Classify Epilepsy and Normal Electroencephalogram (EEG) Signal Using Wavelet Transform and K-Nearest Neighbor

Dewi Rahmawati, Umy Chasanah N.R., Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
dewi16@mhs.if.its.ac.id, umy.rizqi13@mhs.if.its.ac.id, riyanarto@if.its.ac.id

***Abstract*—Epilepsy is a neurological disorder that cannot be predicted and studied. This study propose to classify epilepsy and normal Electroencephalogram (EEG) signal. Stages in the decision-making was done by using a feature extraction and combined with Wavelet Transform (WT). The result from features extraction was implemented dimension reduction method by using Principal Component Analysis (PCA) algorithm. K-Nearest Neighbor (KNN) was implemented using result from dimension reduction stages as features. In this work, 1000 data has been used as training data and 600 data has been used as a data testing. In this experiment, the dataset consist of two sets (A and E) from non-epileptic people and epileptic people. This experimental results also show that the sensitivity, accuracy and specificity of the results are 100%, 99.83% and 99.67%.**

Keywords—Electroencephalogram; KNN; PCA; Wavelet transform

I. Introduction

Epilepsy is a disease that can happen to anyone. Epilepsy is a disorder that occurs in the nervous system of the human brain. It is caused by the activity of neurons that too much until it make many reactions to the sufferer. Epilepsy can occur due to loss of electrical charge that excessive and unexpected in the brain so the reception and transmission of impulses from the brain to other parts of the body is disturbed [1]. Electroencephalogram (EEG) examination can be performed to diagnose a person who has an epileptic saizures because the instrument of EEG can be used for recording indicates electrical activity of the brain. EEG can provide knowledge of disorders of brain activity. Although the occurrence of epileptic seizures seems unexpected, more effort is being focused on developing computational models for automatic detection of epilepsy discharge, which can be used to predict the occurrence of seizures. K-Nearest Neighbor (KNN) is a classification method using the nearest-K neighbor of testing to its training data. This classification method can be used to identify the epileptic seizure, if given the proper training data.

This research was created a classification system, which combines Wavelet Transform (WT) and using KNN as a method of classification. Expected by combining the method can produce a reliable classification system to distinguish normal or epileptic seizures using EEG data. In previous work, EEG signals epilepsy and normal are classification in experiment [2] by using wavelet transform method, classification with euclidean distance composed feature selection on neural network method that called weighted fuzzy membership function and phase space reconstruction. In that study, feature selection is used as reference. The features such as mean, median, variance, standard deviation and mode has been used as the features for the classification [3, 4].

EEG can used to determine mood of a person. To determine mood of a person, KNN classification need to be applied along with wavelet and ICA [5]. This paper has a similar method, but there are differences. The different is this experiment using EEG to determine epilepsy disease of a person, while in previous work used to determine mood. Another difference were this paper used Principal Component Analysis (PCA) to reduce complexity of the feature while in previous work used ICA.

This paper will be divided into several parts, Section 2 describes the details of the EEG dataset that will be used in this study. The proposed methodology, including feature extraction and classification, will be discussed in section 3. And the result disscussion are shown in section 4. Finally, conclusions and development of research will be given in Section 5.

II. Datasets

This study uses the dataset EEG signals from Klinik für Epileptologie, Universität Bonn to classify the normal and epilepsy signal. The dataset consist of 5 sets that one set contains 100 of EEG signals that one signal has a 4097 features. Set A and B is recording from 5 healthy people with eyes closed and open. Set C, D and E are recording from epilepsy patients. In this experiment, used A and E set to classify epilepsy and normal EEG signal. Example of normal EEG signal on the set of A can be seen in Fig. 1, while example of epilepsy signal on the set E can be seen in Fig. 2. In this experiment, 4097 features of EEG signal has been divided into eight sets with 512 features each set, Thus, obtained 800 sub-signals with 512 features are made from 100 signals. Then 500 sub-signals are used as training data, and 300 sub-signals are used as test data, such as in [2].

978-1-5090-5864-8/17/$31.00 ©2017 IEEE

III. METHODOLOGY

Classification of normal EEG signals and epilepsy have several stages. The diagram of proposed method in this experiment can be seen in Fig. 3.

The method used in this study, because it takes the features extraction to obtain a characteristic signal so that the epilepsy and normal signal have significant differences, so the classification can be done accurately, wavelet done for denoising signals then the results of wavelet used for feature extraction to represent distribution frequency of wavelet coefficients of EEG signal. PCA used for reduction signal from features that have been obtained. Classification method used is KNN, because KNN is a method to perform the classification of objects based on the training data that were located closest to the object.

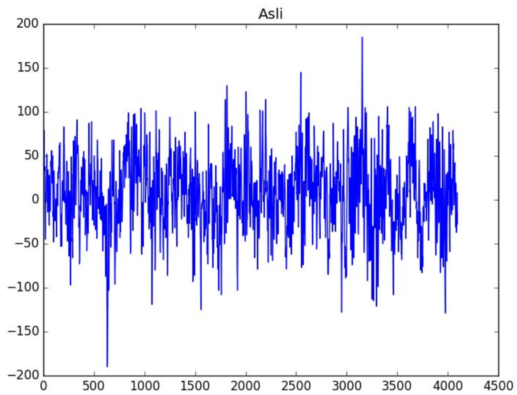


Fig. 1. Normal EEG signal.

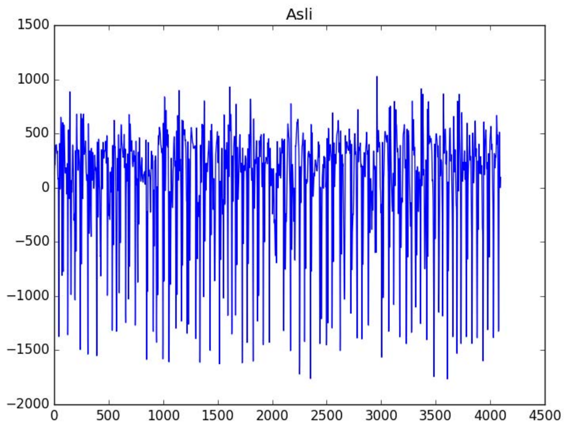


Fig. 2. Epilepsy EEG signal.

A. Feature Extraction

The process of feature extraction was using WT. The method used to compress the signal by preserve the original signal. The purpose is to reduce noise in signal and amplify information to maintain the integrity of the data information in signal [6]. The result of wavelet signal will have a length shorter than the original. WT decipher the signal into sub-band by using low-pass filtering and high-pass filtering of the time domain signal. Low-pass filtering produces approximation and high-pass filtering to produce detailed coefficients, where at the next level of decomposition was done by using the approximation at the previous level. This experiment the used type of wavelet is Daubechies Wavelet order 4 (DB4). The original of signal can be decomposed to a coefficients set that describe content of frequency [7].

Where P_{2i} is smoothing operator, then P_{2i} is digital signal $x(n)$, $i \in Z$ is the set of integral, h_k and g_k are coefficients for corresponding high-pass and low-pass filter. Wavelet coefficients of normal and epilepsy EEG signal shown in Fig. 4.

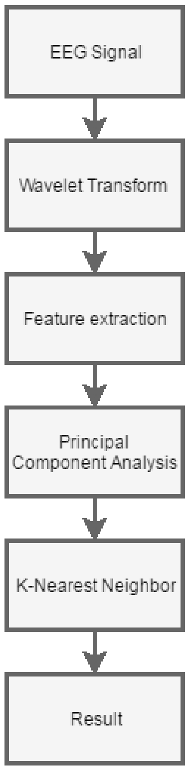


Fig. 3. Diagram of EEG signal classification.

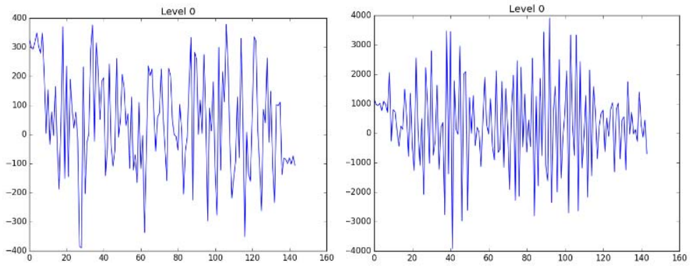
In Fig. 4a, is shows the low-pass filtering results of Wavelet Transform which produces approximation and in Fig. 4f, shows the high-pass filtering results as detail coefficient. Fig. 4, can be seen the wavelet method is highly dependent on the level of decompositionthat will be used. More higher level of decomposition is used, more higher the risk of the signal can be defect. Conversely, more lower the level of decomposition is used, then the signal will be more susceptible to noise.

Having obtained 5 level of signal decomposition and approximation. Furthermore, the results of wavelet used for feature extraction to represent distribution frequency of wavelet coefficients of EEG signal. The features extracted from EEG signal decomposition on the set of E and A differentiated from one another. Therefore, it was concluded, the wavelet coefficient can be presented as a useful parameter in classifying the epilepsy and normal EEG signals.. In this experiment features was extracted with Discrete Wavelet Transform (DWT) as the feature extractor are:

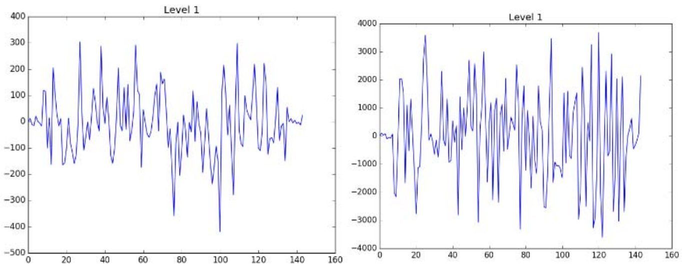
- The highest value wavelet coefficient at sub-signal.

- The lowest value wavelet coefficient at sub-signal.
- The average value wavelet coefficient at sub-signal.
- The standard deviation value wavelet coefficient at sub-signal.

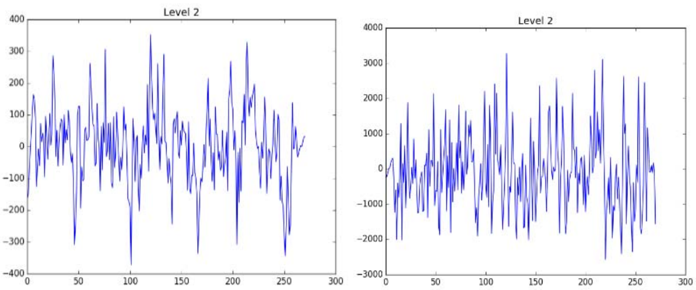
After extracting feature for every level of decomposition (including approximation) of the pieces of the signal, it would get 24 features and was used for classification.



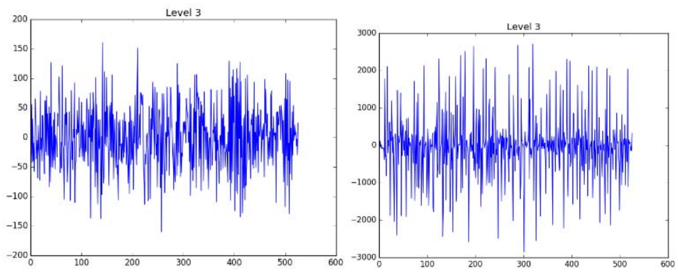
(a) Wavelet coefficient approximation normal and epilepsy.



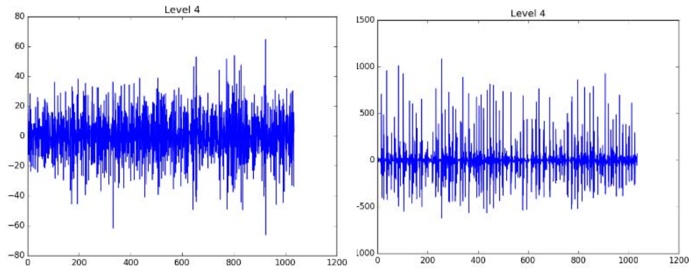
(b) Wavelet coefficient d1 normal and epilepsy.



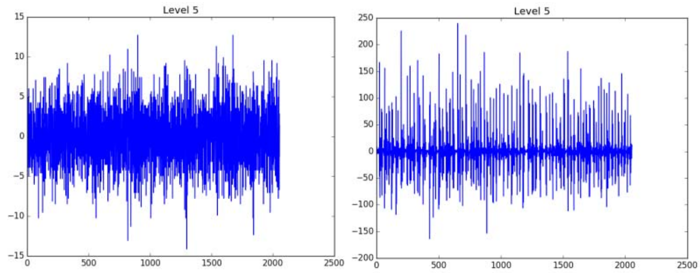
(c) Wavelet coefficient d2 normal and epilepsy.



(d) Wavelet coefficient d3 normal and epilepsy.



(e) Wavelet coefficient d4 normal and epilepsy.



(f) Wavelet coefficient d5 normal and epilepsy.

Fig. 4. Epilepsy EEG signal.

B. *Reduction Dimension*

Before doing the classification process conducted the data extraction from the data reduction features that will be the input of the classification process. PCA used for dimensional reduction for multichannel data and feature extraction. PCA modify the original signal into new principal components (PCs) [8]. By using PCA, the data extraction features that have 24 features can be reduced to d smaller dimensions.

PCA method is used for feature extraction and dimension reduction. In PCA is used to represent the data dimension d in the lower-dimensional space. This will reduce the complexity of space and time. Purpose of the dimension reduction is to represent data in the best dimensions. This technique is mostly useful for segmentation of signals from multiple sources.

Steps to conduct PCA are :

1. Input data
2. Data standardization
3. Compute covariance matrix
4. Compute eigenvalues and eigenvectors of covariance matrix
5. Selecting components and forming feature vectors
6. Getting new dataset.

C. *Classification*

The next stage after the EEG signal performs feature extraction and dimensional reduction that is classified by K-Nearest Neighbor classification method. The KNN data sample is based on distance measurement with the number of patterns stored in memory. Testing data was calculated to each training data using distance method the smallest distance of training data to its testing was assumed the correct class. In research [9] distance function used was cosine similarity because the similarity distance computed are documents. Cosine similarity is one of the many functions used in document classification to find commonalities between multiple documents. While in this distance experiment, the function used in the KNN classification is Euclidean distance. In addition to distance, the nearest sample data sample also can retrieve data from the majority of KNN. Parameter *k* must be chosen in practice.[10]

In KNN, the input data has calculated the distance to the entire training data. After it was taken a number of training data *k* with the shortest distance. The test data was categorized in accordance with the highest data group of a number of data The nearest *k* [11]. Distance measurements used in this study is the Euclidean distance.

Equation (1) are known vectors of two lines $x_i = x_{i1}, x_{i2}, \ldots, x_{in}$ and $x_j = x_{j1}, x_{j2}, \ldots, x_{jn}$ then Euclidean distance of both is:

$$d(k,l) = \sqrt{(X_{k_1} - X_{l_1})^2 + (X_{k_2} - X_{l_2})^2 + \cdots + (X_{k_n} - X_{l_n})^2} \quad (1)$$

In the training phase, the algorithm is just doing the feature vector storage and classification of training sample. In the classification phase, the same features are calculated for testing data. The distance from the new vector of the entire sample training vectors are calculated and the amount of the closest *k* taken.

The way to calculate the KNN algorithm:

1. Determining *k* parameter (the number of nearest neighbors)
2. Calculate the distance Euclidean distance of each object on the data samples provided
3. Sort the objects into groups that have the smallest distance.

IV. Result and Discussion

A. Statistical test

In this study, TP (True Positive) was determined when the epilepsy signals classification as epilepsy signal. FP (False Positive) was determined when the normal signals are classification as epileptic signals. And then TN (True Negative) was determined when the normal signals are classified as an epilepsy signal. FN (False Negative) was determined when the epilepsy signals are classified as normal signal as illustrated in Table I [12]. Furthermore from the confusion matrix can be calculated sensitivity, specificity, and accuracy that defined in (2), (3) and (4) respectively. Sensitivity (True Positive Rate) is measures the proportion of positives that are correctly identified. Which specificity (True Negative Rate) is measures the proportion of negatives that are correctly identified. And accuracy is defined as the degree of closeness between the predicted values to the actual values.

TABLE I. CONFUSION MATRIX

actual		Predicted	
		Epilepsy	*Normal*
	Epilepsy	TP	FN
	Normal	FP	TN

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (2)$$

$$Specificity = \frac{TN}{TN+FN} \times 100 \quad (3)$$

$$Accuracy = \frac{TP+TN}{TN+FN+TP+FP} \times 100 \quad (4)$$

B. Result

In this experiment the normal and epilepsy signals was classified using KNN, given $k = 3$ and $k = 5$ performed using feature extraction WT and dimension reduction using PCA, given $n = 10$ and $n = 15$, Table II are compared to the propose method calculation with different value of *n* at PCA and *k* at KNN. Table III are presented the value from confusion matrix, so that it can be calculated sensitivity, specificity, and accuracy based on (2), (3), (4) respectively from the best result (k=3 and $n = 15$).

TABLE II. COMPARISON OF RESULT

K	N	Sensitivity (%)	Specificity (%)	Accuracy (%)
3	15	99.67	100	99.83
5	15	99.33	100	99.65
3	10	99.33	100	99.65
5	10	99.00	100	99.50

From the above experimental results, the best value of sensitivity 99,67%, specificity 100% and accuracy 99.83%

TABLE III. PROPOSED CLASSIFICATION RESULT

actual		Predicted	
		Epilepsy	*Normal*
	Epilepsy	299	1
	Normal	0	300

Table IV, shown the comparisons of performance results from EEG classification, it shown that feature selection influence on the process of classification.

TABLE IV. COMPARISONS OF PERFORMANCE RESULTS

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
Support Vector Machine (SVM) and Empirical Mode Decomposition (EMD) [13]	98.00	99.40	98.7
Neural Network with Weighted Fuzzy Membership (NEWFM) function with feature selection [2]	96.33	100	98.17
In this work	99.67	100	99.83

V. Conclusion

This study uses a WT as a preprocessing step to extract the useful features of the EEG signal. First, the coefficients of wavelet are made from EEG signals that using WT. Finally, the 24 features extracted from wavelet transform. Furthermore, the signal with 24 of these features conducted the reduction of features using PCA in order to get 15 features. From the EEG signals with 15 of these features conducted using KNN classification with a value of k = 5. The classification method produces a value of sensitivity 99.67%, specificity 100%, and accuracy 99.83%. The advantages of this method lies in feature extraction because it has been observed that, the features extracted from EEG signal decomposition on the set of A and E are different from each other. Therefore, this experiment concluded that the coefficients of wavelet can be presented as useful parameter in classifying epilepsy and normal EEG signal.

References

[1] "What is epilepsy? | Epilepsy Action". [Online]. Available: https://www.epilepsy.org.uk/info/what-is-epilepsy. [Accessed: 19-Mar-2017].

[2] S.-H. Lee, J.S. Lim, J.-K. Kim, J. Yang, and Y. Lee, "Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance", *Comput. Methods Programs Biomed.*, vol. 116, no. 1, pp. 10–25, 2014. https://doi.org/10.1016/j.cmpb.2014.04.012

[3] R. Sarno, M.N. Munawar, B.T. Nugraha, "Real-Time Electroencephalography-Based Emotion Recognition System", *Int. Rev. Comput. Softw. IRECOS*, vol. 11, no. 5, pp. 456–465, May 2016. https://doi.org/10.15866/irecos.v11i5.9334

[4] B.T. Nugraha, R. Sarno, D.A. Asfani, T. Igasaki, and M.N. Munawar, "Classification of driver fatigue state based on EEG using Emotiv EPOC+", *J. Theor. Appl. Inf. Technol.*, vol. 86, no. 3, p. 347, 2016.

[5] M.N. Munawar, R. Sarno, D.A. Asfani, T. Igasaki, and B.T. Nugraha, "Significant preprocessing method in EEG-Based emotions classification", Journal of Theoretical and Applied Information Technology (JATIT*)*, vol. 87, no. 2, p. 176, 2016.

[6] R. Sarno, B.T. Nugraha, M.N. Munawar, "Real Time Fatigue-Driver Detection from Electroencephalography Using Emotiv EPOC+", *Int. Rev. Comput. Softw. IRECOS*, vol. 11, no. 3, pp. 214–223, Mar. 2016. https://doi.org/10.15866/irecos.v11i3.8562

[7] D.R. Wijaya, R. Sarno, and E. Zulaika, "Information Quality Ratio as a novel metric for mother wavelet selection", *Chemom. Intell. Lab. Syst.*, vol. 160, pp. 59–71, Jan. 2017. http://dx.doi.org/10.1016/j.chemolab.2016.11.012

[8] R. Patel, S. Sengottuvel, M.P. Janawadkar, K. Gireesan, T.S. Radhakrishnan, and N. Mariyappa, "Ocular artifact suppression from EEG using ensemble empirical mode decomposition with principal component analysis", *Comput. Electr. Eng.*, vol. 54, pp. 78–86, Aug. 2016. https://doi.org/10.1016/j.compeleceng.2015.08.019

[9] B.Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM", in *2015 International Conference on Data and Software Engineering (ICoDSE)*, 2015, pp. 170–174. https://doi.org/10.1109/icodse.2015.7436992

[10] L. Guo, D. Rivero, J. Dorado, C.R. Munteanu, and A. Pazos, "Automatic feature extraction using genetic programming: An application to epileptic EEG classification", *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10425–10436, Aug. 2011. https://doi.org/10.1016/j.eswa.2011.02.118

[11] T. Wissel and R. Palaniappan, "Considerations on Strategies to Improve EOG Signal Analysis," International Journal of Artificial Life Research, vol. 2, no. 3, pp. 6–21, 2011. https://doi.org/10.4018/jalr.2011070102

[12] T. Fawcett, "An introduction to ROC analysis", *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006. https://doi.org/10.1016/j.patrec.2005.10.010

[13] S. Li, W. Zhou, Q. Yuan, S. Geng, and D. Cai, "Feature extraction and recognition of ictal EEG using EMD and SVM", *Comput. Biol. Med.*, vol. 43, no. 7, pp. 807–816, Aug. 2013. https://doi.org/10.1016/j.compbiomed.2013.04.002