clasification EEG

by Arif M Hindarto

PAPER • OPEN ACCESS

Classification of EEG signals uses the coefficient of wavelet transform and K-nearest neighbor

To cite this article: H Hindarto *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **821** 012036

Classification of EEG signals uses the coefficient of wavelet transform and K-nearest neighbor

H Hindarto[1,*], A Muntasa[2] , A Efiyanti[3]

[1,3] Department of Informatics Engineering, Faculty of Sains and Technology, Universitas Muhammadiyah Sidoarjo, Indonesia.
[2] Informatics Engineering Department, Faculty of Engineering, Universitas Trunojoyo Madura, Indonesia

*Email: hindarto@umsida.ac.id

Abstract. The research that was built was used to explain the application of Electroencephalography (EEG) signal waves. EEG data is used to move the cursor up and the cursor down. Characterization of each EEG signal uses the Wavelet method taken at each the subband of the wavelet process. Wavelet process by taking the average value and the standard deviation value of the wavelet coefficient. The average value and the standard deviation value is used as an EEG feature. K-Nearest Neighbor is used as an identification whether the cursor will move up or vice versa. This study uses of 100 EEG signal data consisting of 50 test data and 50 testing data. The accuracy of identification uses the 80% K-NN method.

1. Introduction

A keyboard or mouse is usually used by someone to move the cursor on a computer screen. Someone who has physical limitations, especially people who do not have hands and feet will be difficult to run computer applications using a keyboard or mouse. Researchers from both inside and outside the country have the idea of overcoming physical limitations by not being able to run computer applications, including using the signal brain that is coiled on the scalp without opening the brain skull. The head brain that is recorded will produce an electrical signal whose current is very weak. This head brain recording is called the Electroencephalography (EEG) signal. The first researcher who tried to produce this EEG signal was German Hans Berger. one of the tools used to communicate with humans is the Brain Computer Interface (BCI), this tool is used to communicate through brain signals. so with this tool, humans can control other equipment by just thinking. BCI is a system that obtains and analyzes nerve signals to create direct communication channels between the brain (EEG signals) and computers. With the BCI tool, the relationship between humans and other equipment does not require the activity of muscle strength [1]. An order from someone to electronic equipment can be responded to using brain signals through the BCI system [2]. Many uses of BCI include playing games through cellphones [3].

The EEG signal is obtained from recording sensors that are affixed to the head and can be used to communicate directly with other equipment through the BCI system. The usefulness of BCI is very much, by using the mind the human brain can control any object or tool [2-4]. To identify the EEG signal is a separate problem, both the method of finding EEG signal characteristics, the method of EEG signal classification, even the problem of the design of inputs and equipment used for applications of BCI systems. Various attributes are used as input of EEG signals for communication with BCI, including alpha rhythm, beta rhythm or activities related to a BCI device. For example BCI by using P300 [2-4].

Research on identifying epilepsy taking EEG signals and the method used is the Fourier Transform method as feature extraction and Neural Network as a classification has been carried out [5]. Research using a combination of slow cortical potential (SCP) and wavelet packet transformation as a feature extraction and classification using the Artificial Neural Network method, so as to obtain an accuracy of 91.47% [6]. EEG signal classification by obtaining an accuracy of 90.80% using the wavelet packet decomposition method as cirri extraction and nerve tissue as an identification process has been carried out [7].

This research tries to find the feature extraction method and the EEG signal classification method. The method used is the DWT method in the process of finding feature extraction and the K-NN method in the classification process. The EEG signal used is the EEG signal whose recording process imagines the cursor movement up and the cursor movement down.

2. Data and methods

2.1. Material

This research uses data from the 2003 BCI competition, the data from the 2003 BCI competition is an EEG signal recording conducted by Dr. Birbaumer and his team at the University of Tuebingen, Germany (Blankertz 2004). EEG signal data is obtained from recording healthy subjects consisting of six electrode sensors that are affixed to the scalp. In the process of recording the subject is asked to imagine the cursor moving up and down. The recording process lasts for 3.5 seconds and is sampled at a frequency of 256 Hz. This trial data set consisted of 268 trial data and 293 test data [8]. In this study EEG signal data taken in the research process were only taken 50 test data and 50 test data.

2.2. Discrete Wavelet Transformation (DWT)

DWT is one of the feature extractions used in the polo recognition process, which aims to reduce the input dimensions. Input data that is too large can result in a long process, so that the necessary representation of features in this case the representation of these features does not reduce the information needed. The selection of features must be done carefully and accurately and can be used as expected information. Many studies that have tried to find suitable feature extraction for EEG signals, one of them is the use of feature extraction using wavelet transforms can be found in the following studies [9, 10].

The representation of EEG signals in the time domain is still imperfect when using representations using the frequency domain. By using the frequency domain, it provides a good description in non-stationary messages, one of which uses wavelet transforms. Wavelet transform is more suitable for feature extraction for EEG signals [11]. Feature extraction using wavelet transform can represent good resolution for low frequency and high frequency. So the scale owned by wavelet can represent a certain violence from the message to be tried. The formula for multi-axis decomposition of the x [n] signal will be schematically shown in Figure 1.

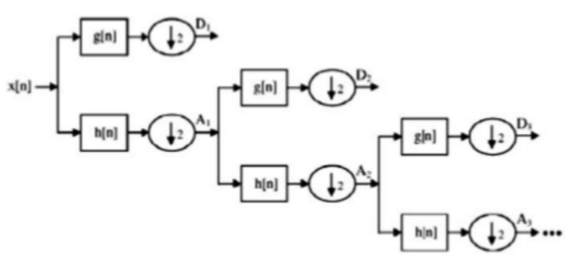

Figure 1: Subband decomposition of discrete wavelet transform

In this study the EEG signal is broken down to breakdown D5. To detect changes in the EEG signal daubechies wavelet sequence 2 is used because it displays the interior more precisely. The feature vector will represent can be obtained from the details of the EEG signal that is calculated and estimated by the wavelet coefficient of the EEG signal. This study takes the statistical value of the wavelet coefficient which can be used to reduce the dimensions of vector features. The statistical characteristics used to represent the frequency-time distribution of EEG signals, namely the average value and the standard deviation value of the wavelet coefficient for each subband.

2.3 K-nearest neighbor classification method

The K-NN algorithm is an easy and simple machine learning algorithm. The essence of the K-NN algorithm is a subject that is close to each other, then the subject has the same characteristics. Thus to find out the characteristics of one object, then we can also predict other objects based on the nearest neighbor. K-NN is the idea that each new instance can be classified based on the most votes of neighbor k, where k is a positive integer, and usually with a small number [12]. The K-NN classification algorithm predicts the category of the test sample according to the training sample k is the closest neighbor to the test sample, and puts it in the category that has the largest probability category [13].

In pattern recognition, the KNN algorithm is a method used to classify objects based on the closest training example in the feature space. KNN is a type of institutional-based learning, or lazy learning where this function is only approached locally and all calculations are deferred to classification [14]. The K-NN classification method has several stages, the first being the k value which is the number of closest neighbors that will determine which new query goes to which class is determined. The second stage, k the nearest neighbor is searched by calculating the distance of the query point with the training point. The third stage, after knowing the distance of each training point with the query point, then see the smallest value. The fourth stage takes the smallest k value, then see the class. The class that is the most is the class of the new query. The distance or distance of a point with its neighbors can be calculated using the Euclidean distance. The euclidean distance is represented as follows [15]:

$$j(a,b) = \sqrt{\sum_{k=1}^{k_n} (a_k - b_k)^2} \tag{1}$$

J (a, b) is the distance between point a which is the point known to its class and b is a new point. The distance between the new point and the training points is calculated and taken by the nearest point. The new point is predicted to enter the class with the highest classification of these points.

3. Results and discussion

This research processes data from EEG signals from 2003 BCI competition data, the data consists of training data and testing data. The initial process is to separate a set of data from 6 channels into data for each channel. Each channel is then processed using the DWT process. From this DWT process, the average value and standard deviation for each subband are taken. Thus the process of finding feature extraction is used to reduce the dimensions of the data, i.e. the initially large amount of data becomes less.

The feature search process via DWT for EEG signals is taken for coefficient values of A5, D5, D4 and D3 (Figure 2). This process is carried out because the coefficients of A5, D5, D4 and D3 enter the frequency of alpha, beta, gamma and tetha waves. Figure 3 is the result of the DWT process which is then taken the average value and the standard deviation of each subband.

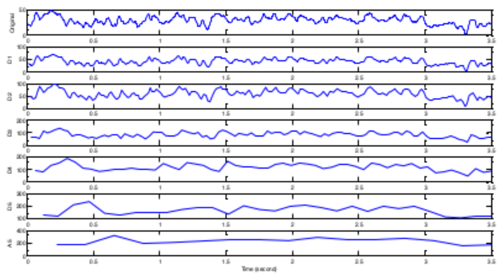

Figure 2. The results of the DWT process

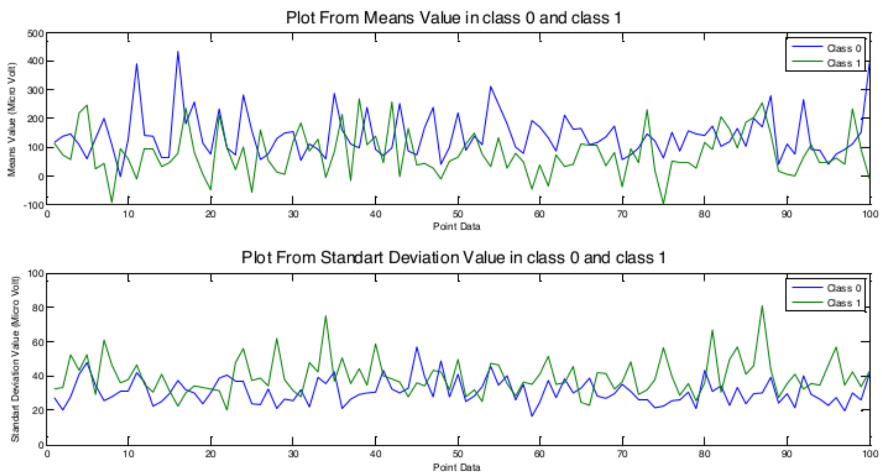

Figure 3. The results of the process of taking the average value and the standard deviation value

To train K-NN, use training data sets, while to verify the accuracy and effectiveness of K-NN test data to detect cursor movements up and down. The test will be carried out with variations in the K value in the KNN system and variations in the training data and test data used. The accuracy in question is the accuracy of the classification results carried out by the program with manual classification results. The level of accuracy can be formulated with Equation 2.

$$Level\ of\ accuracy = \frac{The\ number\ of\ EEG\ signals\ is\ correctly\ classified}{Total\ amount\ of\ test\ data}\ x\ 100\% \qquad (2)$$

In this test variations in the number of k are performed on the KNN function. k is the number of closest neighbors. The k values tested are one, three, five, seven, and nine. Odd values are chosen to avoid similarities in proximity to two different points of class. Because KNN will classify based on the most class voting. From this test the average accuracy value for each k value is determined.

- Classification testing with value k = 1, k = 3, k = 5, k = 7, and k = 9
  The classification results with values k = 1 to k = 9 are shown in Table 1. The test results with k = 1 obtain the highest accuracy with a value of 80% and the average value of the KNN classification of 67%. For k = 3 the highest accuracy value is 72% and the lowest value is 48%. For k = 5 the

highest accuracy value is 76% and the lowest is 48%. For k = 7 the highest value is 72% and the lowest is 48%. For k = 9 the highest accuracy value is 76% and the lowest is 48%.

Tabel 1. Results of testing k = 1, k=3, k=5, k=7 dan k=9

Test	training data	test data	Accuracy K=1	Accuracy K=3	Accuracy K=5	Accuracy K=7	Accuracy K=9
1	50	50	80%	72%	76%	72%	76%
2	50	50	68%	60%	60%	60%	56%
3	50	50	64%	52%	40%	52%	48%
4	50	50	60%	60%	60%	60%	64%
5	50	50	60%	48%	48%	48%	52%
Average			67%	59%	57%	58%	59%

- Comparison of the average value of accuracy
  From all tests with variations in the value of k obtained the highest accuracy value is found at the value of k = 1 with an average level of accuracy of 67%. The lowest accuracy value is found in testing with k = 5 with an average level of accuracy of 57%. Overall the value of accuracy has a value close to 60%. Classification of KNN with the description above, the value of k = 1 is the most optimal k value.

4. Conclusion
This study aims to find a cirri extraction method and a good identification method to recognize EEG signals, namely recording the EEG signal of a subject in imagining right cursor movement and imagining cursor movement to the left. The method used for extracting EEG signal characteristics is using the DWT method and the method used for the identification process is the K-NN method. The results of this study are reaching an accuracy value of 80% in the testing process.

Acknowledgement
The authors express our deep gratitude to Umsida and the Directorate of Higher Education for providing time and money in the research process that is being investigated by researchers.

References
[1] Wolpawa J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *Clin Neurophysiol* vol **113** no 6 pp 767–91.

[2] Vaughan T M 2000 Guest editorial. Brain-Computer interface technology: a review of the second international meeting *IEEE Trans Neural Syst Rehabil Eng* vol **11** no 2 pp 94–109.

[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 *2008 IEEE Symp Comput Intell Games* pp 251–257.

[4] Wolpaw J. R., Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *Clin Neurophysiol* vol **113** no 6 pp 767–91.

[5] Puloma P 2003 Classification of electroencephalogram (EEG) signal based on fourier transform and neural network *A Thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Technology in Electrical Engineering* Department of Electrical Engineering National Institute of Technology Rourkela-769008 (ODISHA).

[6] Baojun W, Liu J, Jing B, Le P, Yan L, Guang L 2005 EEG recognition based on multiple types of information by using wavelet packet transform and neural networks Engineering in Medicine and Biology Society IEEE-EMBS 2005 *27th Annual International Conference* of the Issue Date 17-18 Jan. 2006 pp 5377 – 5380

[7] Wu T, Yan Guo Z, Yang Bang H, Sun H 2007 EEG feature extraction based on wavelet packet decomposition for the brain-computer interface 0263-2241/$ - see front matter _ 2007 Elsevier Ltd. All rights reserved. doi: 10.1016/ j.measurement.2007.07.007.

[8] http://bbci.de/competition/ii/

[9] Dipti U Classification of EEG signals under different mental tasks using wavelet transform and neural network with one step secant algorithm *International Journal of Scientific Engineering and Technology* vol **2**(4) pp 256-259.

[10] Pravin A, Kharat, Sanjay V, Dudul Daubechies Wavelet Neural Network Classifier for the Diagnosis of Epilepsy was transactions on biology and biomedicine.

[11] Rajya Lakshmi M, Prasad Dr T V, Chandra Prakash Dr V 2014 Survey on EEG signal processing methods *International Journal of Advanced Research in Computer Science and Software Engineering* vol **4**(1).

[12] Khamis, Kipruto H S, W C and Kimani S 2014 Application of k-nearest neighbor classification in medical data mining *International Journal of Information and Communication Technology Research* vol **4**(4).

[13] Suguna N and Thanushkodi K 2010 An improved k-nearest neighbor classification using genetic algorithm *International Journal of Computer Science Issues* vol **7** Issue 4 No 2.

[14] Imandoust S B and Bolandraftar M 2013 Application of K-Nearest Neighbor (KNN) approach for predicting economic events*:* Theoritical Background *International Journal of Engineering Research and Applications* vol **3** issue 5.

[15] Pramesti R P A 2013 Identifikasi karakter plat nomor kendaraan menggunakan ekstraksi fitur ICZ dan ZCZ dengan metode klasifikasi KNN *Scientific Repository of Bogor Agricultural University*.

clasification EEG

ORIGINALITY REPORT

11%	11%	0%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	mafiadoc.com Internet Source	6%
2	www.researchgate.net Internet Source	5%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		