Classification of Alcoholic EEG Using Wavelet Packet Decomposition, Principal Component Analysis, and Combination of Genetic Algorithm and Neural Network

Muhammad Saddam, Handayani Tjandrasa, Dini Adni Navastara
Department of Informatics
Faculty of Information Technology, ITS
Surabaya, Indonesia
muhsad.dahlan@gmail.com, handatj@its.ac.id, dini_navastara@if.its.ac.id

***Abstract*— Alcoholism is a disorder characterized by excessive consumption and dependence on alcohol. There are various ways to detect whether a patient is addicted to alcohol, one of them by brain detection using electroencephalograph (EEG). However, the signals generated by the EEG recorder should be prepared to do further processing to detect brain abnormalities automatically. Therefore, this research implements Wavelet Packet Decomposition (WPD) method for feature extraction, Principal Component Analysis (PCA) for dimension reduction, and Back Propagation Neural Network optimized with Genetic Algorithm for alcohol addiction classification. Based on the experiment results, the best performance was 94.00% accuracy with decomposition of 3 levels, taking 30% of the features, and classification using Neural Network and Genetic Algorithm with learning rate of 0.1.**

Keywords—Alcoholism; EEG; Wavelet Packet Decomposition; Principal Component Analysis; Neural Network; Genetic Algorithm

I. INTRODUCTION

Alcoholism is a disorder characterized by excessive consumption and dependence on alcohol. Alcohol itself has become a common health problem and affects 10% of the world's population [1]. Alcohol addiction can cause several negative factors and risks that include psychological disorders, such as stress, depression, anxiety disorder, or schizophrenia. There are various ways to detect whether a patient has been addicted to alcohol, one of them by brain detection using Electroencephalograph (EEG) [2].

EEG signals are widely and clinically used for detection of brain disorders in the world of health. Some studies stated that irregular and complex EEG signals can provide information about the fundamental neural activity in the brain. However, the signals produced by the EEG recorder are considered not ready for further processing in order to detect brain abnormalities automatically. Therefore, there is a need for preprocessing methods for proper feature extraction in order to obtain implicitly stored characteristics of the EEG signal. There are a variety of EEG data preprocessing methods, one of them is Wavelet Transform which decomposes signals into various frequency bands to produce new features, and Principal Component Analysis (PCA) to reduce the dimensions of the feature vector obtained.

There have been many previously formulated classification methods to be able to detect disease based on EEG signals [3,4]. One such classification method is Neural Network. Neural Network alone has the disadvantage that the dominance of the weights and biases parameter that cause a major influence on the performance of Neural Network. Therefore, optimization on the initial weight and bias are needed to support the performance of Neural Network optimally, one of them by implementing Genetic Algorithm (GA) on Neural Network method.

In previous research, the alcoholic classification processing based on EEG data has been previously discussed [5]. However, in the research there was no weight and bias update on the classification of neural network so that the dominance of the weight and bias of this initial set affect the results of classification. Therefore, this research implements Wavelet Packet Decomposition (WPD), Principal Component Analysis (PCA), and Neural Network optimized with Genetic Algorithm (GA) for alcohol addiction classification. Previously, the EEG data will pass through three main stages: preprocessed using Wavelet Packet Decomposition (WPD) method for signal decomposition and feature extraction, Principal Component Analysis (PCA) for feature dimension reduction, and then classification using a combination of Genetic Algorithm and Neural Network.

II. MATERIAL AND METHODS

As described in the introduction, the main objective of this study was about alcoholic classification based on EEG data using Wavelet Packet Decomposition (WPD) for signal decomposition and feature extraction, Principal Component Analysis (PCA) for feature dimensional reduction, and classification using Neural Network optimized by Genetic Algorithm in the livelihood of initial weight and bias. Overall, the main flowchart of this study can be seen in Fig. 1.

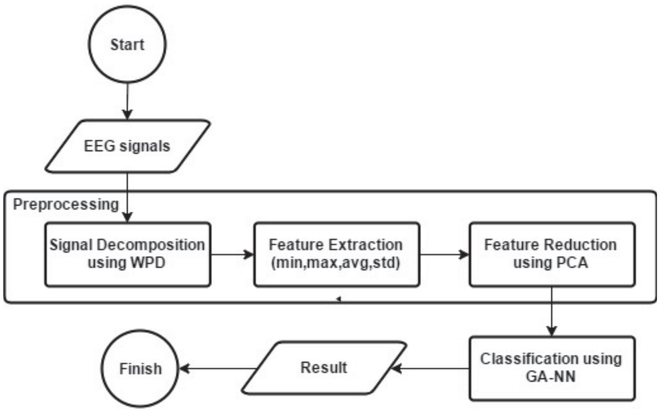

Fig. 1. Main flowchart of the classification system

A. *Electroencephalograph of Alcoholism*

Alcoholism is one of the chronic diseases that include problems including controlling patients for drinking, becoming preoccupied with alcohol, and even continuing to use alcohol as a solution when problems occur. Alcoholic patients usually have to drink more alcohol to get the same effect (physical dependence). This is due to several factors that lie behind a person to consume alcohol until becoming an alcoholic, such as biological factors of psychological, social, and cultural environment. As described in the introduction chapter, one way to detect alcoholism is with EEG recording.

EEG is a procedure of recording the electrical activity of the brain with a sensitive recording device. The EEG will then measure the voltage fluctuations produced by ion currents within the brain's neurons. In a clinical context, EEG refers to the recording of spontaneous electrical activity of the brain over a period of time recorded from many electrodes mounted on the scalp. Neurons in the cerebral cortex emit electrical waves with very small voltages (mV), then flowed to an EEG machine for amplification so that the signal is recorded enough to be captured by EEG readers as Delta, Theta, Alpha, and Beta waves. Fig 2. shows the EEG signals for alcoholic subject and control subject.

B. *Wavelet Packet Decomposition (WPD)*

Wavelet Packet Decomposition (WPD) works by using a wavelet that scales and translates the original nature of the main wavelet called 'mother wavelet' to transform the input signal [4]. WPD will decompose multi-resolution signals divided into low-frequency signals or approximations and high-frequency or detail signals using highpass filters and lowpass filters. Unlike the DWT that only fixed decomposition coefficient approximation, WPD will still decompose at each level of detail coefficient and its approximation to form a full binary tree. In this study, WPD implementation is limited to 3 levels. The topology of 3-level WPD decomposition method can be seen in Fig. 3.

Decomposition results at each derived level, resulting signal length will be reduced by half number of signals at the previous level. WPD signal decomposition result of 3 levels, shown in Fig. 4. After the decomposition of WPD, the next is the feature extraction process by taking the mean, minimum, maximum, and standard deviations of each derived coefficient so there are 56 new features for each training and testing data.

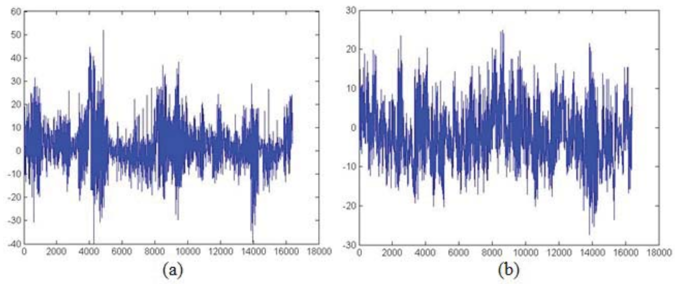

Fig. 2. EEG signals of (a) alcoholic subject and (b) control subject

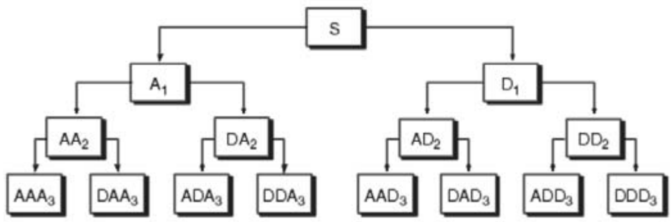

Fig. 3. Topology of WPD

C. *Principal Component Analysis (PCA)*

The PCA is one of the dimensional reduction techniques that converts datasets with large dimensions (dimension - m) to low dimensional orthogonal feature spaces (dimensions - n, where m > n) but retains information from the large dimensions of the original dataset [6]. Essentially the goal of PCA dimensional reduction is intended to facilitate engine classification performance to be more effective, to make complexity lower, and to shorten computing time [1]. The first stage done in PCA is to make a variance-covariance matrix on a diagonal containing the value of variance and the other side containing the value of covariance. The variance-covariance matrix has *n* x *n* dimension where *n* is the number of attributes / dimensions of data, X_i and Y_i are data, and $\bar{X}$ dan $\bar{Y}$ are the mean of each data attribute as can be seen in (1) and (2).

$$\mathrm{var}(x) = \frac{\sum_{n=1}^{n} (Xi - \bar{X})}{(n-1)} \tag{1}$$

$$\mathrm{covar}(x, y) = \frac{\sum_{n=1}^{n} (Xi - \bar{X})(Yi - \bar{Y})}{(n-1)} \tag{2}$$

With the aim of obtaining unrelated and significant components, eigenvectors and eigenvalue within PCA play an important role. The eigenvector and eigenvalue can be seen in (3).

$$\sum X = \lambda X \tag{3}$$

where Σ is the covariance of the matrix, X is a non-zero vector, λ is eigenvalue*,* dan *v* is eigenvector [7]. The next stage of the PC will be sorted according to the eigenvalue compatibility level so that the very first PC presents the most significant variant of the dataset. Dimensional reduction process is done by transforming the initial data to a new set of variables (PC) that are not correlated with each other. In other words, the number of PCs will always be smaller or equal to the number of initial variables [8].

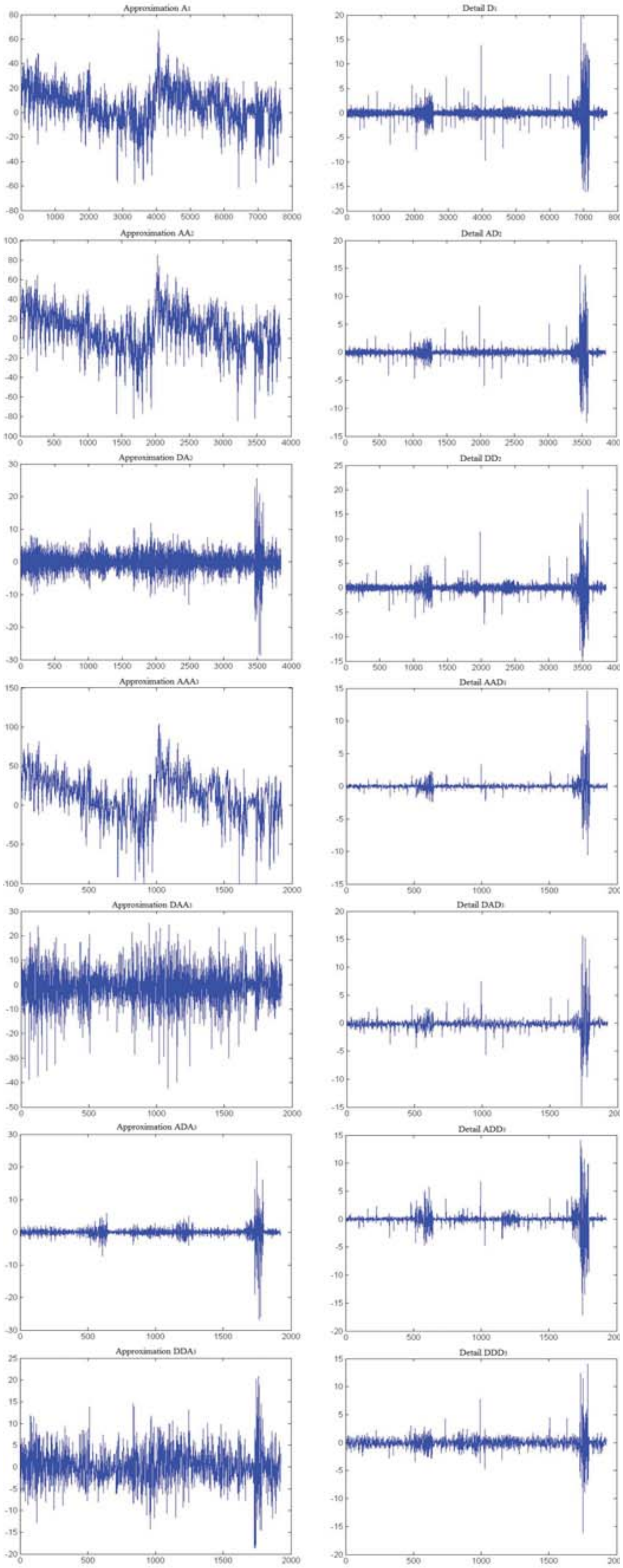

Fig. 4. Three Level of WPD decomposition of alcoholic subject

D. Combination of Genetic Algorithm and Neural Network

Neural Network is a guided training algorithm that has some drawbacks, one of which is having a considerable number of parameters and must be insert manually. The dominance of the number of parameters on the Neural Network lies in the value of weight and bias. The initial weight and initial bias values are randomly assigned which have a major influence on the performance of the Neural Network. So an optimization on the initial weight and bias are needed to support the performance of Neural Network optimally. One method that can optimize the performance of Neural Network is to implement the evolution algorithm, Genetic Algorithm (GA) on the Neural Network method.

The use of GA in addition to finding the best initial weight and bias of a Neural Network is used to speed up convergence. This is used because initialized weights and biases have a significant effect on Neural Network performance and Back Propagation algorithm makes the convergence process slower, requiring suitable optimization algorithms [9]. Network development in this study applies 1 hidden layer consisting of 15 nodes. Using the sigmoid activation function shown in (4).

$$sigmoid(Xi) = \frac{1}{1+\exp(-Xi)} \tag{4}$$

GA will look for the best weight and bias chromosomes generated from GA's important processes and use the error value of Neural Network to measure the performance of the resulting Network model. Network with the smallest error value is the best Network model that will be used for testing process. The first thing to do with the GA-NN combination is to find the initial population using the Neural Network and store the weight and bias values and fitness functions shown in (5).

$$f = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}(t-\bar{a})^2} = \frac{1}{MSE} \tag{5}$$

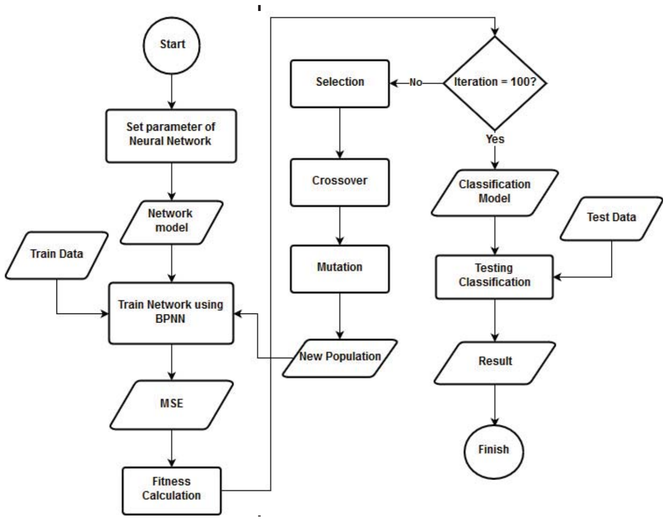

Fig. 5. Flowchart of GA-NN

where N is lot of training data, t is the target output, $\bar{a}$ is the output results, and MSE is Mean Square Error generated by Neural Network [9]. This weight and bias will be used as the initial population to store the value of fitness as a respresentation of each individual. In this study the initial population number = 15 by considering the complexity and computation time. Furthermore, the system will pass the essential processes of GA in the form of selection, crossover, mutation, and evaluation until it reaches the termination criteria = 100 times iteration. Evaluation process in GA by applying weight and bias after GA and re-classification with Neural Network to get fitness value from result of GA process. Selection process from GA using roulette wheel method, crossover process using arithmatic crossover method (6).

$$\begin{cases} X_A^{i+1} = \alpha X_B^{'} + (1-\alpha) X_A^{i} \\ X_B^{i+1} = \alpha X_A^{'} + (1-\alpha) X_B^{i} \end{cases} \quad (6)$$

where α is a parameter that can be chosen constantly $\alpha = 0.9$, i = number of generation produced [9]. Furthermore, the mutation process should be applied with very little probability. This is because if the mutation process is done too often, it will produce a weak individual because the gene configuration in the superior individual will be damaged. This research takes the probability of mutation = 0.095. In this research, the number of chromosomes optimized by GA in each trial will be different depending the number of used features. As described before, this research only use one hidden layer containing 15 nodes and 2 nodes in output layer as it only has 2 classes. Overall, the flowchart of the combination of Genetic Algorithm and Neural Network can be seen in Fig. 5.

III. RESULT

A. EEG Dataset

The EEG dataset used is data from the State University of New York Health Center downloaded at UCI Machine Learning Repository. In each data there are 64 channels with a frequency of 256 Hz per second. In this study, 260 random data were collected, 130 alcoholics and 130 controls, which will be divided into 2 datasets. The first dataset consists of 60 training data and 40 testing data, and the second dataset consists of 120 training data and 40 testing data.

B. PCA Features Selection

The first tests were performed by comparing performance with and without PCA and calculating the performance based on the number of features produced by PCA. In the performance result, it concluded that PCA gives higher accuracy with less running time. It is because the data's dimension of features after PCA has less number of features than without PCA method applied, so the complexity of systems will also be less. The result of PCA performance can be seen in Table I.

Basically, the result of PCA itself is a new feature with smaller dimensions. So, the feature taken after the PCA dimension reduction itself by prioritizing features with the highest eigenvalues. The number of features taken after the PCA method can be seen in Table II. Performance test from PCA itself is by calculating the accuracy generated from each dataset 1 and dataset 2. In this scenario, the classification using combination of Genetic Algorithm and Neural Network with the value of Learning Rate 0.1. The conclusion of the test results shown in Table III is the best accuracy generated with 30% feature retrieval.

It is also concluded that the more features used, the less accuracy will be produced by classification. As described in materials and method, PCA sorted according to the eigenvalue compatibility level so that the very first PC presents the most significant variant of the dataset. It means that using more features obtained from PCA with small value of eigenvalues, will caused less accuracy. The calculation result of the average accuracy is obtained from 5 experiments.

C. GA Performance on Neural Network

In the second test scenario, the performance of Genetic Algorithm (GA) method will be tested on Back Propagation Neural Network method in optimizing the early weight and bias of the classification system. The calculation of the performance itself by counting accuracy and running-time of Neural Network without GA and GA-NN. In this trial scenario, the data used is the dataset after the PCA dimension reduction with 30% features. Based on Table IV, the GA-NN combination method can increase the accuracy rate by 9.00% for dataset 1 and 6.50% for dataset 2 compared to Neural Network method without GA. However, the opposite occurs in the results of the running-time produced indicating the GA-NN combination of GA-NN method has longer running-time

TABLE I. PCA PERFORMANCE

Method	Accuracy (%)		Running Time (s)	
	Dataset 1	*Dataset 2*	*Dataset 1*	*Dataset 2*
PCA	89.00	81.50	44.000	44.451
No PCA	82.50	80.00	120.608	122.001
Differences	**6.50**	**4.00**	**76.608**	**77.550**

TABLE II. NUMBER OF FEATURE PRODUCED AFTER PCA

Quantity (%)	Number of Features
30	17
40	23
50	28
60	34

TABLE III. PERFORMANCE OF EACH FEATURES

Quantity (%)	Accuracy (%)	
	Dataset 1	*Dataset 2*
30	89.00	81.50
40	84.50	78.00
50	80.50	70.00
60	79.00	73.50

results compared to Neural Network without GA until it reaches 45 seconds. This is certainly because the processes of the GA itself has a high enough complexity that requires longer computational time. In addition to calculating accuracy, the BPNN and combination of GA-NN can be seen from the number of iterations to get convergence. Taking into account computation time, the number of iterations is determined at the first of 10,000 iterations for BPNN and 100 for GA-NN. From the experimental results, it can be concluded that GA-NN can reach convergency faster with lower MSE value compared to BPNN method even the number of iteration is less. Although, the running time produced by GA-NN is higher than BPNN because of the complexity of GA method itself. The convergence process of both methods can be seen in Fig. 6 and 7.

D. Learning Rate Parameter on Neural Network

In this scenario, performance calculations are performed on the Neural Network classification method by replacing the constant of the initialized learning rate parameter. The parameters to be used are 0.01, 0.05, 0.1 and 0.2. The performance calculation on this trial scenario by calculating the average accuracy resulting from 5 experiments based on the value of the learning rate specified. The dataset used is data after PCA with 30.00% feature. While the classification method uses the GA-NN combination method. Based on the experimental results shown in Table V, it can be concluded that the highest average accuracy produced based on the value of learning rate 0.1 on dataset 1 of 94.00% and the value of 0.05 in the 2nd dataset of 82.50%.

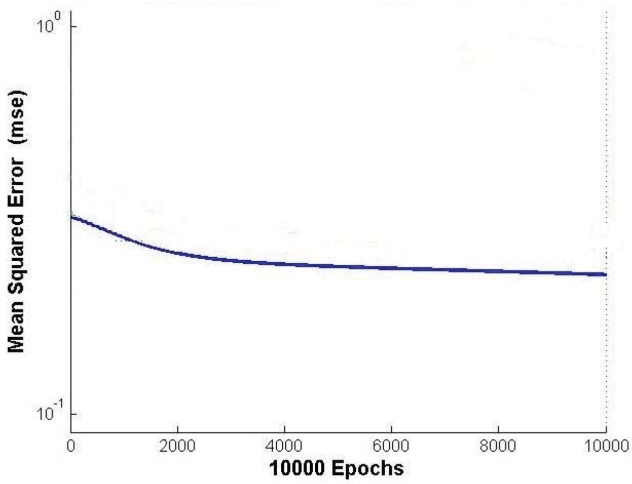

Fig. 6. Convergency of BPNN

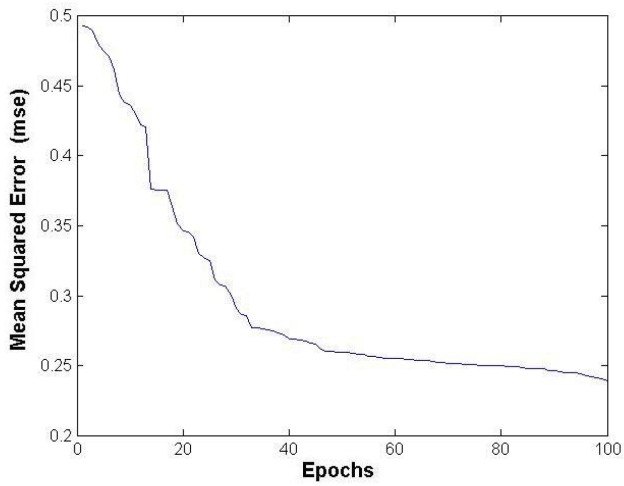

Fig. 7. Convergency of GA-NN

E. Method Comparison

In this scenario, there will be comparison between the method which proposed in this research and the proposed method from the previous research. The previous research was about alcoholism classification based on EEG data using the same dataset with the first dataset and using different methods: Independent Component Analysis (ICA), Wavelet de-noising, and Probabilistic Neural Network (PNN) [5]. Therefore, this scenario compare the highest accuracy using the first dataset only. Based on the experimental results shown in Table VI, our proposed method gives higher accuracy up to 9.00% compare to the previous method.

IV. CONCLUSIONS

The PCA method can be used to reduce the dimensions of EEG data by yielding the highest average accuracy of 89.00% for datasets 1 and 81.50% for dataset 2 with 30% feature capture. Furthermore, the implementation of Genetic Algorithm (GA) can be a suitable algorithm to optimize the initial weight and bias of Neural Network by increasing the accuracy of 9.00% for datasets 1 and 6.50% for dataset 2. However, GA has a high complexity that requires longer computation time up to 45 seconds. In other words, Wavelet Packet Decomposition (WPD), Principal Component Analysis (PCA), and classification using Genetic Algorithm and Neural Network Combinations can be used for pre-processing and

TABLE IV. PERFORMANCE OF GA AND NN

Method	Accuracy (%)		Running Time (s)	
	Dataset 1	*Dataset 2*	*Dataset 1*	*Dataset 2*
NN	80.00	75.00	44.000	45.000
GA - NN	89.00	81.50	45.008	45.451
Differences	9.00	6.50	1.008	0.451

TABLE V. PERFORMANCE FOR EACH LEARNING RATE

Learning Rate	Accuracy (%)	
	Dataset 1	*Dataset 2*
0.01	86.50	80.00
0.05	88.00	82.00
0.1	94.00	79.00
0.2	86.50	81.50

TABLE VI. COMPARISON METHOD

Method	Accuracy (%)
Previous Method [5]	85.00
Proposed Method	94.00
Differences	9.00

classification of alcoholic EEG data with the highest average accuracy results obtained from the trial scenario 94.00% for dataset 1 with learning rate parameter 0.1 and alpha parameter 0.9. Whereas for dataset 2 yields average of highest accuracy of 84.50% with learning rate parameter 0.05 and alpha parameter 0.9. And also, classification of alcoholic based on EEG data using Wavelet Packet Decomposition (WPD), Principal Component Analysis (PCA), and combination of Genetic Algorithm and Neural Network give a higher accuracy than Independent Component Analysis (ICA), Wavelet de-noising, and Probabilistic Neural Network (PNN) methods up to 9.00%. Lastly, this research gives the higher accuracy to detect alcoholic addiction automatically based on EEG data and also useful to reduce human error which can support the clinicians in analysing a lot of data.

REFERENCES

[1] M. R. N. Kousarrizi, A. A. Ghanbari, A. Gharaviri, M. Teshnehlab, and M. Aliyari, "Classification of Alcoholics and Non-Alcoholics via EEG Using SVM and Neural Networks," in *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*, 2009, pp. 1–4.

[2] S. Wang, Y. Li, P. Wen, and D. Lai, "Data selection in EEG signals classification," *Australas. Phys. Eng. Sci. Med.*, vol. 39, no. 1, pp. 157–165, Mar 2016.

[3] H. Tjandrasa and S. Djanali, "Classification of EEG Signals Using Single Channel Independent Component Analysis, Power Spectrum, and Linear Discriminant Analysis," in *Lecture Notes in Electrical Engineering (LNEE)*, Springer, Vol. 387, pp. 259–268, 2016.

[4] H. Tjandrasa, S. Djanali, and F. X. Arunanto, "Feature extraction using combination of intrinsic mode functions and power spectrum for EEG signal classification," in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2016, pp. 1498–1502.

[5] N. T. Rachman, H. Tjandrasa, and C. Fatichah, "Alcoholism classification based on EEG data using Independent Component Analysis (ICA), Wavelet de-noising and Probabilistic Neural Network (PNN)," in *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2016, pp. 17–20.

[6] U. Rajendra Acharya, S. Vinitha Sree, A. P. C. Alvin, and J. S. Suri, "Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9072–9078, Agu 2012.

[7] W. Y. Budiman, H. Tjandrasa, and D. A. Navastara, "Classification of EEG signals using Common Spatial Pattern-Principle Component Analysis and Interval Type-2 Fuzzy Logic System," in *2016 International Conference on Information Communication Technology and Systems (ICTS)*, 2016,pp. 85–89.

[8] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6,pp. 1517–1525, Des 2004.

[9] S. Wang, S. Yin, and M. Jiang, "Hybrid Neural Network Based on GA-BP for Personal Credit Scoring," in *2008 Fourth International Conference on Natural Computation*, 2008, vol. 3, pp. 209–214.