

Principal component analysis implementation for brainwave signal reduction based on cognitive activity

Ahmad Azhari^{1,*}, Murein Miksa Mardhia²

Informatics Department, Universitas Ahmad Dahlan, Indonesia

¹ *ahmad.azhari@tif.uad.ac.id*; ² *murein.miksa@tif.uad.ac.id*

** corresponding author*

ARTICLE INFO

Article history:

Received October 19, 2017

Revised November 22, 2017

Accepted November 22, 2017

Keywords:

Brainwaves

EEG

Cognitive Activity

PCA

Pattern Recognition

ABSTRACT

Human has the ability to think that comes from the brain. Electrical signals generated by brain and represented in wave form. To record and measure the activity of brainwaves in the form of electrical potential required electroencephalogram (EEG). In this study a cognitive task is applied to trigger a specific human brain response arising from the cognitive aspect. Stimulation is given by using nine types of cognitive tasks including breath, color, face, finger, math, object, password thinking, singing, and sports. Principal component analysis (PCA) is implemented as a first step to reduce data and to get the main component of feature extraction results obtained from EEG acquisition. The results show that PCA succeeded reducing 108 existing datasets to two prominent factors with a cumulative rate of 65.7%. Factor 1 (F1) includes mean, standard deviation, and entropy, while factor 2 (F2) includes skewness and kurtosis.

Copyright © 2017 International Journal of Advances in Intelligent Informatics.

All rights reserved.

I. Introduction

Human has the ability to think which comes from the brain. The brain also plays an important role for the human body, everything related to human actions and activities centered from the brain. According to Katona *et. al.* [1], brain can generate electricity. In the human brain, whether in a state of awake or asleep, different frequency changes can be observed in the spectrum of measured electrical signals from the brain. Brain activity that produces electrical impulses can be measured with an EEG (electroencephalograph) device. Based on the frequency of signals, brainwaves can be divided into 5 types of waves, such as delta, theta, alpha, beta, and gamma waves [2][3].

The development of EEG in the biomedical field can be seen from the use of EEG signals in the application of the brain computer interface (BCI) system. BCI Technology is a field of research that develops applications related to control functions to medical diagnostics. BCI is also called neural interface or brain machine interface (BMI), which is a direct communication channel between the brain and a separate device [2]. This study applies BCI technology in the use of sensors that measure and interpret brain activity as a source of input.

Brainwaves can be used as a medium for user identification and authentication [4]. The use of brain waves has several advantages over other biometric authentication systems such as fingerprint and iris scan. Brainwave signals have characteristics and characteristics typical of the individual so it is impossible to have similarities, imitated or read by others [5].

To be able to distinguish the characteristics of each individual and recognize the person's identity appropriately is required recognition of identity. The use of technology that uses biometrics, body parts or human behavior, has been widely developed [4][6]–[17]. The use of brainwave signals is one example of identity recognition by using biometric technology based on body parts.

Characteristics of brainwaves become very strong when a person is exposed to certain stimuli. The use of cognitive tasks is aimed at triggering the specific response of the human brain arising from the cognitive aspect. This development is also referred as cognitive biometrics [5].

Johnson *et. al.* [18] conducted research related to user authentication of computer systems based on brain wave signals using electroencephalogram (EEG). To get the original signal then do some mental tasks. The study involved 15 students with 7 mental tasks performed. The purpose of this study was to measure the robustness of authentication systems of brainwave signals to direct imitation attacks. The data analysis used is divided into 3 stages of data compression, signal differentiation, and subject authentication, as well as identity imitation identity attacks. The results obtained are False rejection rate (FRR) which shows the false and false acceptance rate (FAR) indicates an average of 2%, which means that each subject is authenticated properly authentic on each subject and relatively few experimental subjects can authenticate like any other subject. Half total error rate (HTER) also shows 1% result.

Chai *et. al.* [19] conducted research on brain waves based on mental fatigue. Data were obtained from 65 healthy participants. Mental fatigue is calculated from the large number of eyes open and eyes closed. The methods used include data reduction with principal component analysis (PCA), feature extractor with power spectral density (PSD), and data classification with Bayesian neural network (BNN). The result of data reduction with PCA obtained 6 major components of a total of 26 EEG channel data with a cumulative value of 90%. Classification obtained at open eyes using BNN resulted in sensitivity of 76.8%, specificity of 75.1%, and accuracy of 76%. While the classification obtained at the time of closed eyes using BNN resulted in sensitivity of 76.1%, specificity of 74.5%, and accuracy of 75.3%.

To be able to recognize the pattern of brain activity it is necessary to recognize the pattern of EEG signals. In this study will be implemented the principal component analysis as a first step to reduce the data from the feature extraction results obtained from EEG acquisition.

II. Method

A. Existing Brainwave Data

Our research extends previous experimental work [5]. Data acquisition process is obtained using Neurosky Mindset using a EEG sensor (so-called electrode), placed in FP1 (frontal lobe) position based on the 10-20 system. Electrode placement is based on a 10-20 system [3]. EEG data acquisition was obtained from 6 healthy subjects consisting of 3 men and 3 women. The retrieval is done twice in different time with a sampling frequency of 128 Hz per second. The retrieval process is done for 20 seconds.

B. Cognitive Task

Cognitive activity of the brain is based on several studies related to psychological perceptions [3]–[5][11][18]. This cognitive activity aims to gain a specific response from the brain's cognitive activity (so-called cognitive task). There are nine cognitive tasks involved including breath, color, face, finger, math, objects, password thinking, singing, and sports. The nine types of cognitive task are described below, each task in terms of its instructions for the subjects.

1. Breathing Task (Breath). In this task is focused on breathing. This task is done for 20 seconds while closing the eyes. Subjects are not allowed to perform any movements.
2. Object Counting Color Task (Color). This task is given to remember colors. Subjects are shown some colors in a particular order to remember then subjects are asked to point the colors sequentially according to their memories. This task is done for 20 seconds silently.
3. Simulated Movement Finger (Finger). This task is a simulated task focused on the finger. Subjects are asked to imagine moving your finger without actually moving your finger for 20 seconds while closing your eyes.
4. Simulated Facial Reconstruction (Face). This task focuses on the face of a person known by the subject. Subjects were asked to close their eyes and reconstruct faces for 20 seconds in silent conditions without movement and sound.
5. Simulated Object Reconstruction (Object). This task focuses on the reconstruction of objects in detail. Subjects are given 20 seconds to reconstruct objects while closing their eyes without making a movement and making a sound.

6. **Mathematical Task (Math).** This task serves to perform simple mathematical calculations. Calculations include addition, subtraction, multiplication, and division. Subjects are given 20 seconds to answer questions without making a sound. The false and true answer is ignored in this task.
7. **Simulated Password Recall Task (Pass-thought).** This task focuses on password memories in the form of sentences consisting of a combination of letters and numbers. Subject given a line as a password that must be remembered then the subject is asked to close the eyes and repeat the password without making a sound. This task is done for 10 seconds.
8. **Song Recitation Task (Song).** This task focuses on repeating the lyrics of the song. Subjects were asked to repeat the lyrics of the preferred song for 20 seconds without movement and sound.
9. **Simulated Sport Task (Sport).** This task focuses on the preferred sports movement. Subjects were asked to perform a preferred exercise movement in silence and close their eyes for 20 seconds.

C. Feature Extraction

The feature extraction is done by involving statistical features including mean, skewness, standard deviation, kurtosis, and entropy. EEG data obtained after feature extraction are grouped by three categories including cognitive task, time data collection, and subject. The distribution of data is represented by the mean, the data distribution variation is represented by the standard deviation, the asymmetric data distribution rate is represented by skewness, the high-low distribution of data to the normal distribution is represented by kurtosis, and the randomness measure of the distribution data is represented by entropy. The five statistical features can be calculated using equation (1) to (5).

$$mean = \bar{x} = \frac{1}{N} \sum_{i=0}^N x_i \quad (1)$$

$$standard\ deviation = \sigma = \sqrt{\frac{1}{(N-1)} \sum_{i=0}^N (x_i - \bar{x})^2} \quad (2)$$

$$skewness = s = \frac{\sum_{i=0}^N (x_i - \bar{x})^3}{(N\sigma^3)} \quad (3)$$

$$kurtosis = k = \frac{\sum_{i=0}^N (x_i - \bar{x})^4}{(N\sigma^4)} \quad (4)$$

$$entropy = H = E[-\log_2 P(x)] = -\sum_{i=0}^N P(x) \log_2 P(x) \text{ bits} \quad (5)$$

D. Principal Component Analysis

To obtain the main component of the results of statistical analysis, it is necessary to analyze the main component or better known as principal component analysis (PCA). PCA is a preliminary analysis to be used in the continued analysis of a series of analyzes in a study.

In principle, the use of PCA is formed from new factors that have random properties, then the data can be interpreted according to the factors or components that are formed. In reducing the number of variables, a factor analysis process is needed to create a set of new variables or factors that replace a number of variables from the previous data. Flowchart using the principal component analysis can be seen in Fig. 1. The data reduction steps using PCA are described below.

- a. The first step is to prepare statistical analysis data such as average value, standard deviation, kurtosis, and entropy. The data are grouped into one matrix.
- b. The second step is to find the average value for each component of the matrix data.
- c. The third step is to find the value of the covariance of the matrix. Covariance is obtained by using the equation (6).

$$Cov(x, y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1} \quad (6)$$

- d. The fourth step is to find the correlation value of the covariance matrix by dividing the covariance value of each data with standard deviation according to the data sought.
- e. The fifth step is to search the correlation value for the other data so that the squared matrix is obtained.
- f. The sixth step is to search for eigenvalue and eigenvector values.
- g. The seventh step is to allocate each variable of origin into the factor according to the loading value, if there is a value that is almost the same will be implemented rotation.
- h. The eighth step is squaring and summing up the loading values according to the component that aims to find the value of communalities of each component.
- i. The final step determines which component has the dominant influence in constructing the factor.

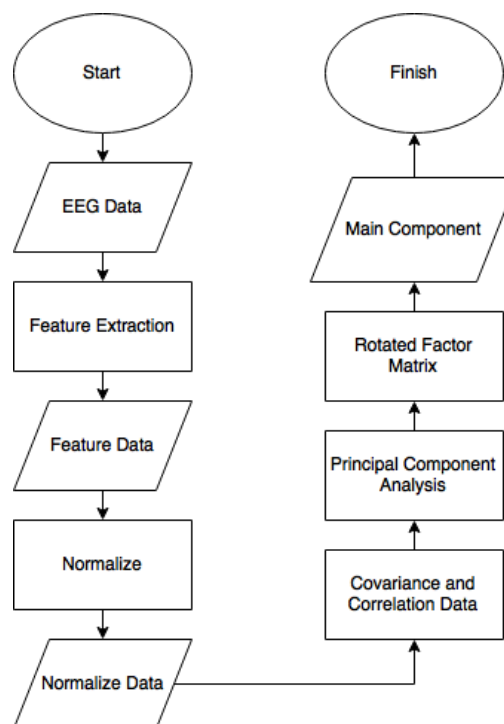


Fig. 1. General Procedure of Proposed Method.

III. Results and Discussion

This section will explain the results of the implementation of principal component analysis (PCA). EEG data from 6 people are grouped based on 9 cognitive tasks and 2 times data retrieval. The data were extracted using statistical features including mean, standard deviation, skewness, kurtosis, and entropy so that 108 variables were obtained for each feature. Due to the diversity of data extracted features it is necessary to normalize the data first. The next step is the implementation of principal component analysis (PCA). The result of feature extraction can be seen in Table 1 and Fig. 2.

Table 1. Feature Extraction

Variable	Observations	Before Normalize		After Normalize	
		Minimum	Maximum	Minimum	Maximum
<i>Mean</i>	108	33.400	42.480	0.085	0.108
<i>Standard Deviation</i>	108	24.481	317.207	0.023	0.299
<i>Skewness</i>	108	-3.085	1.298	-0.645	0.271
<i>Kurtosis</i>	108	-0.304	30.617	-0.007	0.661
<i>Entropy</i>	108	0.349	0.996	0.042	0.120

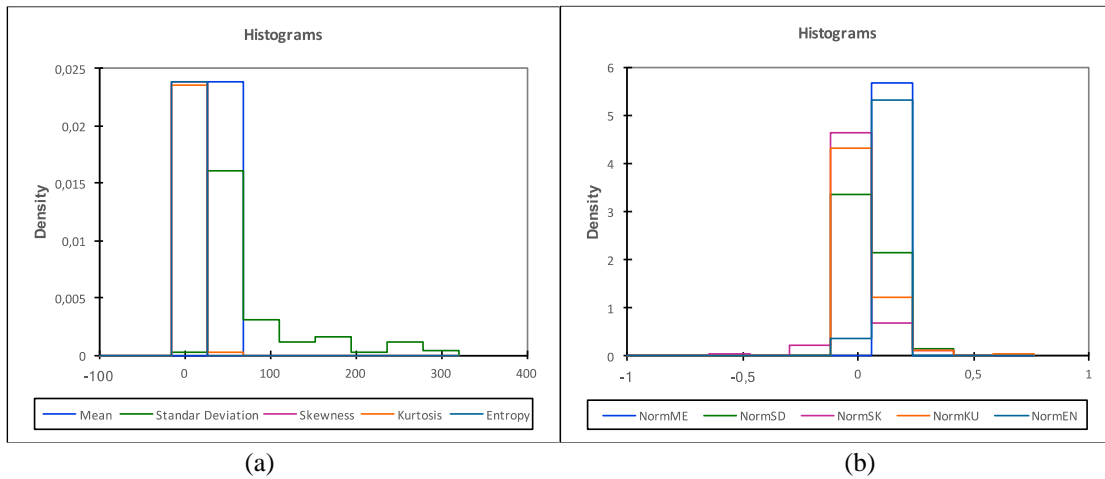


Fig. 2. Histogram Feature Extraction. (a) Raw Data. (b) Normalize Data.

The first step in PCA implementation is to find the correlation value of the covariance matrix by dividing the value of each data covariance with standard deviation according to the data sought. The correlation matrix results shown in Table 2.

Table 2. Correlation Matrix

Variables	Mean	Standard Deviation	Skewness	Kurtosis	Entropy
<i>Mean</i>	1	0.222	-0.058	0.060	0.148
<i>Standard Deviation</i>	0.222	1	-0.211	0.170	0.657
<i>Skewness</i>	-0.058	-0.211	1	-0.524	-0.109
<i>Kurtosis</i>	0.060	0.170	-0.524	1	0.047
<i>Entropy</i>	0.148	0.657	-0.109	0.047	1

The correlation matrix in Table 2 shows the relationship between variables. A high or significant correlation value indicates that both variables are closely related (minimum 0.3, at a significant level of 95%). The correlation matrix results in Table 2 show that there are two variables have the strongest correlation both are skewness-kurtosis and standard deviation-entropy with each correlation value of -0.524 and 0.657.

The second step is to search for eigenvalue and eigenvector values. The eigenvalue and eigenvector shown in Table 3 and Table 4.

Table 3. Eigenvalue

	Eigenvalue	Proportion (%)	Cumulative (%)
λ_1	1.944	38.887	38.887
λ_2	1.342	26.832	65.719
λ_3	0.914	18.271	83.990
λ_4	0.473	9.467	93.457
λ_5	0.327	6.543	100

Based on the Kaiser Criterion Theory, all components having an eigenvalue less than one (<1) will be aborted and an eigenvalue greater than or equal to one (≥ 1) will be maintained [20]. The eigenvalue measure how much variation of the observed variables are explained by factors. Any factor with an eigenvalue greater than or equal to one (≥ 1) explains more variance than a single observed variable.

In Table 3, there are two factors that have eigenvalue value greater than one (>1) with the highest proportion with the cumulative amount of 65.7%. The eigenvalues for successive factors can be viewed in the form of scree plot in Fig. 3. This scree plot can be used to graphically determine the optimal number of factors to retain.

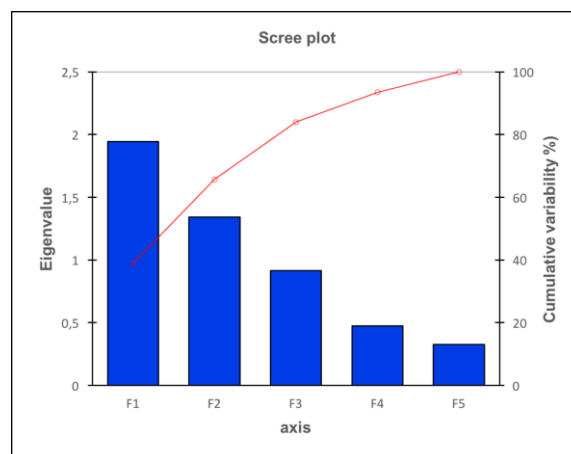


Fig. 3. Scree plot Eigenvalue

Table 4. Eigenvector

	F1	F2	F3	F4	F5
<i>Mean</i>	0.268	-0.195	-0.940	-0.031	-0.078
<i>Standard Deviation</i>	0.585	-0.317	0.170	0.074	0.723
<i>Skewness</i>	-0.418	-0.557	-0.030	0.717	0.028
<i>Kurtosis</i>	0.380	0.603	-0.031	0.693	-0.106
<i>Entropy</i>	0.517	-0.434	0.293	0.003	-0.677

The next step is to allocate each variable of origin into the factor according to the loading value. Factor loadings represent how much a factor explains a variable in factor analysis. Factor loading is shown in Table 5.

Table 5. Loading Factor

	F1	F2	F3	F4	F5
<i>Mean</i>	0.374	-0.226	0.898	-0.022	-0.045
<i>Standard Deviation</i>	0.815	-0.368	-0.163	0.051	0.414
<i>Skewness</i>	-0.583	-0.645	0.029	0.493	0.016
<i>Kurtosis</i>	0.530	0.698	0.029	0.477	-0.061
<i>Entropy</i>	0.720	-0.502	-0.280	0.002	-0.387

Based on the value of eigenvalue which has value greater than one (>1) then will be used two factors of F1 (Factor 1) and F2 (Factor 2). Table 6 shows the variable variance of each factor after varimax rotation. The results show that in Factor 1 and Factor 2 shows a percentage value greater than 30% with a cumulative percentage of 65.7%, while other factors indicate the percentage less than 30%.

Table 6. Percentage of variance after Varimax rotation

	D1	D2	F3	F4	F5
Variability (%)	34,971	30,748	18,271	9,467	6,543
Cumulative %	34,971	65,719	83,990	93,457	100,000

The grouping of the variables on two factors F1 and F2 can be seen in Fig. 4(a), whereas observation of variable distribution on two factors F1 and F2 can be seen in Fig. 4(b).

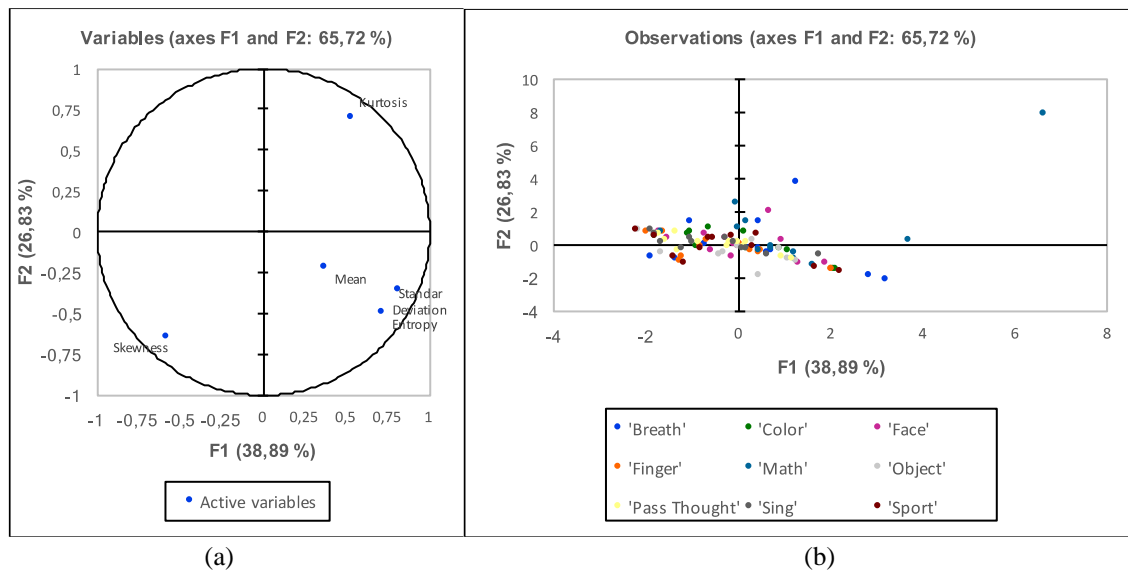


Fig. 4. (a) Principal Component Analysis before rotate. (b) Observations data PCA before rotate.

Prior to the rotation, Factor 1 (F1) includes mean, standard deviation, and entropy with values of 0.374, 0.815, and 0.720, while factor 2 (F2) includes skewness and kurtosis with values of -0.645 and 0.698. Since the distribution of data obtained from F1 and F2 still shows a similar value in skewness and kurtosis, it is necessary to rotate factors by using orthogonal rotation (so-called varimax rotation). The rotation results are shown in Table 7 and Fig. 5.

Table 7. Varimax Rotation

Variables	Varimax Rotation		Proportion (%)	
	F1	F2	F1	F2
Mean	0.436	0.028	10.873	0.050
Standard Deviation	0.879	0.163	44.230	1.721
Skewness	-0.112	-0.862	0.712	48.355
Kurtosis	0.038	0.876	0.081	49.873
Entropy	0.878	-0.002	44.104	0.000

Varimax rotation results in Table 7 show that there is a significant difference in the value of each variable on both factors. The contribution of variables after the Varimax rotation by percentage indicates that in F1 and F2 there are two variables that have a percentage value greater than 40% on each factor including standard deviation and entropy in F1 and skewness and kurtosis in F2.

In Fig. 5 (a), it can be seen that after varimax rotation is applied there are three variables included in Factor 1 (F1) including mean, standard deviation and entropy with values of 0.436, 0.879, and 0.878, while Factor 2 (F2) there are two variables that include skewness and kurtosis with each

value of -0.862 and 0.876. The data distribution results in Fig. 5(b) show more equally distributed spread.

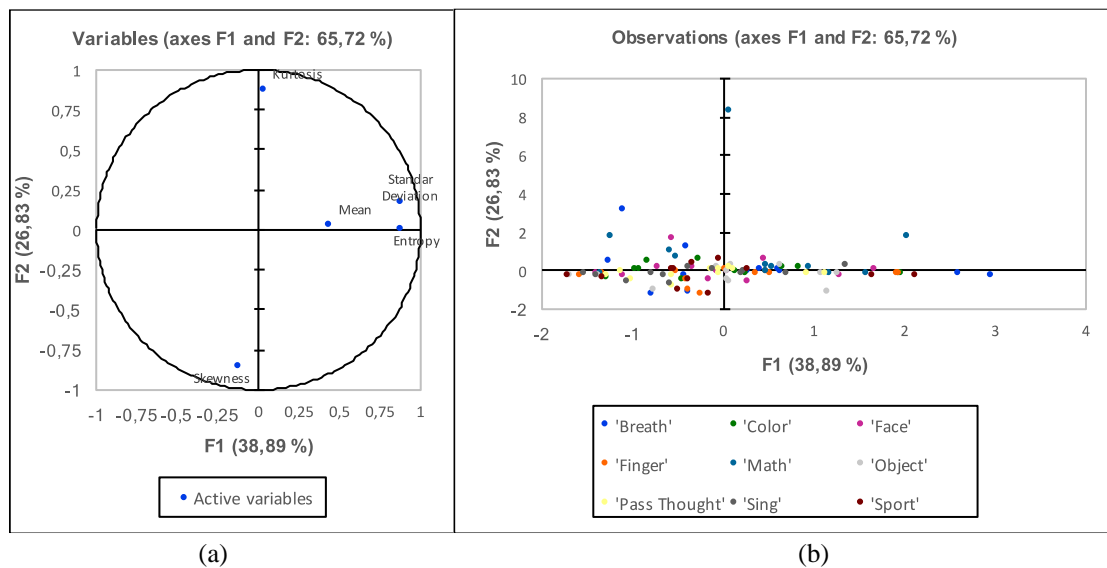


Fig. 5. (a) Principal Component Analysis after rotate. (b) Observations data PCA after rotate.

IV. Conclusion

The results showed that from 108 available datasets, there were five data extraction features including mean, standard deviation, skewness, kurtosis and entropy. These five features are used in principal component analysis. The principal component analysis results show that PCA has succeeded in reducing EEG signals into two main components (PC). Based on the eigenvalues that have been obtained shows there are two factors have a value greater than one (> 1) with the highest proportion with cumulative rate of 65.7%. Three features include the mean, standard deviation, and entropy grouped into Factor 1 (F1) with a percentage of 10.87%, 44.2%, and 44.1%. Two other features include skewness and kurtosis grouped into Factor 2 (F2) with a percentage of 48.35% and 49.87%.

Acknowledgment

This research is supported by LPP Universitas Ahmad Dahlan research grant no. PDP-083/SP3/LPP-UAD/IV/2017.

References

- [1] J. Katona, I. Farkas, T. Ujbanyi, P. Dukan, and A. Kovari, "Evaluation of the NeuroSky MindFlex EEG headset brain waves data," in *Applied Machine Intelligence and Informatics (SAMII), 2014 IEEE 12th International Symposium on*, 2014, pp. 91–94.
- [2] G. Rebolledo-Mendez *et al.*, "Assessing neurosky's usability to detect attention levels in an assessment exercise," in *International Conference on Human-Computer Interaction*, 2009, pp. 149–158.
- [3] N. Inc., *Brainwave Signal (EEG) of Neurosky, Inc.* Neurosky, Inc., 2009.
- [4] J. Klonovs, C. K. Petersen, H. Olesen, and J. S. Poulsen, "Development of a Mobile EEG-Based Feature Extraction and Classification System for Biometric Authentication," Aalborg University Copenhagen, 2012.
- [5] A. Azhari and L. Hernandez, "Brainwaves feature classification by applying K-Means clustering using single-sensor EEG," *Int. J. Adv. Intell. Inform.*, vol. 2, no. 3, pp. 167–173, Nov. 2016.
- [6] M. Diykh, Y. Li, and P. Wen, "EEG Sleep Stages Classification Based on Time Domain Features and Structural Graph Similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, Nov. 2016.
- [7] Z. H. Murat, M. N. Taib, S. Lias, R. S. S. A. Kadir, N. Sulaiman, and M. Mustafa, "EEG Analysis for Brainwave Balancing Index (BBI)," 2010, pp. 389–393.

- [8] R. Zafar, A. S. Malik, H. U. Amin, N. Kamel, S. Dass, and R. F. Ahmad, "EEG spectral analysis during complex cognitive task at occipital," in *Biomedical Engineering and Sciences (IECBES), 2014 IEEE Conference on*, 2014, pp. 907–910.
- [9] A. Patil, C. Deshmukh, and A. R. Panat, "Feature extraction of EEG for emotion recognition using Hjorth features and higher order crossings," in *Advances in Signal Processing (CASP), Conference on*, 2016, pp. 429–434.
- [10] I. Jayarathne, M. Cohen, and S. Amarakeerthi, "BrainID: Development of an EEG-based biometric authentication system," in *Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016 IEEE 7th Annual*, 2016, pp. 1–6.
- [11] B. H. Tan, "Using a low-cost eeg sensor to detect mental states," Carnegie Mellon University, 2012.
- [12] P. Ackermann, C. Kohlschein, J. Á. Bitsch, K. Wehrle, and S. Jeschke, "EEG-based automatic emotion recognition: Feature extraction, selection and classification methods," in *e-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on*, 2016, pp. 1–6.
- [13] Y. Zhong and Z. Jianhua, "Recognition of Cognitive Task Load levels using single channel EEG and Stacked Denoising Autoencoder.pdf," in *Proceedings of the 35th Chinese Control Conference*, 2016, pp. 3907–3912.
- [14] A. Mohamed, K. B. Shaban, and A. Mohamed, "Directed graph-based wireless EEG sensor channel selection approach for cognitive task classification," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2016 International*, 2016, pp. 176–181.
- [15] Z. Pang, J. Li, H. Ji, and M. Li, "A new approach for EEG feature extraction for detecting error-related potentials," in *Progress in Electromagnetic Research Symposium (PIERS)*, 2016, pp. 3595–3597.
- [16] A. Saha, A. Konar, P. Das, B. S. Bhattacharya, and A. K. Nagar, "Data-point and feature selection of motor imagery EEG signals for neural classification of cognitive tasks in car-driving," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–8.
- [17] M. Varela, "Raw EEG signal processing for BCI control based on voluntary eye blinks," in *2015 IEEE Thirty Fifth Central American and Panama Convention (CONCAPAN XXXV)*, 2015, pp. 1–6.
- [18] B. Johnson, T. Maillart, and J. Chuang, "My thoughts are not your thoughts," 2014, pp. 1329–1338.
- [19] R. Chai *et al.*, "Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 4654–4657.
- [20] J. Braeken and M. A. L. M. van Assen, "An empirical Kaiser criterion.," *Psychol. Methods*, vol. 22, no. 3, pp. 450–466, Sep. 2017.