ICIC Express Letters
Volume **14**, Number **12**, December **2020**

ICIC International ©2020 ISSN 1881-803X
pp. **1147–1154**

THE CLASSIFICATION OF ARRHYTHMIA USING THE METHOD OF EXTREME LEARNING MACHINE

Okky Putra Barus and Nico Surantha

Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
JL. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
okky.barus@binus.ac.id; nsurantha@binus.edu

Received March 2020; accepted June 2020

Abstract. *Arrhythmia is a condition in which the heart's electrical impulses have an irregular frequency. Many studies have been done to find an arrhythmia classification algorithm which has satisfactory learning speed and accuracy. However, many of those result in unstable performances in addition to slow learning speed. This research uses the MIT-BIH Arrhythmia Database from the PhysioNet Resource which will be tested through three stages: preprocessing performed by resampling and scaling, feature extraction using Short Time Fourier Transform (STFT), and classification using the method of Extreme Learning Machine (ELM) which has a high learning speed and accuracy. This research is limited to five classes of arrythmia: Paced Beat (PB), Right Bundle Branch Block (RBBB), Congestive Heart Failure (CHF), Atrial Fibrillation (AF) and the normal heart condition Normal Sinus Rhythm (NSR). The parameters of accuracy, sensitivity, and computation time are tested out of 125 test data. The result shows that arrhythmia classification using ELM has 100% accuracy, 100% sensitivity, and average computation time of 0.107 seconds. This result is also compared with another method in supervised learning used in classifications called Support Vector Machine (SVM) to prove that ELM is the suggested approach.*

Keywords: Arrhythmia classification, Extreme Learning Machine (ELM), Short Time Fourier Transform (STFT)

1. **Introduction.** The heart is one of the vital organs which functions to pump blood to the entire human body. According to the World Health Organization, heart diseases are the number one cause of death in the period of 2000-2015 [1]. Data shows that in 2015, approximately 8,756,000 people died from heart diseases. This number increased from the number of deaths caused by the heart in 2010 which was less than six million people during that time.

Heart diseases can increase the risk of high blood pressure and high cholesterol in which both can cause a heart attack. When the coronary arteries shrink, arrhythmia occurs and results in permanent damage of the heart muscles. If left untreated, it can cause stroke. Symptoms of arrhythmia include fatigue and chest pain. Arrhythmia can generally be diagnosed using the Electrocardiogram (ECG).

Many researches in the biomedical field related to the identification of heart disorders have previously been conducted. Khandait et al. identified arrhythmia through the wavelet transform method which has an accuracy of 98.17% [2]. Research on the identification of heart disease using the fuzzy cluster method was also done by Joshy and Pamela [3]. Rama et al. identified heart disorders using the method of signal processing which was successful in describing the heart condition and was able to recognize the identified heart disorder [4]. The development of techniques in classifying medical signals creates

DOI: 10.24507/icicel.14.12.1147

opportunities to develop many algorithms related to that matter, such as using the Lyapunov transformation method, hidden Markov model, fuzzy adaptive resonance theory, and filter analysis.

Classification of ECG signals in previous researches uses the Principal Component Analysis (PCA) method in the feature extraction process and Artificial Neural Network (ANN) back propagation as the classifier. However, the accuracy of these researches reaches a mere 71% at most. This concludes that PCA is less effective to be used as a feature extraction of signals especially those with small matrices, has an insufficiency of training data during the ANN learning process to recognize the pattern variations of heart signals, and it requires a learning speed (learning process time) up to 8 hours for 1,500,000 epoch.

In order to achieve a better accuracy level and to reduce the learning speed, this research implements the ANN method of ELM which is the ANN feed-forward with one hidden layer or more commonly known as Single hidden Layer Feed-forward Neural network (SLFN) [5]. This method also has a better accuracy and learning speed if compared to conventional methods such as exponential smoothing and moving average [6,7]. The goal is to implement ELM in classifying arrhythmia and to prove that ELM is the suggested approach.

This paper is organized as follows: several literature studies pertaining to this research are explained in Section 2, the methodology including all phases and steps taken using the MATLAB tool to build a diagnosis model is described in Section 3, test results including comparison between ELM and SVM are presented in Section 4, and the conclusion follows in Section 5.

2. Literature Study.

2.1. Arrhythmia.

The irregular heart rhythm of this type of heart disease can mean the heart beats too fast, too slow, or simply abnormal. There are numerous types of arrhythmia such as tachycardia, atrial flutter, supraventricular tachycardia, ventricular tachycardia, ventricular fibrillation, bradycardia, and atrial fibrillation. This research is limited to the following five classes.

a) Paced Beat (PB). Ventricular muscle which controls conduction of the heartbeat and generates slower heartbeats, ranges between 30 to 50 bpm (beats per minute).
b) Right Bundle Branch Block (RBBB). Delayed activation of the right chamber, which causes the right chamber to contract later than the left chamber.
c) Congestive Heart Failure (CHF). The function of the heart to circulate oxygen to the entire body is not able to provide the body's oxygen needs adequately.
d) Atrial Fibrillation (AF). Electrical changes are irregular and very fast in the atrium.
e) Normal Sinus Rhythm (NSR). Rhythm occurs continuously and periodically, heartbeat ranges 60 to 100 bpm.

2.2. Electrocardiogram (ECG).

This medical tool is used together with the electrocardiograph which are notes and graphs obtained from the human body through electrodes and captures potential emission of the heart's bioelectrical wave. Figure 1 shows an example of a normal ECG schematic.

The ECG wave is represented by the letters P, Q, R, S, and T. QRS complex wave is the wave group as a result of the depolarization of the right and left ventricles. Morphology, amplitude, and the complex duration of QRS are used to diagnose arrhythmia.

2.3. Short Time Fourier Transform (STFT).

An enhanced mathematical methodology is derived from Discrete Fourier Transform (DFT) to explore the instantaneous frequency as well as the instantaneous amplitude of localized waves with time-varying characteristics [8]. On this algorithm, signals are sampled in a specific time range. The

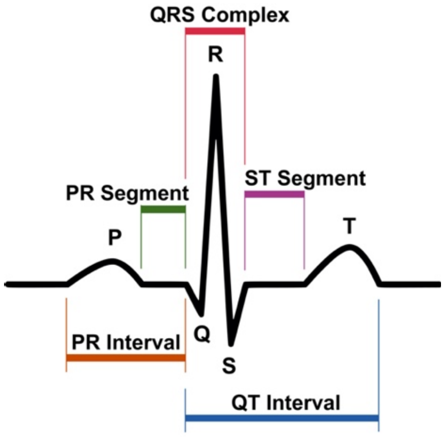

FIGURE 1. Normal ECG schematic

signals are rendered in the frequency domain. Subsequently, the signals are resampled for t seconds and the position of the signal will be known in the frequency and time domains.

The equation below where $x(\tau)$ is Fourier transformation inverse, h is a linear system, τ is impulse at time t, and $d\tau$ are superimposed (added up) for each input time τ

$$STFT(t, f) = \int x(\tau)h * (\tau - t)e^{-2j\pi\tau}d\tau \tag{1}$$

maps out the signal in two dimensions that is the frequency and time domains as seen in Figure 2.

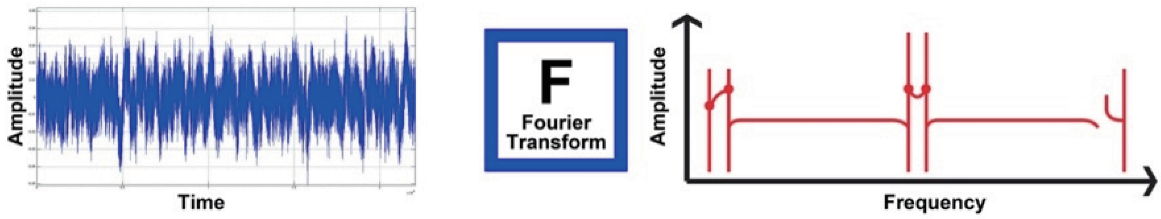

FIGURE 2. STFT diagram block

2.4. **Artificial Neural Network (ANN).** This system has been widely implemented for pattern classification and regression problems due to its ability to obtain the function of non-linear prediction models which describes the relationship between dependent and independent variables by using the provided input samples [9]. Despite having numerous advantages such as a better approach ability and a simple network structure, it has an inaccurate learning rate as well as inaccurate selection of total hidden neurons.

2.5. **Extreme Learning Machine (ELM).** This method is a feed-forward ANN which consists of a single hidden layer known as Single hidden Layer Feed-forward Neural Network (SLFN). It has an astounding and substantial characteristic but is different from the learning algorithm that is based on the popular gradient for feed-forward neural network. The characteristics are as follows [10-12].

a) The learning phase can be completed in mere seconds for numerous applications.
b) The generalization performance is far better than learning based on gradient, for example, in most cases of backpropagation.

ELM has the tendency to accomplish simple solutions without being affected by trivial matters. Its learning algorithm is far simpler compared with the learning algorithm of a feed-forward neural network in most cases. However, differing from the learning algorithm based on gradient which only works for differentiated activation function, the ELM algorithm can be implemented on SLFN with the purpose of training SLFN with ample amount of activation functions and not having differentiation.

2.6. **Support Vector Machine (SVM).** This method was first introduced in 1992 and has a better performance in application fields such as bioinformatics, handwriting recognition, disease classification diagnosis, as well as text classifications [13]. SVM is a classification and regression learning method with an orientation towards an effortless computation and currently has achieved great success as a result of being the state-of-the-art classifier [14]. If used effectively, this method has a better performance and optimizes the best prediction parameters [15]. SVM is essentially a binary classifier [16]. However, a one-against-all multiclass SVM is used in this research due to various classes of arrhythmias.

3. **Methodology.** This research analyzes the accuracy of the algorithm to generate a prediction and is expected to generate a model accuracy on arrhythmia diagnosis in which the accuracy measurement will use the confusion matrix.

3.1. **Preprocessing.** The data taken from the MIT-BIH Database is in the amount of 125 training data and 125 test data with 25 data for each class of arrhythmia. During this stage, the number of samples on the data is uniformized to 300 samples. Subsequently, a scaling process is done to position the signal in order to have the same amplitude range of the values $(+1)$ and (-1) without altering the form of the input signal. The result is a matrix in the dimensions of 1×300.

3.2. **Feature extraction.** The STFT algorithm used here is in the format S = spectrogram (x, window, n-overlap, nfft), where S is complex matrix with nfft line and k column, and x is length Nx complex signal. Each S column contains a calculation of the short-term which contains the time-localized frequency of x. Time increases along S column and frequency increases along the line. Based on testing, the ideal performance tuning STFT parameters occurs on window 20, n-overlap 19, and nfft 599. Thus, the value of k is:

$$k = \operatorname{fix}\left(Nx - \frac{\textit{n-overlap}}{\textit{window} - \textit{n-overlap}}\right) = \frac{300 - 19}{20 - 19} = 281 \tag{2}$$

Consequently, a matrix as a result of the STFT transformation is obtained in 281×300 dimensions. The following process is to determine the maximum and minimum values of each column which results in a 1×600 matrix with a maximum value of 300 on the left and minimum of 300 on the right.

$$\{\ \max_1 \quad \max_2 \quad \cdots \quad \max_{300} \quad \min_1 \quad \min_2 \quad \cdots \quad \min_{300}\ \} \tag{3}$$

As this research uses 125 data, the resulting matrix is 125×600 as seen below.

$$\begin{pmatrix} \max_{1-1} & \max_{2-1} & \cdots & \max_{300-1} & \min_{1-1} & \min_{2-1} & \cdots & \min_{300-1} \\ \max_{1-2} & \max_{2-2} & \cdots & \max_{300-2} & \min_{1-2} & \min_{2-2} & \cdots & \min_{300-2} \\ \max_{1-125} & \max_{2-125} & \cdots & \max_{300-125} & \min_{1-125} & \min_{2-125} & \cdots & \min_{300-125} \end{pmatrix} \tag{4}$$

The maximum and minimum values on this 600-columned matrix will be inputted into the ELM ANN.

3.3. **Classification.** At this final stage, the feature extraction results are identified using the ELM algorithm as the classifier. The algorithm is designed to identify the features of the heart's data signal which have been previously stored in the database by performing two phases.

3.3.1. *Learning phase.* This phase is performed repeatedly until optimum values of the 125 test data are obtained. It is expected that the value of the total accuracy achieves 100%. The following are the 3 steps done at this phase, if given the training data of activation function g and total hidden nodes L.

a) Randomly determine the weight vector input w_i and influence factor of the hidden node ith B_i, $i = 1, \ldots, L$. For example, if $L = 800$, $x_j = 600$ and 125 training data, then it results in matrix

$$w_i = [B_i \times x_i][x_i \times x_j] \tag{5}$$

$$w_i = [800 \times 125][125 \times 600] = [800 \times 600] \tag{6}$$

b) Calculate the output matrix on hidden layer H. Continuing from the above, then matrix

$$H = \left[w_i \times x_j'\right] + b_i = [800 \times 600][600 \times 125] + b_i = [800 \times 125] \tag{7}$$

b_i result is adjusted.

c) Calculate the output weight

$$\beta = H * T = [800 \times 125] * [125 \times 5] = [800 \times 5] \tag{8}$$

H is Moore-Pemrose generalized inverse on the output matrix of the H hidden layer. This ELM learning process is eventually done to search for the output value of β, thus the fast learning time duration. After the process is completed, the system will store three values: b_i hidden node bias, w_i input value, and β output value. These three values will be used at the testing phase.

The parameters for ELM learning are as follows.

- Sigmoid activation function which is the function value between 0 and 1. The equation is

$$f(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

- The total number of hidden node L in which the default is 20. The value of L will be changed six times (20, 100, 200, 800, 1000, and 1600) after obtaining the best accuracy at the feature extraction results with certain n-overlap.

3.3.2. *Testing phase.* The ELM input network derives from 600 STFT feature extraction results – 300 maximum values and 300 minimum values. The three values which have been stored during the learning phase are used at this stage and it is expected that the total accuracy value of the test data achieves the highest possibility. The test system is explained further in the following section.

4. **Test System and Analysis Simulation Results.**

4.1. **Test accuracy.** The following is the results of characteristic recognition after the feature extraction process using STFT with the aforementioned ideal performance tuning parameters and ELM (with total $L = 800$ and sigmoid activation function).

As seen in Table 1, the system did not find any errors and all 125 data were correct. Thus, the accuracy of the system for the above STFT and ELM parameters reaches 100%.

4.2. **System sensitivity.** The accuracy results of the characteristic recognition on the above test system show that one of the system performances was successful in recognizing data. In the medical field, another test parameter is often used to show system performance which is known as sensitivity. The following is the results of characteristic recognition after the feature extraction process using STFT with the aforementioned ideal performance tuning parameters and ELM (with total $L = 800$ and sigmoid activation function).

TABLE 1. Accuracy results of characteristic recognition using ELM

Data	**Known as**				
	AF	**CHF**	**NSR**	**RBBB**	**PB**
AF	25	0	0	0	0
CHF	0	25	0	0	0
NSR	0	0	25	0	0
RBBB	0	0	0	25	0
PB	0	0	0	0	25
Accuracy	$(125/125) \times 100\% = 100\%$				

Based on Table 2, the test data has a larger total of TP (True Positive) than FN (False Negative). The system simulation has a sensitivity level of 100%.

TABLE 2. Sensitivity results of characteristic recognition using ELM

Data	**TP**	**TN**	**FP**	**FN**
AF	25	0	0	0
CHF	25	0	0	0
NSR	0	25	0	0
RBBB	25	0	0	0
PB	25	0	0	0
Total	100	25	0	0
Sensitivity = TP/(TP + FN) × 100%	$100/(100 + 0) \times 100\% = 100\%$			

4.3. **Computation time.** This test system shows the average computation time test results at the time of the feature extraction and classification (through 125 times of tests with the ideal performance tuning STFT parameters, $L = 800$ and sigmoid activation function).

TABLE 3. Average time of the computation system during test

Parameter	**Average time (seconds)**
STFT feature extraction	0.037
ELM classification	0.070
Total time average	0.107

4.4. **ELM and SVM prediction comparison.** Table 4 compares the rate of test processes using ELM and SVM.

TABLE 4. Comparison of test processes using ELM and SVM

Method	**Accuracy**	**Sensitivity**
ELM	**100%**	**100%**
SVM	97.59%	97.56%

Because the accuracy of this method is seen from the test process results, the comparison of ELM and SVM can only be seen from test process results only. Average accuracy generated by ELM is 100%, whereas average accuracy generated by SVM is 97.59%. Based on those numbers, it can be concluded that the average accuracy of the ELM performance is better than the SVM performance.

SVM is intended to achieve structure minimalization by using margin concepts, unlike criteria functions from other pattern recognition algorithms which aim at minimalizing empirical errors. In contrast, ELM performs a learning algorithm which analytically calculates the optimum value between the output layer and hidden layer using the linear matrix equation. Thus, ELM is the preferable method.

5. **Conclusions.** Classification of arrhythmia using ELM with a feature extraction process using STFT to precede it is the suggested approach. This research proves that the accuracy rate of the test system is 100% as it shows that the system recognizes all 125 test data correctly and the system sensitivity is also 100% for the corresponding arrhythmia classes. It also shows that the total time average of the overall process is 0.107 seconds, with average time of the feature extraction process using STFT for all test data being 0.037 and average time of the classification process using ELM being 0.070. During the learning process, it is found that the results presented above are obtained if the performance tuning is set to window 20, n-overlap 19, nfft 599, total hidden nodes L 800 (9th test) and 1600 (10th test), and sigmoid activation function which are both 100% respectively. The more hidden nodes there are, the better are the training and test accuracy results but the slower the learning speed is.

For future researches, comparisons with other methods at the preprocessing stage can be done with morphology filtering, Finite Impulse Response (FIR) filter, or polynomial curve algorithm to correct bad signals. Additionally, comparisons with other methods at the feature extraction stage can be done as well with Durbin or Prony to extract more specific features.

Acknowledgment. This work is supported by the Directorate General of Strengthening for Research and Development, Ministry of Research and Technology, Republic of Indonesia as a part of Penelitian Dasar Unggulan Perguruan Tinggi Research Grant to Binus University entitled "Sistem dan Aplikasi Portabel Pendeteksi Gangguan Irama Jantung Menggunakan Sinyal ECG Berbasis Machine Learning dan Cloud Computing" or "Portable ECG-Based Heart Arrhythmia Prediction System and Application Using Machine Learning and Cloud Computing".

REFERENCES

[1] World Health Organization, *The Top 10 Causes of Death*, https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, 2018.

[2] P. D. Khandait, N. G. Bawane and S. S. Limaye, Features extraction of ECG signal for detection of cardiac arrhythmias, *International Journal of Computer Applications Proceedings on National Conference on Innovative Paradigms in Engineering and Technology*, no.8, 2012.

[3] N. Joshy and D. Pamela, Heart diseases identification system using fuzzy cluster algorithm, *International Journal of Engineering Research & Technology*, vol.3, no.4, pp.1217-1220, 2014.

[4] V. Rama, C. B. Rama Rao and N. Duggirala, Analysis of signal processing techniques to identify cardiac disorders, *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Controlengineering*, vol.3, no.6, pp.159-164, 2015.

[5] Z. L. Sun, T. M. Choi, K. F. Au and Y. Yu, Sales forecasting using extreme learning machine with applications in fashion retailing, *Decision Support System*, vol.46, no.1, pp.411-419, 2008.

[6] V. Cerqueira, L. Torgo and C. Soares, Machine learning vs statistical methods for time series forecasting: Size matters, *arXiv abs/1909.13316*, 2019.

[7] F. Pan, H. Zhang and M. Xia, A hybrid time-series forecasting model using extreme learning machines, *The 2nd International Conference on Intelligent Computation Technology and Automation*, vol.1, pp.933-936, 2009.

[8] S. Haykin and B. V. Veen, *Signals and Systems*, Wiley, Hoboken, NJ, USA, 1999.

[9] R. Singh and S. Balasundaram, Application of extreme learning machine method for time series analysis, *International Journal of Computer, Electrical, Automation, Control and Information Engineering and Proc. of World Academy of Science, Engineering and Technology*, vol.11, no.11, pp.3394-3400, 2007.

[10] G. B. Huang, Q. Y. Zhu and C. K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing*, vol.70, nos.1-3, pp.489-501, 2006.

[11] J. J. Pangaribuan, Mendiagnosis penyakit diabetes melitus dengan menggunakan metode extreme learning machine, *Information System Development*, vol.2, no.2, pp.32-40, 2016.
[12] O. P. Barus, Perbandingan metode extreme learning machine dan backpropagation untuk mengklasifikasi phishing website, *Informatics Engineering Research and Technology*, vol.1, no.1, pp.1-8, 2019.
[13] F. Li, Comparison of the primitive classifiers without features selection in credit scoring, *International Conference on Management and Service Science*, pp.1-5, 2009.
[14] K. Z. Huang, H. Yang, I. King and M. R. Lyu, *Machine Learning – Modeling Data Locally and Globally*, Berlin Heidelberg, Zhejiang University Press, Hangzhou and Springer-Verlag Gmbh, 2008.
[15] C. Dewi and R.-C. Chen, Random forest and support vector machine on features selection for regression analysis, *International Journal of Innovative Computing, Information and Control*, vol.15, no.6, pp.2027-2037, 2019.
[16] G. Xie, S. Wang and K. K. Lai, Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches, *Journal of Air Transport Management*, vol.37, pp.20-26, 2014.