it is hoped that input data can support system pe rformance to work optimally [5, 8]. PCA can detect QRS waves by reducing the number of samples exceeding the limit > 200 || <-200, because the main characteristic of VT disease is the QRS wave that passes the limit [1]. The sampling resulted in a varying number of samples per record. Differences in the number of samples will be uniformed by taking the least number of samples per record. Uniformity is done because the PCA process can only be used in the same sample. The final process of this PCA stage produces a new variable that is principal component (PC). In the feature extraction process of this research do not use other methods such as linear discriminant analysis, linear preserving projection or the others because the PCA method signals can be reduced to small components that can detect R waves according to the characteristics of ventricular tachyarrhitmia heart signals. In this research, 10, 20, and 30 are the PC parameter values whose accuracy have tested.

D. ANN Backpropagation (Levenberg Marquardt)

Backpropagation (BP) algorithm is one example of ANN Multi-Layer Perceptron (MLP). Backpropagation is used to train the network to gain balance in recognizing the patterns used during the training and an apparent response to input patterns similar to the training pattern. In the training process using

Levenberg Marquardt algorithm to initiate the scalar μ value to speed up the training process. The author uses the Levenberg Marquardt (LM) method with multilayer neural networks because the LM training method requires a smaller number of iterations than the training backpropagation algorithm (BPA) method in achieving minimum errors. In this research, 0.0010, 0.0012, and 0.0014 are the parameter μ values whose accuracy have tested. Besides that, there are hidden neurons that are used whose accuracy have tested that 3, 4, and 5.

IV. RESULTS AND DISCUSSION

To acquire ECG feature extraction signals, QRS weak detection is first conducted, shown in figure 6, figure 7 and figure 8.

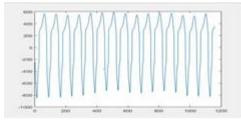


Fig 6. Detection of QRS Wave

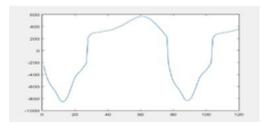


Fig 7. Cutting to 120 sample features

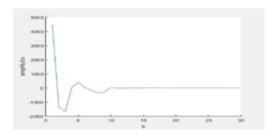


Fig 8. Feature Extraction using PC=30

Based on input data, there were 2500 samples / 34 records in VT data and 1280 samples / 18 records in NSR data. After preprocessing using the median filter, a variety of sample size features were obtained. The number of diverse samples makes the sample feature size not the same so that it will complicate the reduction process because input at the PCA stage requires a uniform sample size feature. Therefore, the author will equate the size of the sample feature by taking the least sample to guide the size of the sample feature. As seen result of median filter, the least number of samples is 120 samples. So, the authors set the sample feature size for each record to be 120 samples. All samples will be deducted by 120 samples and records that have more than 120 samples, the rest will be used as new records. The process is also carried out on NSR data. So that later all data will have a feature size of 120 samples in each record. The following are the record results and samples obtained by each class after the sample is cut:

TABLE II
RESULTS OF RECORD AMOUNTS AND SAMPLES AFTER CUTTING SAMPLES

	Jumlah Record	Jumlah Sampel
Data VT	180	@120
Data NSR	180	@120

Detection of QRS waves on VT can be done by taking samples between > 200 or < -200 because R peak passes the limit. The sampling results can be seen in Fig.6 where the QRS wave is generated or widened. The sampling will produce different samples of each data, therefore the uniformity of the sample is done by cutting the sample to 120 samples per record. after that, the data will be reduced to the sample feature to the

specified PC. The result of sample cutting and sample feature reduction with PCA can be seen in Fig.7 and Fig.8.

Several data testing scenarios based on train or test data split using parameter principal component=10, 20, and 30, hidden neuron=3, 4, and 5, μ value=0.0010, 0.0012, and 0.0014 with 180 data normal and 180 data VT are used in VT detection. Train or test data split is a method used to divide general data into training and testing data. This research uses three data division schemes: training and testing data with each 50:50 percent in scheme 1, 60:40 percent in scheme 2, and 80:20 percent in scheme 3 at random. This division is done to test research method performance for several data compositions with a comparison of training data greater than testing data, test data greater than training data, and equally Thus, each scheme has different accuracy levels that can be compared to obtain the scheme with the best accuracy. Comparison of training data and test data on the scheme can be seen in table 3, table 4, and table 5.

TABLE III SYSTEM ACCURACY BASED ON SCHEME 1 TABLE IV SYSTEM ACCURACY BASED ON SCHEME 2

60:40 Percent

Scheme 2

TABLE V SYSTEM ACCURACY BASED ON SCHEME 2

Scheme 1.		50:50 Percent	
PC	Hidden Neuron	μ Value	Accuracy
		0,001	80,56%
	3	0,0012	82,22%
		0,0014	79,44%
		0,001	77,78%
10	4	0,0012	81,67%
		0,0014	87,78%
		0,001	75%
	5	0,0012	80,56%
		0,0014	78,33%
		0,001	70,56%
	3	0,0012	78,33%
		0,0014	80,56%
		0,001	80%
20	4	0,0012	80%
		0,0014	80%
		0,001	80%
	5	0,0012	75%
		0,0014	74,44%
		0,001	81,67%
30	3	0,0012	62,78%
		0,0014	66,11%

PC	Hidden Neuron	μ Value	Accuracy
		0,001	84,03%
	3	0,0012	81,94%
		0,0014	89,58%
		0,001	86,81%
10	4	0,0012	87,50%
		0,0014	84,72%
		0,001	74,31%
	5	0,0012	77,08%
		0,0014	79,86%
		0,001	77,08%
	3	0,0012	79,17%
		0,0014	77,08%
		0,001	75,69%
20	4	0,0012	76,39%
		0,0014	76,39%
		0,001	77,78%
	5	0,0012	81,25%
		0,0014	78,47%
		0,001	58,33%
30	3	0,0012	76,39%
		0,0014	64,58%

Scheme 3.		80:20 Percent	
Hidden Neuron	μ Value	Accuracy	
	0,001	83,33%	
3	0,0012	81,94%	
	0,0014	86,11%	
	0,001	91,67%	
4	0,0012	87,50%	
	0,0014	87,50%	
	0,001	88,89%	
5	0,0012	86,11%	
	0,0014	80,56%	
	0,001	86,11%	
3	0,0012	79,17%	
	0,0014	77,78%	
_	0,001	91,67%	
4	0,0012	90,28%	
	0,0014	88,89%	
_	0,001	86,11%	
5	0,0012	79,17%	
	0,0014	77,78%	
	0,001	77,78%	
3	0,0012	84,72%	
	0,0014	87,50%	
	Hidden Neuron 3 4 5	Hidden Neuron	

		0,001	75,56%
	4	0,0012	78,33%
		0,0014	78,89%
		0,001	70,56%
	5	0,0012	70%
		0,0014	67,22%
AVRG		76,79%	

		0,001	69,44%
	4	0,0012	78,47%
		0,0014	68,06%
		0,001	74,31%
	5	0,0012	77,78%
		0,0014	79,86%
AVRG		1	77,49%

		0,001	69,44%
		-,	**,****
	4	0,0012	79,17%
		0,0014	75%
		0,001	77,78%
	5	0,0012	77,78%
		0,0014	76,39%
	AVRG		82,82%

Proportion data on scheme 1 with 50 percent of training data and 50 percent of testing data. By looking at the result above, the best PC, hidden neuron and μ value when PC=10, hidden neuron=4, and μ value= 0.0014 with accuracy is 87.78%.

Proportion data on scheme 2 with 60 percent of training data and 40 percent of testing data. By looking at the result above, the best PC, hidden neuron and μ value when PC=10, hidden neuron=3, and μ value= 0.0014 with accuracy is 89.58%.

Proportion data on scheme 3 with 80 percent of training data and 20 percent of testing data. By looking at the result above, the best PC, hidden neuron and μ value when PC=10 and 20, hidden neuron=4, and μ value= 0.001 with accuracy is 91.67%.

Different levels of accuracy obtained for the three dataset division schemes on this research. Out of overall data, the scheme 3 have the highest accuracy at 91.67% with PC=10 and 20, hidden neuron=4, μ value=0,001 and training:testing data at 80:20 percent. Scheme 3 also has the highest average accuracy compared to other schemes at 82.82%. In scheme 1 and scheme 2 only produce of average accuracy at 76.79% and 77.49%.

Several factors that influence the system results are:

1) The process of cutting data signal and QRS wave detection

The detection of QRS waves is done manually by taking samples at R peak that exceeds the upper limit of 200 or below -200. Though QRS waves can be detected in other ways. In addition, the sample cutting into 120 samples is also very manual, because the least number of samples obtained at that time as many as 120 samples. It happens to be done sampling uniform on PCA process. But it can not be denied, that there are other ways that can be used in addition to cutting the sample is by adding a value of 0 on the sample of each record. So, the sample size will be uniform.

2) Parameter of testing

The parameters used may affect the accuracy. In this study only use the parameters specified are PC=10, 20, and 30, hidden neurons=3, 4, and 5, the μ value=0.0010, 0.0012, and 0.0014. Determination of the parameter value is obtained by trial and error. The next research can use other parameter values to improve performance.

3) Dataset division

Test and test data are divided into three schemes using trial / trial. The average accuracy for schemes 1, 2, and 3 was 76.79%, 77.49%, and 82.82%. It seems that the more training data used, the greater the probability of system accuracy. Proven on the outcome of scheme 3 with training and testing division data at 80:20 percent has the highest accuracy of 91.67% with PC=10 and 20, hidden neuron=4, μ value=0.001. Scheme 2 has the highest accuracy of 89.58% with PC=10, hidden neuron=3 and 20:80 percent dataset. Scheme 1 has the highest accuracy rate at 87.78% with PC=10, hidden neuron=4, μ value=0.0014 and 50:50 percent dataset.

There are some recommendations for further system development in detecting VT based on the test results analysis:

- 1. Using other methods for preprocessing, feature extraction, and classification.
- 2. In this research, author does signal cutting method so manually, therefore, developing better and more accurate signal cutting method.
- 3. Using other parameters to produce precise accuracy.

V. CONCLUSION

In this study, VT detection system based on feature extraction and classification. The system structure is obtained by using PCA method and levenberg marquardt backpropagation. This study uses PC parameter=10, 20, and 30, hidden neurons=3, 4, and 5, the μ value=0.0010, 0.0012, and 0.0014 and split train / test to obtain the highest accuracy. Schemes 1, 2, and 3 have an increased accuracy of 87.78%, 89.58%, and 91.67%. The average accuracy in schemes 1, 2, and 3 also increased by 76.79%, 77.49%, and 82.82%. The highest accuracy is found in schematic 3 with the train data: 80:20 percent test data is 91.67% with parameter PC =10 and 20, hidden neuron=4 and μ value=0.001. It was concluded that the diagnosis of VT by the PCA and Levenberg Marquardt Backpropagation methods was sufficient for a medical examination. Because the PCA method is good enough for reducing and detecting QRS waves and can classify the data correctly by using backpropagation modification of Levenberg Marquardt algorithm. In this research, there are many shortcomings, therefore there needs to be an improvement for the next research using other methods in terms of Preprocessing, feature extraction and classification, such as DWT by hybrid between GA and ANN or using RR Interval and K-neighrest Neighbor etc.

REFERENCES

- [1] Abbas Rahat, Aziz Wajid, Arif Muhammad, "Prediction Of Ventricular Tachyarhhythmia In Electrocardiogram Signals Using Neural Network And Modified Nearest Neighbour Method", Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS).
- [2] F.N. Mohamad, M.S.A. Megat Ali, A. H. Jahidin, M.F. Saaid, and M.Z.H. Noor, "Principal Component Analysis and Arrhythmia Recognition Using Elman Neural Network", IEEE Control and System Graduate Research Colloquium, Aug 2013.
- [3] D. Ge, L. Sun, J. Zhou, and Y.Shao, "Discrimination of myocardial infarction stages by subjective feature extraction", Comput. Meth. Prog. Blomed., vol. 95, no.3, pp.270-279, Sep 2009.
- [4] https://www.physionet.org/cgi-bin/atm/ATM
- [5] Adiwijaya, Deteksi Kanker Berdasarkan Klasifikasi Microarray Data, Media Inform. Budidarma, vol. 2, no. 4, pp. 181–186, 2018.
- [6] Adiwijaya, U. N. Wisesty, E. Lisnawati, A. Aditsania, D. S. Kusumo. Dimensionality Reduction using Principal Component Analysis for Cancer Detection based on Microarray Data Classification, Journal of Computer Science 14(11), 2018.
- [7] Destiani, D.K. and Utama, D.Q. Study of Wavelet and Line Search Techniques on Modified Backpropagation Polak-Ribiere Algorithm for Heart Failure Detection. In 2018 6th International Conference on Information and Communication Technology (ICoICT) (pp. 319-323), IEEE, May 2018.
- [8] I. Jekova, "Shock advisory tool: Detection of life-threatening cardiac arrhythmias and shock success prediction by means of a common parameter set," Biomed. Signal Process. Control, vol. 2, pp.25–33, Jan 2007.
- [9] Ashley EA, Niebauer J, "Caridology Explained," London: Remedica., 2004.
- [10] Achmad Rizal, "Perbandingan Skema Dekomposisi Paket Wavelet untuk Pengenalan Sinyal EKG," JNTETI, Vol.4, No.2. Mei 2005.
- [11] Adiwijaya, Wisesty, U.N. and Nhita, F. Studi of line search techniques on the modified backpropagation for forecasting of weather data in Indonesia. Far East Journal of Mathematical Sciences, 86(2), p.139, 2014.
- [12] Resiandi, Kartika, and Dody Qori Utama. "Detection of Atrial Fibrillation Disease Based on Electrocardiogram Signal Classification Using RR Interval and K-Nearest Neighbor." 2018 6th International Conference on Information and Communication Technology (ICoICT). IEEE, 2018.
- [13] Manuel B., Tricahyono D., "Classifying Electronic Word of Mouth and Competitive Position in Online Game Industry", Journal of Data Science and Its Applications (JDSA), 1(1), pp.20-27. 2018