

# sinyal EEG

*by* Madiun Madiun

---

**Submission date:** 08-Apr-2019 03:15PM (UTC+0700)

**Submission ID:** 1107990899

**File name:** jurnal\_1.docx (144.69K)

**Word count:** 3388

**Character count:** 17707

## Identification of Electroencephalograph signals using Sampling Technique and K - Nearest Neighbor

I Hindarto, Zarif Muntasa, E Ade Effiyanti

<sup>1</sup>Universitas Muhammadiyah Sidoarjo,

<sup>2</sup>Universitas Trunojoyo Madura

hindarto@umsida.ac.id

### Abstract

Electroencephalograph is a device that can help humans observe and analyze the results of electrical waves produced by neurons in the brain. The results of reading the tool are called Electroencephalogram (EEG), besides being able to help diagnose physician for medical therapy, it is also developed for the Brain-Computer Interfacing (BCI) application. BCI is a method that allows humans to be able to control an external system without direct contact with the system. Research on communication between humans to control external equipment has been widely investigated, including research on brain activity to control a cursor on a computer screen. This study focuses on feature extraction for Electro Encephalo Graph (EEG) signals using the sampling technique. K - Nearest Neighbor is used as a classification of EEG signals to determine whether the cursor moves up or down. The data used are EEG data originating from the 2005 BCI competition (BCI 2005 Competition). Decision making is done to classify the cursor movement up and down the cursor movement. The research data uses 250 EEG signal file training data and 50 from EEG signal file testing data, so that the whole becomes 300 EEG signal data files. The best results with K = 3 values obtained for the classification of EEG signals using K-MN are 70% of the signal data tested.

*Keywords:* EEG, sampling technique, BCI, K-MN, Cursor

### 1. INTRODUCTION

The activity of recording electricity in the brain for electro physiological monitoring is called Electroencephalograph (EEG). Electrode placement is located along the scalp. The EEG measures voltage fluctuations that result from ion currents in brain neurons. In the clinical context, EEG refers to the recording of spontaneous electrical activity of the brain over a period of time, as recorded from several electrodes placed on the scalp [1]. Diagnostic applications generally focus on the potential associated with EEG spectral content.

Placing electrodes on the scalp follows a predetermined system, namely the 10-20 system. The right and good electrode placement is the main requirement to get good and reliable EEG recordings. Besides that the cleanliness of the scalp, the condition of the electrode, the EEG machine and the compliance of the subjects during recording are also very influential to get good results. Hans Berger stated that the human brain has continuous electrical activity that can be recorded. Brain activity can be possible to send commands to electronic equipment with the help of the Brain Computer Interface (BCI) [2]. Most BCI uses spontaneous mental activity (for example, imagining moving a finger, hand, or the entire arm, etc.) to produce a distinguishable electroencephalogram (EEG) signal [3]. EEG signals that can be distinguished are then converted into external actions. Over the few years, various evidences have evaluated the possibility of recognizing some mental tasks of EEG signals [4]. However, how to improve the performance of EEG signal recognition in signal processing is still a major problem. The introduction procedure mainly includes feature extraction and classification, where feature extraction plays an important role for classification, this study mainly focuses on feature extraction.

At present, the feature extraction method for EEG motor images mainly includes several methods including using the Fast Fourier transform (FFT) method [5] [6], the

Welch method using the windowed Fourier signal spectrum transformation is used to calculate the Fourier spectral feature. This method only uses frequency information and does not use time domain information so that it is a major disadvantage to this method. The combination of frequency information and time domain information can improve the performance of EEG signal classification in research [7]. Autoregressive (AR) is a method of the AR spectrum, band power is calculated in several bands frequency and the amount of power is used as an independent variable [8]. In addition, the features that are used are the AR or multivariate autoregressive (MVAR) model [9]. Time-frequency analysis by Wang et al. Uses time-frequency analysis as a useful tool for oscillating EEG components during motor image [10]. As we all know, the oscillating EEG component produced during the motor image is the time and frequency involved, therefore, this method gets promising results. However, the oscillating EEG component can cause a shift in slow cortical potential. A combination of two related signals may be used to improve the information extracted. The time-frequency method only considers oscillating EEG components. Utilizing the wavelet transformation coefficient, that is, extracting the wavelet transformation coefficient on a useful frequency band according to transcendent information [11]. However, the EEG production mechanism is rather complicated, so it is difficult to get accurate and rather inflexible transcendent information. Likewise with the sampling technique, the sampling technique (TS) method is applied to detect features of EEG data. One of the main important techniques in statistics is sampling. The characteristics of the EEG signal can only be known to be taken with the sample size. In statistics, there are various types of sampling techniques used [12]. The characteristics of EEG signals in two classes are in normal people and people who suffer from epilepsy using Simple Random Sampling (SRS). Feature extraction with the RBF kernel is obtained from both classes. The LS-SVM classification results obtained were 80.31% for training data and 80.05% for testing data [13]. Based on the background above, this study was designed as follows, the second part describes the material and the EEG signal method used in EEG feature search, the third part describes the results of feature extraction and the EEG signal classification process, and the fourth part describes the conclusions of this study.

## 2. MATERIALS AND METHODS

### 2.1. material

The 2003 BCI competition data comes from Dr. Birbaumer and his team at the University of Tuebingen, Germany (Blankertz 2004) are EEG signal datasets taken in this study [17]. Healthy subjects are recorded using Six EEG channels using a sampling rate of 256 Hz and a recording length of 3.5 seconds. The results of each experiment for each channel were 896 samples. In the process of recording subjects are asked to imagine moving the cursor up or moving the cursor down on the computer screen. Subjects received visual feedback from SCPs (feedback phase). The dataset consists of 268 training data and 293 trial data, according to the 2003 BCI description.

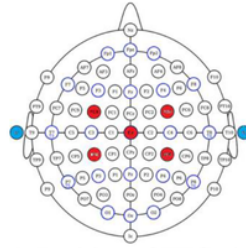


Figure 1. Position of EEG electrodes

Figure 1 shows the position of the electrode attached to the head. There are 6 sensors used, namely Cz for channel 1, A2-Cz for channel 2, FC3 for channel 3, CP3 for channel 4, FC4 for channel 5 and PC4 for channel 6 [17].

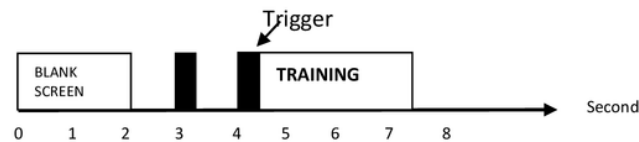


Figure 2. Timing pattern

This research takes one channel from the 6 channels installed, namely channel 3, each channel has two different classes. The data used for each class were 688 data for training and 293 data for trial data. The data used in this study are 250 training data and 50 trial data .

## 2.2. Sampling technique

In this study extraction The characteristics used are the Sampling Technique method. The EEG signal is divided into four sub-EEG signals. Of the four sub-EEG signals determined the maximum value, the minimum value, the average value and the standard deviation value to be extracted from the characteristics of the EEG signal as shown in figure 3.

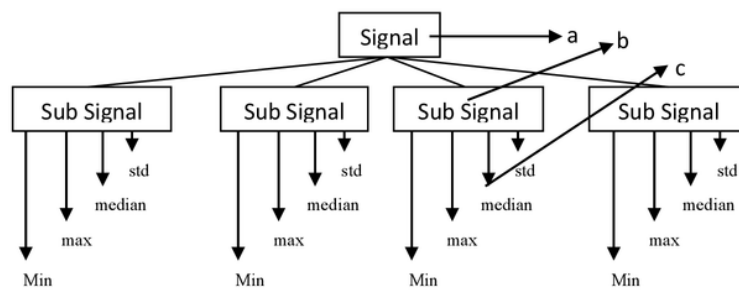


Figure 3. a. Signal sample, b. sub signal samples and c. Feature Selection

There are 300 EEG signal data files taken in this study. One EEG signal file data has 896 data points. One EEG signal is divided into four sub-signals. So that each sub signal has 214 points of data. The EEG sub signal data is searched for the minimum

value, maximum value, average value and standard deviation value. There are 20 points of data obtained from 4 x 4 sub-signals.

### 2.3 K-Nearest Neighbor Classification

K-NN is a simple machine learning algorithm. This is only based on the idea that an object that is 'close' to each other will also have similar characteristics. This means that if we know the characteristics of one object, we can also predict other objects based on their closest neighbors. K-NN is an advanced improvisation of the Nearest Neighbor classification technique. This is based on the idea that each new example can be classified by the majority vote of a neighbor  $k$ , where  $k$  is a positive integer, and usually with a small number [14]. The K-NN classification algorithm predicts the sample's closest category according to the training sample  $k$  which is the closest neighbor to the test sample, and enters into the category that has the largest probability category [15].

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 instance-based learning, or lazy learning where this function is only approached locally and all calculations are deferred to classification [16]. 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 value, then see the class. The class that is the most is the class of the new query. Near or far points with neighbors can be calculated using the Euclidean distance. *Euclidean distance is represented as follows :*

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

$J(a, b)$  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. New points are predicted to enter the class with the most classifications of these points.

### 3. Results And Discussion

Data used from BCI completion 2003 Data set Ia. The data set Ia consists of 6 channels (electrodes affixed to the scalp totaling 6 electrode sensors, resulting in 6 EEG signal channels). The large amount of data will cause the old computing process to be caused by a lot of data being processed, so that with a few features it results in a fast computing process. In this study, one EEG signal is divided into four sub-signals. The four sub-signals are taken maximum values, minimum values, average values and standard deviation values. The maximum value, minimum value, average value and standard deviation value of the Sampling Technique process are used as feature extraction for the identification process.

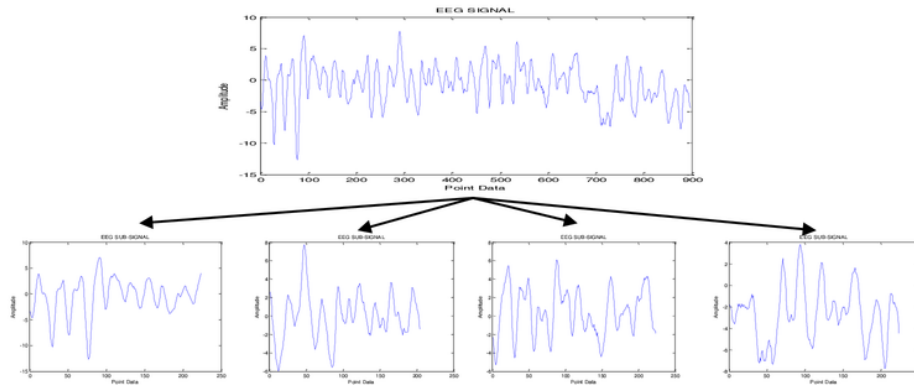


Figure 4. Results of the Sampling Technique process on EEG signals

Figure 4 is an EEG recording and the division process into four sub-signals. The four sub-signals are taken the maximum value, the minimum value, the average value and the standard deviation value for each sub-signal.

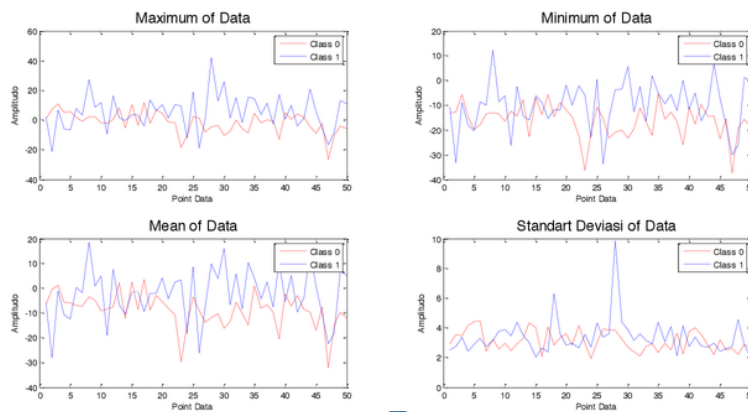


Figure 5. Comparison of class 0 and class 1 to the maximum value, minimum value, average value and standard deviation value of each sub signal

In Figure 5 shows that for values of maximum values, minimum values, average values and standard deviation values for each sub signal. Each signal for the cursor movement up and down the cursor has a difference value. With differences in values that are not the same, it shows that the level of classification by taking the maximum value, minimum value, average value and standard deviation value is quite good. Classification using K-NN is implemented using the maximum value feature, minimum value, average value and standard deviation value from the sampling technique as input. In this study, the training set numbered 250 EEG signal data and test data of 50 EEG signal data. table 1 is the distribution of sample classes in the training and test data collection. Data obtained from different subjects in the training process is way to improve the ability of K-NN. To train K-NN, training data sets, while to verify the accuracy and effectiveness of K-NN test data to detect cursor movements up and down.

Table 1. Sampling of class distribution on training data and test data

Class	Training set	Test set
Up cursor	125	50
Down cursor	125	

The test will be carried out with a variation of the K value in the KNN system and variations in training data and test data used. The accuracy in question is the accuracy of the classification results carried out by the program with the results of manual classification. The level of accuracy can be formulated with equation (2).

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

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 classes. 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 k Value = 1

The classification results with K = 1 are shown in Table 2. The test results with K = 1 obtained an average accuracy of KNN classification of 54.4%. With the lowest accuracy 48% and the highest accuracy 64%.

Table 2. Results of testing k = 1 classification

Test	Amount of training data	Amount of test data	Right	wrong	accuracy
1	50	50	13	12	52%
2	50	50	13	11	52%
3	50	50	14	11	56%
4	50	50	16	9	64%
5	50	50	12	13	48%
Average					54,4%

#### Classification Test with k = 3

The classification results with K = 3 values are shown in Table 3. The test results with K = 3 obtained an average accuracy of 64%. The highest level of accuracy is found in the fifth test of 76%. The lowest level of accuracy is found in the third test of 56%.

Table 3. Results of testing k = 3 classification

Test	Amount of training data	Amount of test data	Right	wrong	accuracy
1	50	50	14	11	56%
2	50	50	14	11	56%
3	50	50	16	9	64%
4	50	50	19	6	76%
5	50	50	17	8	68%
Average					64%

### Classification Test with k = 5

Hasil klasifikasi dengan nilai K= 5 ditunjukkan pada Tabel 4. Dari hasil pengujian dengan K=5 didapatkan rata-rata akurasi sebesar 54,4%. Akurasi tertinggi sebesar 64% terdapat pada pengujian keempat. Akurasi terendah sebesar 44% terdapat pada pengujian kesatu dan kedua.

Table 4. Results of testing k = 5 classification

Test	Amount of training data	Amount of test data	Right	wrong	accuracy
1	50	50	11	14	44%
2	50	50	11	14	44%
3	50	50	15	10	60%
4	50	50	16	9	64%
5	50	50	15	10	60%
Average					54,4%

### Classification Test with k = 7

The classification results with k = 7 are shown in Table 5. Testing with k = 7 produces an average accuracy of 60.8%. The highest accuracy results are found in the fifth test with a value of 68% accuracy. The lowest accuracy of 52% is in the third test.

Table 5. Results of testing k = 7 classification

Test	Amount of training data	Amount of test data	Right	wrong	accuracy
1	50	50	15	14	60%
2	50	50	15	14	60%
3	50	50	13	10	52%
4	50	50	16	9	64%
5	50	50	17	10	68%
Average					60,8%

### Classification Test with k = 9

The results of classification testing with k = 9 values are shown in Table 6. Classification results with a value of k = 9 indicate an average accuracy value of ten testing times of 62.4%. Of the five tests there were three tests which produced 64% accuracy, namely in the second, fourth and fifth tests. The lowest test value of 60% is found in the first and third tests.

Table 6. Results of testing k = 9 classification

Test	Amount of training data	Amount of test data	Right	wrong	accuracy
1	50	50	15	14	60%
2	50	50	16	14	64%
3	50	50	15	10	60%
4	50	50	16	9	64%
5	50	50	16	10	64%
Average					62,4%

### Comparison of Average Accuracy Values

Of all the tests with variations in the K value the highest accuracy value was found in the value of k = 3 with an average level of accuracy of 64%. The lowest accuracy value is found in tests with k = 1 and k = 5 with an average level of accuracy of



54.4%. Overall the value of accuracy has a value close to 60%. Classification of KNN with  $k = 1$  is very susceptible to noise this results in a low level of accuracy. With the above description, the value of  $k = 3$  is the most optimal  $k$  value.

### Conclusion

This study can be concluded that the proposed KNN classification method can classify EEG signals with an accuracy of 76% at the best neighbor value  $k = 3$ . The feature extracted for EEG signal classification uses the sampling technique.

### Acknowledgement

The author thanks the Chairman of the Muhammadiyah Siarjo University for providing time for research that the researchers are working on and the Directorate of Research and Community Service, Directorate General of Research Strengthening, and Development of the Republic of Indonesia Ministry of Research, Technology and Higher Education which supports funding for this research.

### Bibliography

- [1] H. D. Lux, "Neurophysiological Basis of the EEG and DC Potentials," *Electroencephalogr. Basic Princ. Clin. Appl. Relat. Fields*, pp. 24–28, 1981.
- [2] J. R. Wolpaw *et al.*, "Brain-computer interface technology: a review of the first international meeting.," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–73, Jun. 2000.
- [3] M. Planck *et al.*, "Support Vector Channel Selection in BCI Support Vector Channel Selection in BCI," vol. 51, no. 120, pp. 1003–1010, 2003.
- [4] J. D. R. Millán, "Adaptive brain interfaces for communication and control," *HCI Spec. Issue User Interfaces Age*, vol. 113, pp. 767–791, 2003.
- [5] T. A. H. Fatehi and A.-B. R. Suleiman, "Features Extraction Techniques of EEG Signals For BCI Application," *Naif Arab Univ. Secur. Sci.*, pp. 35–40, 2011.
- [6] Shakshi and R. Jaswal, "Brain Wave Classification and Feature Extraction of EEG Signal by using FFT on Lab View," *Int. Res. J. Eng. Technol.*, vol. 3, no. 7, pp. 1208–1212, 2016.
- [7] B. D. Mensh, J. Werfel, and H. S. Seung, "BCI Competition 2003--Data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals.," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1052–6, Jun. 2004.
- [8] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, 2000.
- [9] C. W. Anderson, E. A. Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalogram during mental tasks," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 3, pp. 277–286, 1998.
- [10] T. Wang, J. Deng, and B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns," *Clin. Neurophysiol.*, vol. 115, no. 12, pp. 2744–2753, 2004.
- [11] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *J. Neurosci. Methods*, vol. 123, no. 1, pp. 69–87, 2003.
- [12] Cochran, W. G.: *Sampling Techniques*. Wiley, New York, 1977.

- [13] Siulyand Li, Yanand Wen, Peng, "Classification of EEG signals using sampling techniques and least square support vector machines".In: 4th International Conference on Rough Set and Knowledge Technology (RSKT2009), 14-16 Jul 2009.
- [14] Khamis, H.S. Kipruto, W.C. dan Kimani, S. 2014. *Application of k-NearestNeighbor Classification in Medical Data Mining*. International Journal of Information and Communication Technology Research, Vol. 4, No, 4.
- [15] Suguna, N. dan Thanushkodi, K. 2010. *An Improved k-Nearest NeighborClassification Using Genetic Algorithm*. International Journal of ComputerScience Issues, Vol. 7, Issue 4, No 2.
- [16] Imandoust, S.B. dan Bolandraftar, M. 2013. *Application of K-Nearest Neighbor(KNN) Approach for Predicting Economic Events: Theoretical Background*. International Journal of Engineering Research and Applications, Vol. 3, Issue 5.
- [17] <http://bbci.de/competition/ii/>

# sinyal EEG

---

## ORIGINALITY REPORT

---

19%

SIMILARITY INDEX

15%

INTERNET SOURCES

16%

PUBLICATIONS

%

STUDENT PAPERS

---

## PRIMARY SOURCES

---

- 1** Wu Ting, Yan Guo-zheng, Yang Bang-hua, Sun Hong. "EEG feature extraction based on wavelet packet decomposition for brain computer interface", Measurement, 2008  
Publication **5%**
  - 2** H Hindarto, A Muntasa, S Sumarno. "Feature Extraction ElectroEncephaloGram (EEG) using wavelet transform for cursor movement", IOP Conference Series: Materials Science and Engineering, 2018  
Publication **3%**
  - 3** [eprints.undip.ac.id](http://eprints.undip.ac.id)  
Internet Source **3%**
  - 4** SungTae Lee, HyungWoo Park, Myung-Jin Bae, Kwanho Park. "P-112: Sound Quality Improvement of a Flat-Panel Display Front Speaker: Emotional and EEG Test of Exciter Speaker", SID Symposium Digest of Technical Papers, 2017  
Publication **1%**
-

5	Wei, Chih-Chiang. "Comparing lazy and eager learning models for water level forecasting in river-reservoir basins of inundation regions", <i>Environmental Modelling &amp; Software</i> , 2015.	1%
Publication		
6	<a href="http://www.laseeb.org">www.laseeb.org</a>	1%
Internet Source		
7	<a href="http://eprints.usq.edu.au">eprints.usq.edu.au</a>	1%
Internet Source		
8	<a href="http://pnrsolution.org">pnrsolution.org</a>	1%
Internet Source		
9	<a href="http://www.ijcic.org">www.ijcic.org</a>	1%
Internet Source		
10	Nguyen, Thanh, Abbas Khosravi, Douglas Creighton, and Saeid Nahavandi. "EEG data classification using wavelet features selected by Wilcoxon statistics", <i>Neural Computing and Applications</i> , 2015.	<1%
Publication		
11	<a href="http://www.mdpi.com">www.mdpi.com</a>	<1%
Internet Source		
12	Daniel Mareci, Daniel Sutiman, Romeu Chelariu, Florin Leon, Silvia Curteanu. "Evaluation of the corrosion resistance of new ZrTi alloys by experiment and simulation with	<1%

an adaptive instance-based regression model",  
Corrosion Science, 2013

Publication

---

13

Abbas Rohani, Mojtaba Mamarabadi. "Free alignment classification of dikarya fungi using some machine learning methods", Neural Computing and Applications, 2018

Publication

---

<1%

14

RENHUAN YANG. "FEATURE EXTRACTION OF MOTOR IMAGERY EEG BASED ON WAVELET TRANSFORM AND HIGHER-ORDER STATISTICS", International Journal of Wavelets Multiresolution and Information Processing, 2010

Publication

---

<1%

---

Exclude quotes      On

Exclude matches      < 10 words

Exclude bibliography      On