Computer Engineering and Applications Vol. 8, No. 3, October 2019

Accuracy Improvement of Incidence Level Detection Based on Electroencephalogram Using Fuzzy C-Means and Support Vector Machine

M. Ramdhany*, S. Wahjuni, S. N. Neyman

Department of Computer Science, Faculty of Math and Science, Institut Pertanian Bogor
**ramdhanymarwan@gmail.com*

ABSTRACT

Some jobs require that the concentration level be maintained for a longtime during work time. Lack of sleep will in disruption of someone concentration level. To find out a human concentration level can be done by recording his/her brain waves. This research uses Electroencephalography (EEG) technology which functions to capture human brain waves. The focus of this study is to build a model of the detection system of a human concentration level. The research datasets are data from brain wave recording using Neurosky Mindwave Mobile which has extracted in 19 features. Data will then be labeled using cluster techniques namely Fuzzy C-Means to become data to be input into the classification process using Support Vector Machine (SVM). The classification results show an accuracy of 98.34%. That results show FCM can be used to automatically label EEG data properly.

Keywords: ANN, EEG, Clustering, FCM, SVM

1. INTRODUCTION

Some jobs require concentration level be maintained for a longtime during work time. Work that results in less sleep will result in disruption of the human concentration level. Disorders of the sleep cycle can result in a decrease in alertness, performance, and fatigue levels [1]. Symptoms of drowsiness, namely the transition between conscious and sleep, where the condition makes all the senses experience a decline in function resulting in decreased concentration while working or not working. This condition is a factor that contributes to accidents that occur on the road [2]. Because the decrease in concentration is one of the causes of many accidents, research is needed for early detection of the level of alertness in order to minimize accidents that occur. Based on BPS data in the last 5 years, there have been consistent figures above 95,000 accidents. In 2017 there were 103,228 total accidents. Therefore, it is necessary to do research related to the detection of the level of alertness to minimize the risk of the accident. According to [3], driving while drowsy can lead to a number of problems, namely, longer reaction times, reduced alertness, and a reduction in information processors which results in a decision-making error. Furthermore, [3] suggest that there are two categories of methods that are widely used to detect drowsiness. The first category is using a digital image process that uses optical sensors or video cameras to detect eye activity from the conscious to fall asleep by measuring the time of the lid closure, eye movement, and movement of the head. First category has been carried out [4] who made a sleep detection system using an algorithm that specifically detects the condition of the driver eyelids and will raise a warning when drowsy. That research has obstacles, namely increasing adaptability to changes in brightness, and guaranteeing technological change and lifestyle. Second category is focusing on the physiological measurements of driver, such as heart-rate variability (HRV), electrooculographic (EOG), and electroencephalogram (EEG). The approach with the second category has advantages in making accurate assessments of levels of alertness and faster processing methods.

EEG signals themselves can be recorded with medical devices or more simply like Neurosky Mindwave Mobile. Neurosky Mindwave Mobile has been used in several studies, especially in the field of brain-computer interfaces (BCI). [5] conducted research on BCI using morse password from EEG signals obtained from Neurosky Mindwave Mobile. This morse password uses data obtained from eye movements. The average accuracy obtained is 96.3% using k-nearest neighbor and dynamic time wrapping. Another study of BCI was carried out [6] who developed a real-time BCI system to recognize the direction of right and left movements based on what the subject imagined in his mind. The multiscale PCA (MSPCA) method used to clean EEG signal data is then classified using the classification method, Rotation Forest (RoF). The results of the analysis of the study offline system showed that the use of MSPCA and rotation forest classification method produced the highest average accuracy of 89.6%. Some of these studies show that consumer level sensors such as Neurosky Mindwave Mobile can be used for research scales.

Another classification algorithm used to determine EEG accuracy is Support Vector Machine. Research [7], using SVM classification on EEG data obtained from Physionet. The study aims to develop software for the classification of EEG data to detect sleep and awake conditions through EEG signals using the autoregressive (AR) model as a feature extraction method using SVM as a classifier. The study used three sub-bands, namely theta, alpha, and beta. Test data in the form of numbers 0 for sleep and 1 for awake. The method used to estimate AR coefficients is the Yule-Walker method and the Burg method. The results of the study were, the average accuracy, precision, and sensitivity of the Yule-Walker method were 80.60%, 78.19%, and 77.59%, respectively, while the foreign Burg method was 94.01 %, 95.70%, and 93.39%. [8] uses the Hierarchical Clustering method and classification using Artificial Neural Network (ANN) which gets an accuracy of 75.37%. Other related research on EEG as an early warning system was conducted [9] which used Artificial Neural Network (ANN) as a classifier which was initialized into 3 classes namely sleeping, waiting and standby to get an accuracy of 77.03%. According to [10], there are two clustering methods namely partitioning and hierarchical clustering. One algorithm from partitioning is Fuzzy C-Means. This algorithm was first introduced by Jim Bezdek in 1981. The selection of FCM is based on the characteristics of an uncertain level of alertness. The level of vigilance when asked directly to consumers produces a subjective answer. Then the cluster technique is used on data that does not yet have a label. Research related to EEG using FCM was carried out [11] use FCM for EEG recognition in single-trial motor imagery (MI). According to [11] approach using FCM is an adaptive approach that is suitable for clustering non-stationary biomedical signals.

This research continues the research of [8] by using FCM for data clustering and SVM as a classifier. Data clustering processes that do not yet have labels use the Fuzzy C-Means (FCM) method to produce a level of alertness class which will be

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

used as a comparison output using the Support Vector Machine (SVM) classification method.

2. RESEARCH METHOD

This research is conducted with the following stages :

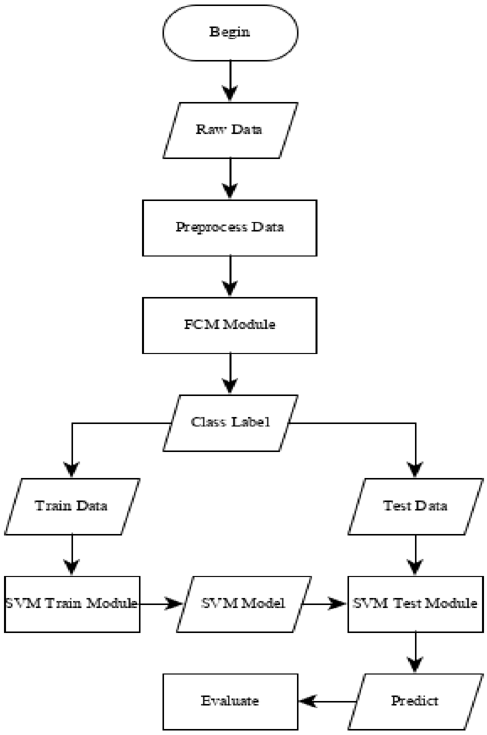

FIGURE 1. Research stages

2.1 RAW DATA

The data used is data used from the research of [8]. The data is in the form of EEG data from 12 people. Conditioning of the alert state of the subject represents standby, sleepiness, and sleep. Each conditioning is carried out twice recording with a duration of five minutes each. The actual class which is the subject of alertness of the subject is ignored in advance so that data that does not yet have a label obtained from the Neurosky Mindwave Mobile tool that has passed the feature extraction process uses Daubechies type 2 Discrete Wavelet Transform (DWT) with 4 levels of decomposition. After using the wavelet, 19 features and 5760 data will be obtained.

The data has 19 features extracted namely meanA4, meanD4, meanD3, meanD2, meanD1, stdA4, stdD4, stdD3, stdD2, stdD1, rmsA4, rmsD4, rmsD3, rmsD2, rmsD1, ratioA4D4, ratioD4D3, ratioD3D2 and ratioD2D1 and ratioD2D1.

2.2 PREPROCESSING

Preprocessing data uses a standardization process so that the data to be used is in a certain small distance. Data will use standard standardization with Equation (1).

$$x_{\text{new}}=\frac{x\text{-min}}{\text{max-min}} \tag{1}$$

where x_{new} is the normalized data value, x is the old data value, min is the lowest data of a feature, and max is the highest data of a feature. Each value from each row of data will be normalized so that it gets a range of values between 0 and 1. Data that has gone through preprocessing will produce different data from the previous data, the preprocessing purpose itself so that the data to be used can be clean and the distance between the data is not far apart.

2.3 FCM MODULE

According to [12], several parameters are needed for the FCM algorithm according to Table 1. These parameters are the initial parameters to get the cluster center.

TABLE 1.
FCM Parameters

Parameter(s)	Value
Data	Data n x m (5760x19)
Number of cluster (C)	3
Fuzzifier (w)	2
Stop criteria (error)	0.005
Number of Iterations (MaxIter)	1000

The first iteration of FCM must first determine the μ_{ik} matrix randomly with $\sum_{k=1}^{c}\mu_{ik}=1$ meaning that the number of membership values for a data cluster must equal 1. After getting the initial membership value matrix, then will find the cluster center with Equation (2).

$$V_{kj}=\frac{\sum_{i=1}^{n}\left(\mu_{ik}\right)^{w} * X_{ij}}{\sum_{i=1}^{n}\left(\left(\mu_{ik}\right)^{w}\right.} \tag{2}$$

μ_{ik} is the random membership value that has been given before, and X is the data $n\ x\ m$. V_{kj} as the initial cluster center will be a reference to get the objective function of each iteration with Equation (3).

$$P_t=\sum_{i=1}^{n}\sum_{k=1}^{C}\left(\left[\sum_{j=1}^{m}\left(X_{ij}-V_{kj}\right)^2\right]\left(\mu_{ik}\right)^{w}\right) \tag{3}$$

The next step is to improve the value of membership value μ_{ik} with equation (4).

 ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

$$\mu_{ik}=\frac{\left[\sum_{j=1}^{m}\left(X_{ij}-X_{kj}\right)^{2}\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(X_{ij}-X_{kj}\right)^{2}\right]^{\frac{-1}{w-1}}} \tag{4}$$

Iteration 1 will get a new μ_{ik} value which will then be repeated in the next iteration. Criteria for terminating iterations are two conditions, If (P_t-$P_{t-1}|<error$) or (t> MaxIter) then stop, and if not, then t = t + 1, repeating the cluster center count (Vkj).

After the stop criteria are reached, the membership value will be obtained for each data consisting of 3 clusters with each weight.

2.4 SVM TRAINING AND TESTING MODULE

Training module is the stage of making a model for classification. SVM classification used in this study applies three kernels, namely linear, polynomial, and radial base function (RBF).

TABLE 2.
SVM Kernels

Kernel	Function
Linear	$\langle x, x' \rangle$
Polynomial	$(\gamma\langle x, x'\rangle+r)^d$
RBF	$exp\ (-\gamma\|x-x'\|^2)$

Output of the model will be saved as a prediction class which will then be evaluated. The evaluation mechanism of the data obtained from the system testing phase will be analyzed using confusion matrix. Accuracy values can be calculated using Equation 5.

$$\text{Accuracy}=\frac{\sum \text{correct classification test data}}{\sum \text{test data}}\times 100\% \tag{5}$$

3. RESULT AND ANALYSIS

3.1 PREPROCESSING

Based on preprocessing data using Equation (1). Obtained as in Table 3. The data represents 19 features that will be used in the clustering process.

TABLE 3.
Preprocessing result using Equation (1)

No	meanA4	meanD4	meanD3	meanD2	meanD1	..	ratioA4D4	ratioD3D2
0	0.019	0.117	0.201	0.109	0.080	..	0.066	0.495
1	0.018	0.057	0.097	0.066	0.055	..	0.152	0.410
2	0.048	0.064	0.102	0.061	0.052	..	0.193	0.460
3	0.013	0.078	0.112	0.068	0.056	..	0.102	0.456
:	:	:	:	:	:	-		
5760	0.020	0.014	0.012	0.001	0.001	..	0.358	0.726

3.2 FCM MODULE

The features previously obtained are then clustered using Equation (2). Fuzzy C-Means Clustering (FCM) algorithm found in the python sklearn library. The data will be divided into 3 classes (clusters) according to the representative class according to [8]. Parameters used are parameters from [13] and [8] to obtain clustering results using FCM such as Table 4. Data that has gone through the clustering process will have a cluster label which is a representation of each class of data, namely 0 to sleep, 1 to sleepy and 2 to standby.

TABLE 4.
FCM Membership Degree

Data n-	Membership degree			Cluster
	0	1	2	
1	0.171490764	0.1003457	0.631461941	0
1000	0.016056691	0.942833176	0.041110133	1
2000	0.094391696	0.15827626	0.747332044	2
3000	0.725562121	0.076932986	0.197504894	1
4000	0.308020156	0.083429066	0.608550778	2
5760	0.045612851	0.839508155	0.114878994	1

3.3 DATA SELECTION

Training data used is 80% of the total 5760 data or 4608 data. While the test data is used as much as 20%, namely 1152 data. Distribution of training data and test data using random sampling of the total data.

3.4 SVM TRAINING AND TESTING MODULE

Previously selected training data will enter the learning process in SVM with the sklearn library before being tested with test data. The parameters to be used are parameters from [13] based on Table 5.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

TABLE 5.
SVM Parameters

Parameters	Value
Kernels	Linear, RBF.polynomial, sigmoid
Degree	3
Gamma	0.05
Number of iterations	unlimited

The results of the identification based on the chosen model resulted in the distribution of data in the form of confusion matrices as in Table 6. Calculation of accuracy was obtained from each data model correctly identified.

TABLE 6.
Confusion matrix result using FCM and SVM

Predicted cluster	Real cluster			Total
	0	1	2	
0	157	1	7	165
1	0	465	9	474
2	0	2	511	513
Total	157	468	527	1152

$$\text{Accuracy}=\frac{157+465+511}{1152} * 100\%=98.35\%$$

TABLE 7.
Result others SVM kernel

Kernel	Predicted result (%)
Linear	98.45
Polynomial	93.05
RBF	94.7

Prediction results from the SVM method using linear kernels are the highest compared to the rbf, polynomial, and sigmoid kernels which reach 0.9835 or 98.35%.

4. CONCLUSION

Clustering problems in previous research can be overcome by using a fuzzy clustering algorithm, Fuzzy C-Means Clustering (FCM). FCM can improve prediction accuracy. Using the ANN classifier gets an accuracy of 97.91% where the previous research received 75.37%. SVM as a comparison classifier gets an accuracy of 98.35%. This shows that FCM can group EEG data properly.

REFERENCES

[1] A. Jones, G. Schwartz, and H. S. Choi, "Using brain-computer interfaces to analyze EEG data for safety improvement," *Team for Research in Ubiquitous Secure Technology,* 2010.

[2] R. R. Knipling and J. S. Wang, *Crashes and fatalities related to driver drowsiness/fatigue*. National Highway Traffic Safety Administration Washington, DC, 1994.

[3] C. T. Lin, R. C. Wu, S. F. Liang, W. H. Chao, Y. J. Chen, and T. P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Transactions on Circuits Systems I: Regular Papers,* vol. 52, no. 12, pp. 2726-2738, 2005.

[4] H. Ueno, M. Kaneda, and M. Tsukino, "Development of drowsiness detection system," in *Vehicle Navigation and Information Systems Conference,* 1994, pp. 15-20: IEEE.

[5] A. H. Brilian, H. Tjandrasa, and C. Fatichah, "Pengenalan sandi morse dari sinyal electroencephalogram yang direkam perangkat *neurosky mindwave* menggunakan *dynamic time warping*," *JUTI,* vol. 14, no. 1, pp. 63-71, 2016.

[6] J. Kevric and A. Subasi, "The impact of Mspca signal de-noising in real-time wireless brain computer interface system," *Southeast Europe Journal of Soft Computing*, vol. 4, no. 1, 2015.

[7] Y. H. Mahendra, H. Tjandrasa, and C. Fatichah, "Klasifikasi data eeg untuk mendeteksi keadaan tidur dan bangun menggunakan autoregressive model dan support vector machine," *JUTI: Jurnal Ilmiah Teknologi Informasi,* vol. 15, no. 1, pp. 35-42, 2017.

[8] A. Wicaksono, "Deteksi Tingkat Kewaspadaan Berbasis *Electroencephalogram* Menggunakan *Hierarchical Clustering* dan Jaringan Syaraf Tiruan," Computer Science, IPB University, 2017.

[9] S. Wahjuni and M. H. Perdana, "The Development of Eeg Based Alertness Level Detection System Using Artificial Neural Network," *International Journal of Tomography and Simulation*, vol. 32, no. 3, pp. 23-36, 2019.

[10] R. Suganya and R. Shanthi, "Fuzzy c-means algorithm-a review," *International Journal of Scientific Research Publications*, vol. 2, no. 11, p. 1, 2012.

[11] W.-Y. Hsu *et al.*, "Unsupervised fuzzy c-means clustering for motor imagery EEG recognition," *Int J Innovative Comput Inform Control,* vol. 7, pp. 4965-4976, 2011.

[12] S. Kusumadewi and H. Purnomo, *Aplikasi logika fuzzy untuk pendukung keputusan*. Yogyakarta: Graha Ilmu, 2004.

[13] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of machine learning research*, pp. 2825-2830, 2011.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)