Journal of Engineering Science and Technology
Vol. 14, No. 2 (2019) 578 - 588

WAVELET SUB-BAND ENERGY FOR FEATURE EXTRACTION OF ELECTRO ENCEPHALO GRAPH (EEG) SIGNALS

HINDARTO[1,*], ARIF MUNTASA[2]

[1]Departement of Informatics Engineering, Universitas Muhammadiyah Sidoarjo, Jl. Mojopahit 666B Sidoarjo, East Java, 61271, Indonesia
[2]Departement of Informatics Engineering UTM - East Java - Indonesia
*Corresponding Author: hindarto@umsida.ac.id

Abstract

This research describes the application of back propagation Neural Network as a classification and Discrete Wavelet Transformation for feature extraction by taking energy values on each sub-band of the Electro Encephalo Graph (EEG) signal wave. The purpose of this study was to identify the EEG signal used in the cursor movement. The aim of this study was to provide new communication and control options for people with severe motor disability. The data used are EEG data derived from BCI competition 2003 (BCI Competition 2003). This data contains class 0 data (for upward cursor movement) and class 1 (for downward cursor movement). Decision-making is done in two stages. In the first stage, the energy values in each discrete wavelet sub-band are used to extract features of the EEG signal data. This feature is as the input on back propagation Neural Network. The second stage is identifying process into two classes (class 0 and class 1) of EEG signal data files. There are 260 training data files of EEG signals and 293 of the EEG signal data file testing, so the total is 553 data files of EEG signals. The result obtained for the classification of EEG signals is 73.5% of the tested signal data.

Keywords: Back propagation, BCI, Discrete wavelet, EEG.

1. Introduction

To move a cursor on the computer screen, someone usually needs a keyboard or mouse to run it. This is not possible with someone who does not have a hand or someone who can move his hand. Initially, it may be just wishful thinking, however, the creative and revolutionary ideas of researchers both local and foreign ones to be able to move the cursor without using the hands always appear.

Hans Berger was a German psychologist, in 1929 he claimed the existence of a weak electrical currents generated by the recording of the brain without opening the brain. The results of brain recording can be painted on a paper. He named the brain recording with Electroencephalography (EEG), so it can connect the brain and the object, which is being controlled by the thought by using a tool called Brain Computer Interface (BCI). BCI is a system that can analyse and acquire neural signals to create a communication channel between the computer and the brain. BCI can be shaped into systems provided by human muscles [1]. With BCI, someone can make a command to an electronic device using the brain [2]. Playing a simple game can also be done with BCI a by system [3].

Some studies take samples of data based on data sets from BCI Competition 2003 - Data Set It (EEG signal data to move the cursor up and down controlled by the human mind). Among them were Mensh, BD, Werfel, J, Seung, HS, in 2004. Their data consisted of four channels and four features (two of the average of the SPC and two of the gamma band power). The results of the classification process were 88.7% [4]. Subsequent research was by Wang et al. [5] in 2005. Researchers used two channels and four features by combining slow cortical potentials (SCPs) and wavelet packet transforms. The results of the classification process is 91.47% [5]. Other researchers were Ting et al. [6] in 2007, in which, the research they conducted were using six channels and took seventeen features with neural network as a classification process. The result of the classification process is 90.80% [6]. In 2005, Sun and Zhang [7] used the 2003 competition research data using six features and used seven features, namely RMS, spectral centroid, bandwidth, zero crossing rate, spectral roll-off frequency, band energy ratio and delta spectrum magnitude with Bayesian as a classification process. The result of the classification process is 90.44% [7]. In 2010, Kayikcioglu and Aydemir [8], in addition, used BCI 2003 competition data using one channel (channel 1) as experimental data and took 2 features using polynomial fitting method by taking feature of h value and b coefficient with KNN as the classification process. The result of the classification process is 92.15% [8].

Prochazka et al. [9] presented the segmentation for EEG signal and analysed using harmonic wavelet transform with the EEG signal feature extraction using the wavelet method. Many researchers used the wavelet method for EEG signal feature extraction. Therefore, there is a feature for a scale of 1, 2 and 3, which includes three frequency bands with different time scales of resolution [9]. Analysing EEG signal recording against epileptic patients using wavelet transform [10] is done by taking the value of the minimum, maximum, average and median of wavelet transforms for feature extraction of EEG signals against the epilepsy disease [11].

This study presented a new approach based on Artificial Neural Networks (ANN). This can be used for classifying cursor movements. The signal processing method using the Wavelet feature presents Transform EEG signals to move the cursor up or down on the computer screen while (at the same time) the SCP is recorded. ANN are used to classify cursor movements when energy as features retrieved from a sub-band Wavelet Transform is used as input.

2. Materials and Methods

This study used EEG signal dataset, which was taken from the BCI 2003 competition data. Six EEG channels were used and recorded from a healthy subject. Then, regarding sampling rate, we used 256 Hz with a recording time of every 3.5 second. Each experiment was done in each channel containing 896 samples. Subjects were tested to imagine about the moving cursor up or down on the computer screen while (at the same time) the SCP was recorded. Then, subjects received a visual feedback from SCPs as feedback phases. For the analysis, we divided the dataset into training (containing 268 experiments) and trials (containing 293 experiments). This experiment was done based on BCI 2003 Ia in literature [12-18].

2.1. Wavelet transform

The Wavelet theory brings an integrated framework for a number of techniques developed for various signal processing applications. In particular, it is interesting for non-stationary signal analysis, such as EEG, as it provides an alternative to the classic short time Fourier Transform (STFT) or Gabor transformation. The fundamental difference is that, unlike STFT, which uses a single analysis window, Wavelet Transform (WT) uses short windows at high frequencies and long windows at lower frequencies. This is similar to "Constant Q" or the relative bandwidth of the conventional bandwidth [12, 13].

2.2. Discrete Wavelet Transform (DWT)

Stationary signals are signals that do not change much over time. In signal processing, all stationary can use Fourier Transform method. However, EEG has many signals. EEG signals can contain non-stationary signals. Therefore, the Fourier transform is not ideal to apply for EEG signal. To overcome this, the wavelet method can be used. In wavelet analysis, the signal relating to different probing functions can be used. This analysis leads to a decisive equation for continuous wavelet transform (CWT):

$$W(a,b) = \int_{-\sim}^{\sim} x(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

Symbol b is used in x (t) and the action variable to vary the time scale of the probing function is ψ. If it is larger than one, the value of wavelet function (ψ) can be withdrawn at all times and if it is not smaller than one (however, still positive) contact function. The probing function can be one of the several different functions. And, the oscillatory form is always taken with the term "wavelet". Complex conjugation operation, and normalization factor $1/\sqrt{|a|}$ ensuring the same energy for all a values. In applications requiring bilateral transformations, the transformation produces the minimum number of required coefficients accurately. Discrete Wavelet Transforms (DWT) achieve this economically by limiting variations in scale and translation, usually to power 2. Mostly for signal processing and image applications, DWT-based analysis is best described in the form of a filter bank. The use of a group of filters to divide the signal into various spectral components is called sub-band coding. This procedure is classified as the multi-resolution decomposition of the signal $x[n]$. The first filters h [.] is the discrete, high-pass in nature. And, the second filters g [.] is a reflective, low-in-nature version. The bottom output of the sample from the first high-pass and low-pass filters provides detail, approximately $A1$ and respectively $D1$ [10, 14].

By using DWT, the proper wavelet selection and number of decomposition levels are very important in signal analysis. The dominant frequency component of the signal is used to select the number of decomposition levels. The decomposition rate is chosen so that portions of the signal are correlated well with the frequency required for signal classification to be maintained in the wavelet coefficients. The number of levels is chosen to be 5 because the EEG signal has no useful frequency components above 30 Hz. Thus, the signal is decomposed into details $D1$-$D5$ and one last approach, $A5$. These detailed estimates and records are reconstructed from the Daubechies 4 (DB4) wavelet filter [15, 16]. The extracted wavelet coefficient provides a concise representation showing the distribution of EEG signal energy in time and frequency. To further reduce the dimensions of the extracted feature vector, the EEG signal characteristic extraction is obtained by decomposing the signal up to 5 levels using discrete wavelet transforms. The wavelet function used is db4. Illustration of decomposition of a 5-degree EEG signal with a 256 Hz snap frequency is shown in Fig. 2. Each EEG signal is decomposed up to 5 levels to obtain detailed signals $D1$, $D2$, $D3$, $D4$, and $D5$ and approximation signals A5. The average decomposition energy of the detailed signal per sub-band is calculated by Eq. (2):

$$E_{Di} = \sum \frac{Di(k)^2}{Lenght\ Di} \tag{2}$$

where: k = 1, 2, ..., Length D_i., i = 1, 2,...N=5

The average decomposition energy of the approximation signal A5 is calculated by Eq. (3):

$$E_{A5} = \sum \frac{A_5(k)^2}{Lenght\ A_5} \tag{3}$$

where: k = 1, 2, .. . Length A_5

Since it will be the neural network input, the average energy of each decomposition signal is normalized by dividing the largest average energy between the decomposition average energy in each signal:

$$E_{nj} = \frac{E_i}{Maks\ (E_{Di}, E_{A5})} \tag{4}$$

where: j = 1, 2, 3, …..n=4

As a result of normalization, the value of extraction of Enj properties is between 0 and 1.

2.3. Back propagation neural network

ANN have certain performance characteristics, such as biologists neural network. Thus, it can be used as an information processing system [19, 20]. ANN has been used and developed as a generalization of the mathematical model in human cognition or biological nerves. The neural network is characterized by an interconnected pattern between neurons (design), the method of weighting the connection (learning or algorithm), and its activation function. The learning about counterfeit neural networks includes three stages: forward propagation, backward propagation and weight changes [17]. There are two main learning parameters in the reversal of learning rate α and momentum μ. The rate of learning is used to regulate the rapidity of learning. Momentum is used to avoid significant changes in weight due to different data from others.

The classification process consists of two stages, first the learning process or training, both testing and testing process. The training process is done and used to get the best weight value by obtaining the smallest value in error condition from the desired output target. Back propagation algorithm to change the weight value uses an output error with backward direction. Previously you have to do the forward propagation stage to get an error value.

Neurons will be activated using the sigmoid activation function. If the backward output in the hidden layer is not the same as the desired output, it will be forwarded to the input layer. Calculation of the level of error is done to analyse the number of errors that occur between the actual data and forecasting data. Calculation of forecasting error rates uses mean square error (MSE). In the back propagation algorithm, the error value can be minimized by using Eq. (5):

$$MSE = \frac{\sum_{k=1}^{a}(t_k - y_k)^2}{a} \tag{5}$$

where: a = total data, t_k = target input value, y_k = Output actual.

The back propagation training algorithm is as follows :

- Step 0: Initialize all weights with small random numbers, epoch = 1, test the learning rate (α), specify the number of units on the hidden screen (p), and specify the termination conditions. Termination conditions are maximum epoch and target error.
- Step 1: If the epoch is not in the maximum condition (epoch ≠ max epoch) and target error is less than MSE, do steps 2 - 9.
- Step 2: For each pair of training data (1 to α where α is the amount of training data), do steps 3 - 8.
- Phase I: Forward propagation.

Step 3: Each input unit receives a signal and passes it to the hidden unit above it.

Step 4: Calculate all output in the unit hidden z_j (j = 1, 2, ..., p).

$$z_{net_j} = v_{j0} + \sum_{i=1}^{n} x_i v_{ji} \tag{6}$$

$$z_j = f\left(z_{net_j}\right) = \frac{1}{1+e^{-z\_net_j}} \tag{7}$$

Step 5: Calculate all network output in the unit

y_k(k = 1, 2,....., m)

$$y_{net_k} = w_{k0} + \sum_{j=1}^{p} z_j w_{kj} \tag{8}$$

$$y_k = f\left(y_{net_k}\right) = \frac{1}{1+e^{-y\_net_k}} \tag{9}$$

Phase II: Backwards Propagation

Step 6: Calculate the factor δ output unit based on errors in each output unit y_k, (k = 1, 2, ..., m)

$\delta_k = (t_k - y_k)f'(y\_net_k) = (t_k - y_k)\, yk(1 - y_k)$ (10)

Calculate the rate of change in weight w_{kj} with the acceleration rate α

$\Delta w_{kj} = \alpha\delta_k z_j$ (11)

where $k = 1, 2, ..., m$; $j = 0, 1, ..., p$ (11)

Step 7: Calculate the factor δ hidden units based on errors in each hidden unit $Z_j(j= 1, 2, ..., p)$

$\delta\_net_j = \sum_{k=1}^{m} \delta_k w_{kj}$ (12)

$\delta_j = \delta_{net_j} f'\left(z_{net_j}\right) = \delta_{net_j} z_j\left(1 - z_j\right)$ (13)

Calculate the rate of change in weight v_{ji} with the acceleration rate α

$\Delta v_{ji} = \alpha\delta_j x_i$ (14)

Phase III: Modification in weight

Step 8: Calculate all weight changes. Change the weight of the line leading to the output unit

$W_{kj}(new) = w_{kj}(now) + \Delta w_{kj}$ (15)

where $k= 1, 2, ..., m$; $j = 0, 1, ..., p$

Change the line weight to the hidden unit:

$V_{ji}\,(new) = v_{ji}\,(now) + \Delta v_{ji}$ (16)

where $j = 1, 2, ..., p$; $i = 0, 1, ..., n$

Step 9: Update the epoch value

epoch = epoch + 1

and calculate mean squared error (*MSE*)

$MSE = \frac{\sum_{k=1}^{a}(t_k - y_k)^2}{a}$ (17)

Generally in the training process, the results of back propagation training will not produce MSE = 0 (with a lot of training data), however, most people are quite satisfied with the results of MSE = 0.1. If the MSE value is increased from 1 to 0.000000001 it will increase the rate of understanding, and back propagation training will be faster, however, if the understanding rate is too large it will have an impact on unstable algorithms (MSE does not decrease or even rises), which means this network does not recognize patterns and researchers will look for the best MSE values.

Part of the ANN is back propagation, where the design can be seen in Fig. 1. To research the input count of 4 neurons, with three hidden layers, and two classes as output.

In this study, the data classification process is conducted by separating the EEG signals into two parts, the data for the training process about 268 vector data and data for the data testing process about 293 data. This network has the input 4 (×1, ×2, ×3, ×4) derived from the DWT feature, hidden layer 1 has 10 nodes ($z1$, $z2$, ... $z10$), hidden layer 2 has 20 nodes ($w1$, $w2$, ... $w20$), hidden layer 3 has 10 nodes

($r1$, $r2$, ... $r10$) and binary type outputs for condition identification ($y1$, $y2$). Network design in this research is shown in Fig. 1. This figure shows output pattern with 2 target output in binary form. The patterns used in this study are shown in Table 1.

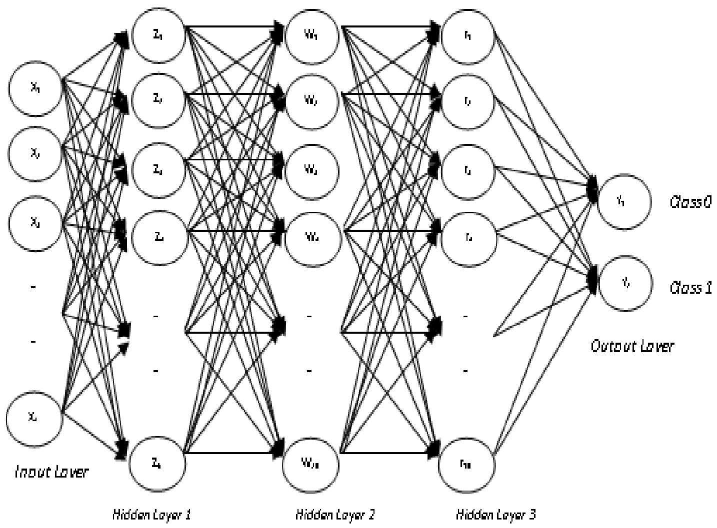

Fig. 1. Back propagation neural network design with 3 hidden layers.

Table 1. Output vector patterns.

No	Data Classification	Output Patterns
1.	Up Cursor Movement	0
2.	Down Cursor Movement	1

3. Results and Discussion

This study explains the detection of cursor movement of EEG signals obtained from BCI dataset Competition 2003. The EEG data is calculated using DWT as a feature extraction of EEG signals. In the DWT Process, the value of the EEG feature is derived from energy at the frequency of the DWT sub-band.

In Fig. 2, the presented EEG recording is divided into sub-band frequencies. It can be written as wavelet coefficients in $A5$, $D5$, $D4$, and $D3$ uses DWT. Wavelet sub-band frequencies of (0-4 Hz), (4-8 Hz), (8-16 Hz) and (16-32 Hz) are extracted to become EEG signal feature sets. The following features are used to explain the processing time. Then, the frequency distribution of the observed signal is the energy of the wavelet coefficients in each sub-band.

Figure 3 shows that energy in each sub-band for class 0 and class 1 has a different value. Different scores of values indicate that the classification rate by taking the energy value is good enough. Classification using Artificial Neural Network Back propagation is implemented using the energy value feature of the DWT process as input. In this study, the training set is about 260 sample data and trial set is about 293 sample data. In addition, we obtained 260 data samples (from normal subjects) for channel 1, which were used as training data. Then, we used

293 sample data (from normal subjects) for each channel for testing data. The class distribution of the prepared sample data in training and testing is shown in Table 2. To increase the capability of back propagation, two parameters (such as training and testing) were constructed by data obtained from different subjects. The data set for the training process is used as training for backpropagatioan, while for the testing data set is done as the accuracy as well as efficiency of back propagation in detecting cursor movements (for both up and down).

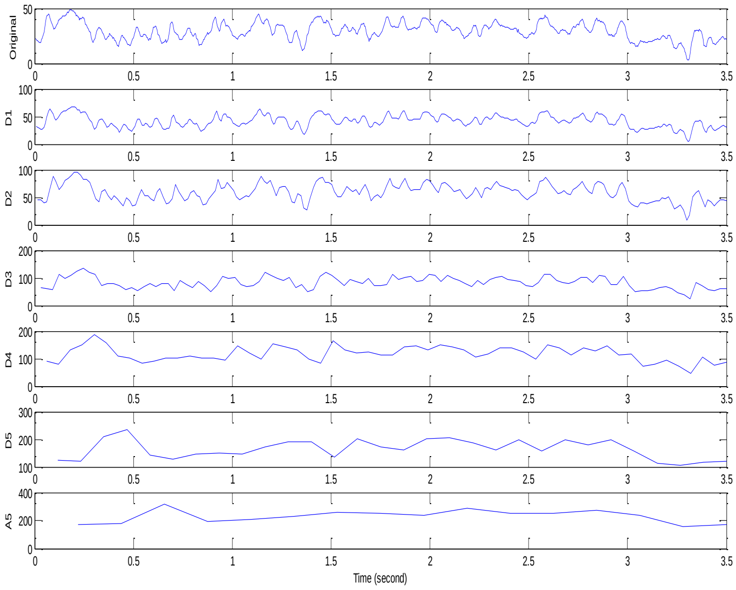

Fig. 2. Detailed coefficients of EEG signal taken from a healthy subject.

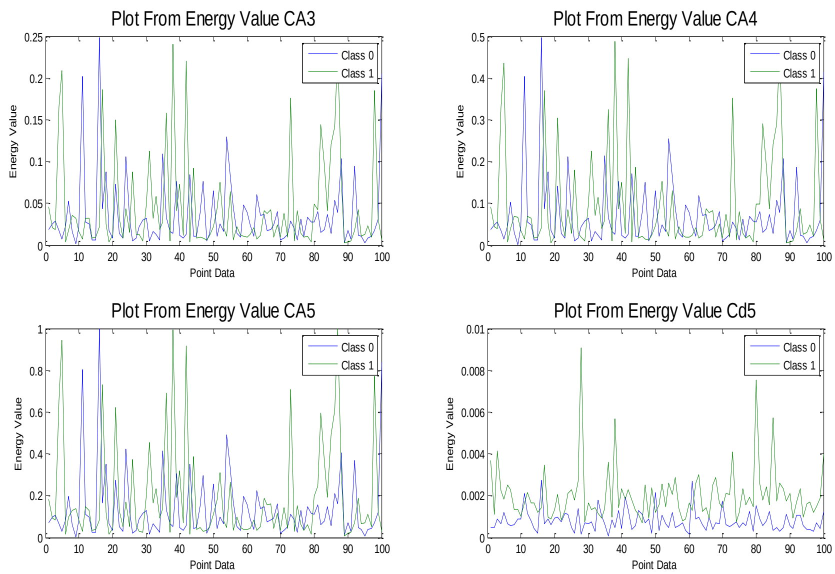

Fig. 3. Energy of the wavelet coefficients in each sub-band.

Figure 4 shows 260 training data from channel 1 in 831 training period. Then, the step size for the adaptation parameter has an initial value of between 9.97 and 10. Performance back propagation using 3 hidden layers is able to perform the training process by passing the minimum error limit, so that it has 100% accuracy of the training process.

Table 2. Class distribution of the samples in the training and test data sets.

Class	Training set	Test set
Up Cursor (class 0)	130×6 Channel	293× 6
Down Cursor (class 1)	130×6 Channel	channel (mix)

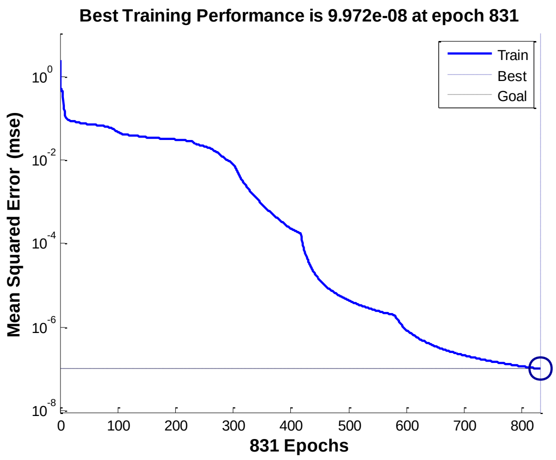

Fig. 4. Performance training of ANN using 3 hidden layers.

Table 3 presents the occupancy of channel 4 with a good degree of accuracy compared to other channels. Table 4 shows the degree of accuracy of the various hidden layers with the highest level of accuracy shown in back propagation with either 3 hidden layers that reach the level of accuracy of 73.5%. According to table 4, by using 3 hidden layers in back propagation, 73.5% accuracy value from the testing process can be achieved.

Table 3. Back propagation accuracy results with 3 hidden layers for all channels.

	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5	Channel 6
Accuracy	70,0 %	72,7 %	72,3 %	73,5 %	72,2 %	71,0 %

Table 4. ANN performance against different numbers of hidden Layer.

	MSE (1 Hidden Layer)	MSE (2 Hidden Layer)	MSE (3 Hidden Layer)
Time	33 seconds	86 seconds	154 seconds
Iteration	1000	624	831
MSE	$1{,}70.10^{-2}$	$9{,}98.10^{-8}$	$9{,}97.10^{-8}$
Accuracy	73,0 %	72,9 %	73,5 %

Based on the research by Ting et al. [6], the level of accuracy is still lower. However, both research studies used the neural network method for the classification process. For that, we need another method to be applied by researchers to improve classification accuracy and in addition, looking for different feature extraction models namely Wavelet Transform.

4. Conclusion

This paper introduces the Discrete Wavelet to extract features by taking energy values on each sub-band. The process of classifying EEG signals is divided into two classes, class 0 and class 1. This study uses 553 EEG signal data files for training and testing. The accuracy of back propagation classification reached 73.5% of test data. The future research will examine the suitable techniques for feature extraction and EEG signal classification, so the accuracy level of the command in moving the cursor would be better. The results obtained will be compared with the methods already studied.

Acknowledgements

We thank Universitas Muhammadiyah Sidoarjo for giving the time and facilitating our research. Our special gratitude to the Directorate of Research and Community Service, Directorate General of Research, Research and Development of the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for supporting the fund in this research.

References

1. Wolpaw, J.R.; Birbaumer, N.; McFarland, D.J.; Pfurtscheller, G.; and Vaughan, T.M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6), 767-791.
2. Wolpaw, J.R.; Birbaumer, N.; Heetderks, W.J.; McFarland, D.J.; Peckham, P.H.; Schalk, G.; Donchin, E.; and Quatrano, L.A. (2000). Brain-computer interface technology: A review of the first international meeting. *IEEE Transactions on Rehalibitation Engineering*, 8(2), 164-173.
3. Pour, P.A.; Gulrez, T.; AlZoubi, O.; Gargiulo, G.; and Calvo, R.A. (2008). Brain-computer interface: Next generation thought controlled distributed video game development platform. *Proceedings of the IEEE Symposium Computational Intelligence and Games. Perth, Australia*, 251-257.
4. Mensh, B.D.; Werfel, J.; and Seung, H.S. (2004). BCI Competition 2003-Data set Ia: Combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals. *IEEE Transactions on Biomedical Engineering*, 51(6), 1052-1056.
5. Wang, B.; Jun, L.; Bai, J.; Peng, L.; Li, G.; and Li, Y. (2005). EEG recognition based on multiple types of information by using wavelet packet transform and neural networks. *Proceedings of the IEEE 27th Annual Conference on Engineering in Medicine and Biology*. Shanghai, China, 5377-5380.
6. Ting, W.; Guo-zheng, Y.; Bang-hua, Y.; and Hong, S. (2005). EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement*, 41(6), 618-625.

7. Sun, S.; and Zhang, C. (2005). Assessing features for electroencephalographic signal categorization. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing.* Philadelphia, United States of America, 417-420.
8. Kayikcioglu, T.; and Aydemir, O. (2010). A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data. *Pattern Recognition Letters*, 31(11), 1207-1215.
9. Prochazka, A.; Kukal, J.; and Vysata, O. (2008). Wavelet transform use for feature extraction and EEG signal segments classification. *Proceedings of the 3rd International Symposium on Communications, Control, and Signal Processing.* St. Julians, Malta, 719-722.
10. Adeli, H.; Zhou, Z.; and Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods*. 123(1), 69-87.
11. Garg, S.; and Narvey, R. (2013). Denoising and feature extraction of EEG signal using wavelet transform. *International Journal of Engineering Science and Technology (IJEST)*, 5(6), 1249-1253.
12. Barford, L.A.; Fazzio, R.S.; and Smith, D.R. (1992). An introduction to wavelets. Hewlett-Packard Instruments and Photonics Laboratory, *HPL-92-124*, 29 pages.
13. Chavan, A.S.; and Kolte, M. (2011). EEG signal preprocessing using wavelet transform. *International Journal of Electronics Engineering*, 3(1), 5-10.
14. Murugappan, M.; Ramachandran, N.; and Sazali, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering*, 390-396.
15. Nakate, A.; and Bahirgonde, P.D. (2015) Feature extraction of EEG signal using wavelet transform. *International Journal of Computer Applications*, 124(2), 21-24.
16. Alomari, M.H.; Awada, E.A.; Samaha, A.; and Alkamha, K. (2014). Wavelet-based feature extraction for the analysis of EEG signals associated with imagined fists and feet movements. *Computer and Information Science*, 7(2), 17-27.
17. Zilouchian, A. (2001). Fundamentals of neural network. *Intelligent Control Systems using Soft Computing Methodologies*, Chapter 2, 42-62.
18. BCI Competition II. (2018). BCI Competition II. Retrieved November 1, 2018, from http://www.bbci.de/competition/ii/.
19. Farobie, O.; and Hasanah, N. (2016). Artificial neural network approach to predict biodiesel production in supercritical Tert-Butyl Methyl Ether. *Indonesian Journal of Science and Technology*, 1(1), 23-36.
20. Abdullah, A.G.; Sopian, W.W.; Arasid, W.; Nandiyanto, A.B.D.; Danuwijaya, A.A.; and Abdullah, C.U. (2018). Short-term peak load forecasting using PSO-ANN methods: The case of Indonesia. *Journal of Engineering Science and Technology* (JESTEC), 13(8), 2395-2404.