Journal of Engineering Science and Technology
Special Issue on ACSAT, October (2021) 101 - 110

ELECTROENCEPHALOGRAM EXTRACTION AND SELECTION SYSTEM REPRESENTATIVE FEATURES

AHMED RASHID ALKHUWAYLIDEE[1,*], HADI RATHAM AL GHAYAB[2], MUHANAD AL-HASHIMI[3] MOHAMMED A. KASHMOOLA[4]

[1]College of Information Technology, Mazaya University College, Iraq
[2]Faculty of Health, Engineering and Sciences, University of Southern Queensland, QLD 4350, Australia,
[3]Tikrit University, Salah Al Din Governorate, Iraq
[4]University of Al-Hamdaniya, Iraq
*Corresponding Author: dr.ahmedrm83@gmail.com

Abstract

The brain signals are related to the human's activity that reflected by Electroencephalogram (EEG). EEG recordings are complex signals in general that is being non-stationarity and non-linear. In last decade, this signal was studied by many researchers as EEG contains important information about the body's activities. In this study, we proposed a new analysis and classification scheme that employs discrete wavelet transform (DWT) and Information Gain (InGain) denoted as DWT-InGain method. For the study of EEG signals, first step, DWT is applied to analyse the EEGs into frequency bands. Secondly, all bands are divided into windows, and from each window we extract most common statistical features. After that, the InGain is utilised to select the most important features. At the end, the extracted and selected features are used to feed the most popular classifier, support vector machine (SVM) to evaluate the execution of the proposed DWT-InGain scheme. This method is tested on a benchmark EEG database and obtains great results for five different epileptic EEG pairs in term of accuracy, sensitivity and specificity. The consequences of the proposed system might help the doctors, researchers and experts to reveal the epileptic seizures.

Keywords: Brain signal data, DWT, Electroencephalogram (EEG), Information gain, Support vector machine.

1. Introduction

The most important part of the human body is the brain that is controlled human's activity. Through number of electrodes located in the scalp, the brain functions have reflected in waves as can be seen in an electroencephalograph (EEG) chart. Although, the EEG includes huge information of human's activities that make them important test to study these activities, analysing EEG data is a very challenging task due to their characteristics [1, 2]. It is really fatigued to obtain and then choose the representative attributes from EEG recordings. Nevertheless, to analyse the EEG signals, several methods were applied. Many of these approaches place under the time domain, frequency domain and time-frequency domain [3].

Extracting features from EEG signals, time domain approaches have been extensively implemented to the EEG analysis. The roughly popular time domain methods are linear prediction, Principal analysis of components and independent analysis of components, used for time series analysis [4-7]. However, frequency domain methods, which are parametrically and non-parametrically spectral analysis approaches, have been applied for EEG analysis to Identify EEG signal epileptic seizures [8, 12]. The spectral estimation methods are the analysis of data in the frequency domain by computational and transformational methods by Fourier [3].

Other time-frequency domain methods that used to relocate the signals continually through the period of time to extract a set of coefficients at each moment such as wavelet transforms. These methods have been used to analyse the EEG data. Three types of Wavelet Transformation, Discrete Transformation Wavelet (DWT), Continuous Transformation Wavelet, and wavelet packet decomposition have been widely applied in biomedical signals field. Murugappan, Sharma and Alickovic [13-15] employed DWT in their work for EEG classification. However, the wavelet transform methods are limited to reduce the huge size of EEG data.

This research presents a potent feature extraction technique to analyse and classification the EEG recordings. The system is combined the discrete wavelet transform (DWT) and information gain (InGain) technique that is denoted as DWT-InGain. A Support Vector Machine (SVM) is implemented to test the functional extraction system's output. In sections 2-3 demonstrate the specifics of our proposed technique. Section 4 explaining the experimental results and data used and also in Section 5 has the conclusion of this method provided.

2. Methodology

This Section provides an exhaustive description of the proposed feature extraction system. Figure 1 shows the structure of the suggested feature extraction and dimension reduction system. In order to extract and select the most representative features for EEG classification, the DWT and InGain methods are used. The proposed DWT-InGain approach with SVM are carried out the EEG classification process evaluation.

2.1. Discrete wavelet transforms (DWT)

DWT approach is the specific case with a wavelet transform those analyses and efficiently measures a signal dependent on time and frequency and provides a compact signal output. In this research study, the epileptic EEG data are divided into various bands, based on the decomposition level. In decomposition first level, the EEG signals are transformed into two numerically vectors in same length, which are (LPF_1) low pass filter and (HPF_1) high pass filter. Sequentially, previous LPF is decomposed into HPF_2 and LPF_2 that is the level 2 of decomposition. This procedure is continued to the end level of decomposition of EEG signals. The duration of each stage of LPF and HPF is equivalent and half preceding LPF_n length [16]. Figure 2 represents the work of DWT method of EEG decompose. In this study, we decomposed the EEG recordings into five frequency bands: gamma (γ), beta (β), alpha (α), theta (θ) and delta (δ).

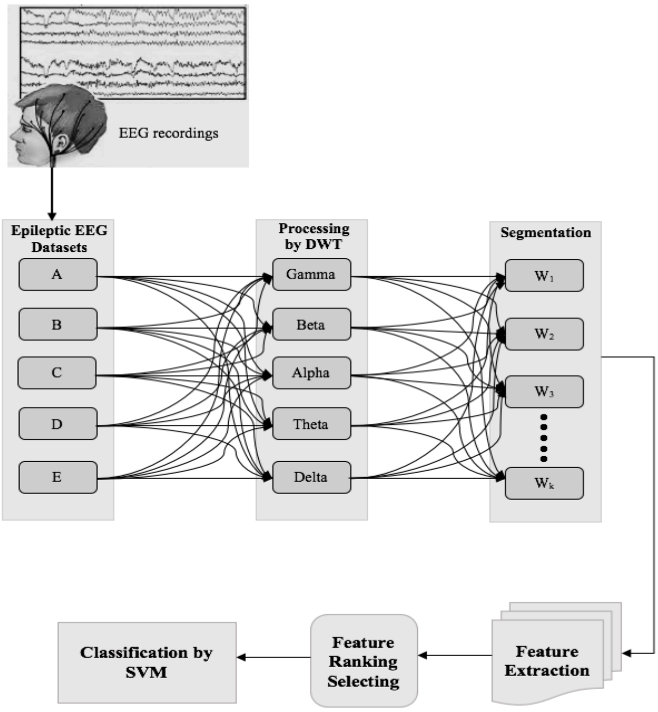

Fig. 1. Block diagram of the suggested method for EEG analysis and classification.

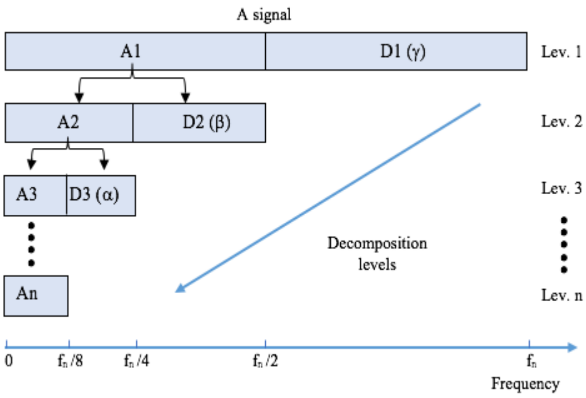

Fig. 2. Operation of EEG analysis using the DWT.

2.2. Segmentation bands

Generally, the frequency bands for EEG decomposition are non-stationary, however, the segmentation of EEG bands into number of small parts called windows can make these bands of EEGs quasi-stationary. These windows are determined empirically based on the Eq. (1):

$$W = floot\ \frac{B}{fs} \tag{1}$$

where *W* refers to the number of windows, *B* is the frequency bands, and *fs* refers to the number of feature set of each window.

Next, we extract statistical features from each window and reset as one vector, called FWDn as shown in Eq. (2):

$$FWD_n = \sum_{i=1}^{W} S(i)\,; \qquad n = 1, 2, \ldots\ldots B \tag{2}$$

where FWD_n refers to the statistical feature set from each frequency band, *n* is the frequency bands, and *w* refers to the number of windows.

2.3. Statistical features

As the EEG recordings have huge size of important data for the human's activities, ideally the most important information is the statistical features [17]. From each window of the bands, this study extracts and puts nine statistical features into a set to reduce a huge amount of EEG. These statistical features are *maximum* (*Max*), *minimum* (*Min*), *mean* (*Mea*), *median* (*Med*), *mode* (*Mod*), *first quartile* (*q1*), *second quartile* (*q2*), *range* (*Ran*) and *standard deviation* (*Std*) as can be seen in Table 1. A set of features with 4097 data dots, in 36 dimensions, is obtained from every band using the DWT method. The following phase in the proposed system is used these features.

Table 1. Equations of the extracted statistical features.

$Max = max[Ts_n]$	(3)	$q1 = \frac{1}{4(N+1)} th$	(8)
$Min = min[Ts_n]$	(4)	$q2 = \frac{2}{4(N+1)} th$	(9)
$Mea = \frac{1}{n}\sum_{1}^{n} Ts_n$	(5)	$Ran = Max - Min$	(10)
$Med = \left(\frac{N+1}{2}\right)^{th}$	(6)	$Std = \sqrt{\sum_{n=1}^{N}(Ts_n - M)\frac{2}{n-1}}$	(11)
$Mod = 3Med - 2\frac{Ts_n}{N}$	(7)		

where Ts_n= 1, 2, …., *n*, is a time series; *N* refers to the number of data points, *M* is the mean of the sample.

2.4. Information gain

The information gain technique (InGain) is applied to enhance classification sets by reducing the dimension of the key features and removing non-relative data from the classification process [18]. The InGain is employed to rank the feature set and from the huge amount of data provided by the EEG signals select the most important features. The technique relies on the concept of entropy from the theory of information, as indicates in Eq. (12) [19]:

$$Info(X) = -\sum_{j=0}^{n} P_j \, log_2\left(P_j\right) \tag{12}$$

where Pj is the prior probability for the *j*th discretised value of X. The entropy of X after observing another variable Y is then defined as:

$$Info\left(\frac{X}{Y}\right) = -\sum_{r=1}^{m} \frac{|y_r|}{|X|} \sum_{j=0}^{n} P\frac{x_j}{y_j} \, log_2\left(P\frac{x_j}{y_j}\right) \tag{13}$$

The InGain is calculated as the difference between the original information and the new information after divided on Y. It is given as following equation:

$$InGain = Info(X) - Info\left(\frac{X}{Y}\right) \tag{14}$$

In this study, each feature in the different EEG signals bands is classified based on a decreasing adaptive threshold using the InGain method to choose the most important statistical features.

2.5. Support vector machines (SVM)

The SVM has been developed by Cortes and Vapnik as a well-known classification method. [20]. SVM 's latest vibration is a non-linear support vector machine, kernel-quadratic, which is called QSVM and which Dagher has proposed [21], is used in this study. This method distinguishes the feature set as described in [19]in a nonlinear way. The key explanation for using the SVM classifier in this analysis is that SVM offers a solution that gives a really strong overall sample if the parameters are correctly chosen. That means selecting suitable parameters, the SVM can be robust, although the using sample has some bias [22].

3. Measurements Tools

Several measurement tools were conducted to evaluate the efficiency of the suggested DWT-InGain approach. Fold cross validation method is one of the tools, which is used to divide the input datasets into *k* parts or subsets called folds [2]. The processing of this method is repeated to acquire *k* folds. The test of the SVM classifier is evaluated one-fold in every iteration and *k*-1 folds are applied as training set to train the classifier. Table 2 presents a sample size of training and testing sets. An average accuracy is achieved from the operation of the cross-validation approach. *K* folds are utilized to test the classifier and the accuracy rate is counted through fivefold cross validation as is seen in the Eq. (15):

$$\text{Accurcy rate} = \frac{\sum \text{Performance}}{\text{FiveFolds}} \tag{15}$$

In this research, accuracy measurement was employed also to evaluate the proposed feature extraction system. The accuracy calculated as the Eq. below [7, 23]:

$$Accuracy = \frac{\sum \text{True positives} + \sum \text{True negatives}}{\sum \text{All samples}} \times 100 \tag{16}$$

However, other measurement tools applied in this research were sensitivity and specificity as shown in Eqs (17) and (18), respectively [23].

$$\text{Sensitivity} = \frac{\sum \text{True positives}}{\sum \text{All positive samples}} \times 100 \tag{17}$$

$$\text{Specificity} = \frac{\sum \text{True negatives}}{\sum \text{All negative samples}} \times 100 \tag{18}$$

4. Experimental Results and Discussions

In this research, a publicly available database is used, which is widely tested in many of the research work such as [1, 2, 7], [23-28], and the Bonn University, Germany was collected this database [29, 30]. There are five different EEG (A-E) databases usable. Sets A and B have been collected from five healthy subjects with open eyes and closed. C-E sets have been collected from five separate patients. In epileptic patients free of seizures, sets C and D were recorded. During active seizures, Set E was taken from epileptic subjects. EEG recordings with a 12-bit resolution is digitalized to 173.61 Hz. Every dataset (A, B, C, D and E) therefore contained 100 channels, a sample length of 4096 with a duration of 23.6 seconds each to avoid multi-channel EEG continuous recording after a visual artifact's inspection. [12]. The DWT method was applied to decompose the five classes of epileptic EEG signals into five frequency bands (γ, β, α, θ and δ). To extract the representative features, each band was divided empirically into four windows and extract nine statistical features from each window. In the next phase, the obtained feature set were processed over the InGain technique to select the most important features and put them in a set. The key features were obtained and forwarded to the SVM classifier. Here, several cases for epileptic EEG signals database were tested. In addition, the experiment results were implemented using MATLAB R2017b.

In this experiment, we applied the proposed features extraction system on five epileptic pairs, {A vs. E}, {A vs. D}, {B vs. E}, {C vs. E}and {D vs. E}as appeared in Table 2. The proposed system was conducted and evaluated through several evaluation tools. Table 3 presents the average accuracies of different epileptic cases.

Table 2. Number of the training and testing sets of each case for five frequency bands used in this research.

Method	Case	Parameter	Frequency bands				
			γ	***β***	***α***	***θ***	***δ***
DWT_InGain	A vs. E	**Total**	4837	4733	4263	3925	3892
		Training	3628	3550	3197	2944	2919
		Testing	1209	1183	1066	981	973
	A vs. D	**Total**	4323	4463	4475	4142	4210
		Training	3242	3347	3356	3106	3157
		Testing	1081	1116	1119	1036	1053
	B vs. E	**Total**	4734	4691	4283	3920	3936
		Training	3550	3518	3212	2940	2952
		Testing	2284	1173	1071	980	984
	C vs. E	**Total**	5142	5033	4445	3881	3701
		Training	3856	3775	3334	2911	2776
		Testing	1286	1258	1111	970	925
	D vs. E	**Total**	5128	5328	4854	4305	4438
		Training	3846	3996	3640	3228	3328
		Testing	1282	1332	1214	1076	1110

That were achieved from each frequency band by applied the SVM classifier. In Table 3, the DWT-InGain feature extraction system based on the SVM classifier obtained a 99.98%, 94.46%, and 100% overall classification accuracy in γ, β, α, θ and δ frequency bands, respectively, of {A vs. E} case. In case No. 2, the highest accuracy is in γ frequency band with 98.96%, while the lowest accuracy achieved in same case is in β band with 88.66%. In {B vs. E} case, the proposed system yielded a 100% average classification accuracy in both θ and δ bands, which was the highest accuracy compared with other frequency bands. In same case, which is No. 3, the second highest accuracy is achieved from γ frequency band, however, β frequency bands has a lowest overall accuracy with a 95.35%. From case No. 4 observed that the DWT-InGain achieved a 99.94%, 95.35%, 100%, 99.98%, and 100% accuracy rate in γ, β, α, θ and δ frequency bands, respectively. In addition, the proposed system yielded a 99.88% average accuracy in γ and α bands, which the highest in case No. 5.

Table 3. Classification accuracy of each frequency band of the DWT-InGain method for different cases of epileptic EEG data.

No.	Case	Accuracy (%)				
		γ	***β***	***α***	***θ***	***δ***
1	**A vs. E**	99.98	94.46	100	100	100
2	**A vs. D**	98.96	88.66	95.22	92.9	96.87
3	**B vs. E**	99.83	95.52	99.02	100	100
4	**C vs. E**	99.94	95.35	100	99.98	100
5	**D vs. E**	99.88	92.94	99.88	99.77	99.64

The sensitivity measurement tool, the proposed feature extraction system obtained high sensitivity average score with more than 98.93% in *γ* band for all epileptic cases. On the other hand, the proposed system yielded a 100% sensitivity rate in α, θ and δ frequency bands for some cases as can be shown in Fig. 3, which presents a performance of the proposed DWT-InGain system with SVM classifier on five cases. The low results are achieved in β frequency band for all five epileptic cases. However, compared among five cases, the case {A vs. D} is achieved sensitivity rates with the proposed system as it seen in Fig. 3. The set A in case {A vs. D} was analogous to the set D that led to the low sensitivity scores compared with the sensitivity rates for other cases.

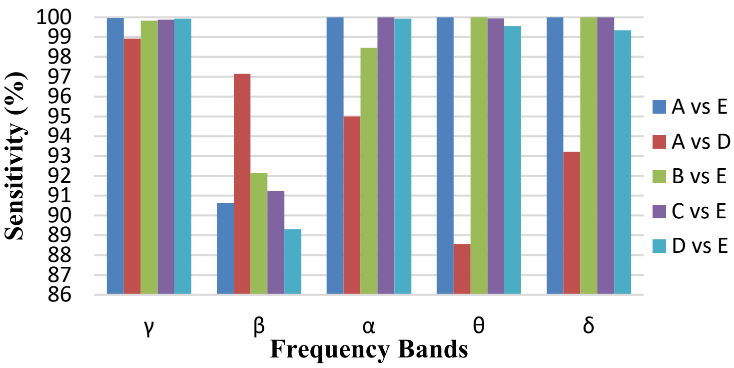

Fig. 3. Testing of the proposed DWT-InGain approach with SVM classifier on five epileptic cases in term of sensitivity.

Additionally, Table 4 provides overall specificity measurement scores in the four-fold cross validation of each frequency band through the proposed system for five epileptic EEG cases. From Table 4, the highest results were achieved in δ band with 100% specificity rates in most studied epileptic cases. As a result, from Tables 3 and 4 and Fig. 3, the DWT-InGain feature extraction system can provide perfect result over the SVM classifier in γ frequency band for all epileptic EEG cases.

Table 4. Specificity score of each frequency band of the DWT-InGain feature extraction system for five cases of epileptic EEG signals.

No.	Case	Specificity (%)				
		γ	*β*	*α*	*θ*	*δ*
1	**A vs. E**	100	98.71	100	100	100
2	**A vs. D**	98.87	83.78	95.19	96.97	99.6
3	**B vs. E**	99.85	99.29	99.53	100	100
4	**C vs. E**	100	99.22	100	100	100
5	**D vs. E**	99.84	96.66	99.84	100	100

5. Conclusion

This research presented a new scheme for EEG feature extraction. A set of features was extracted through the DWT-InGain method implementation and was used as input to the popular machine learning classifier (SVM). The tested of suggested scheme was evaluated by the accuracy, sensitivity and specificity measurement tools. From the evaluation results, the DWT-InGain feature extraction scheme based on the SVM classifier has ability to classify the epileptic seizures with a reasonable performance. The DWT-InGain technique can assist the specialists to decompose a missive size of EEG recordings and to extract the most representative features. In the future, the suggested DWT-InGain system will be adapted to analyse and classify the online EEG time series rather than offline EEG data.

References

1. Al Ghayab, H.R.; Li, Y.; Siuly, S.; and Abdulla, S. (2019). A. feature extraction technique based on tunable q-factor wt for brain signal classification. *Journal of Neuroscience Methods*, 312, 43-52.
2. Al Ghayab, H.R.; Li, Y.; Abdulla, S.; and Wen, P. (2017). Developing a tunable q-factor wavelet transform based algorithm for epileptic eeg feature extraction. *In International Conference on Health Information Science*, 45-55.
3. Acharya, U.R.; Sree, S.V.; Swapna, G.; Martis, R.J.; and Suri, J.S. (2013). Automated eeg analysis of epilepsy: A review. *Knowledge-Based Systems*, 45(2013), 147-165.
4. Joshi, V.; Pachori, R.B.; and Vijesh, A. (2014). Classification of ictal and seizure-free eeg signals using fractional linear prediction. *Biomedical Signal Processing and Control*, 9(2041), 1-5.
5. Artoni, F.; Delorme, A.; and Makeig, S. (2018). Applying dimension reduction to eeg data by Principal Component Analysis reduces the quality of its subsequent *Independent Component Decomposition. NeuroImage*, 176-187.

6. Chaitanya, G.; Sinha, S.; Narayanan, M.; and Satishchandra, P. (2015). Scalp high frequency oscillations (HFOs) in absence epilepsy: an independent component analysis (ICA) based approach. *Epilepsy Research*, 115, 133-140.
7. Al Ghayab, H.R.; Li, Y.; Abdulla, S.; Diykh, M.; and Wan, X. (2016). Classification of epileptic EEG signals based on simple random sampling and sequential feature selection. *Brain Informatics*, 3(2), 85-91.
8. Zhang, Y.; Liu, B.; Ji, X.; and Huang, D. (2017). Classification of EEG signals based on autoregressive model and wavelet packet decomposition. *Neural Processing Letters*, 45(2), 365-378.
9. Wu, Y.; Jiao, B.; Wu, Z.; Zhen, J.; Jia, Q.; Zhang, H.; Guan, B.; and Wang, S. (2015). Autoregressive spectral analysis of cortical electroencephalographic signals in a rat model of post-traumatic epilepsy. *Neurological Research*, 37(11), 959-966.
10. Petroff, O.A.; Spencer, D.D.; Goncharova, I.I.; and Zaveri, H.P. (2016). A comparison of the power spectral density of scalp EEG and subjacent electrocorticograms. *Clinical Neurophysiology*, 127(2), 1108-1112.
11. Myers, M.H.; Jolly, E.; Li, Y.; de Jongh Curry, A.; and Parfenova, H. (2017). Power spectral density analysis of electrocorticogram recordings during cerebral hypothermia in neonatal seizures. *Annals of Neurosciences*, 24(1), 12-19.
12. Al Ghayab, H.R.; Li, Y.; Siuly, S.; and Abdulla, S. (2018). Epileptic eeg signal classification using optimum allocation based power spectral density estimation. *IET Signal Processing*, 12(6), 738-747.
13. Murugappan, M.; Ramachandran, N.; and Sazali, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering*, 3(4), 390-396.
14. Sharma, R.; Pachori, R.B.; and Acharya, U.R. (2015). An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures. *Entropy*, 17(8), 5218-5240.
15. Alickovic, E.; Kevric, J.; and Subasi, A. (2018). Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomedical Signal Processing and Control*, 39, 94-102.
16. Acharya, U.R.; Vidya, K.S.; Ghista, D.N.; Lim, W.J.E. Molinari, F.; and Sankaranarayanan, M. (2015). Computer-aided diagnosis of diabetic subjects by heart rate variability signals using discrete wavelet transform method. *Knowledge-Based Systems*, 81, 56-64.
17. Siuly, S.; Kabir, E.; Wang, H.; and Zhang, Y. (2015). Exploring sampling in the detection of multicategory EEG signals. *Computational and Mathematical Methods in Medicine*, Volume 2015, Article ID 576437 .
18. Deng, J.D.; Simmermacher, C.; and Cranefield, S. (2008). A study on feature analysis for musical instrument classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(2), 429-438.
19. Han, J.; Pei, J.; and Kamber, M. (2011). D*ata mining: concepts and techniques*. Elsevier.
20. Cortes, C.; and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-97.

21. Dagher, I. (2008). Quadratic kernel-free non-linear support vector machine. *Journal of Global Optimization*, 41(1), 15-30.
22. Auria, L.; and Moro R.A. (2008). Support vector machines (SVM) as a technique for solvency analysis. *German Institute for Economic Research Discussion Paper Series*.
23. Al Ghayab, H.R.; Li, Y.; Siuly, S.; and Abdulla, S. (2018). Epileptic seizures detection in eegs blending frequency domain with information gain technique. *Soft Computing*, 23(1), 227-239.
24. Siuly, J.; and Li, Y. (2014). A novel statistical algorithm for multiclass eeg signal classification. *Engineering Applications of Artificial Intelligence*, 34, 154-167.
25. Hassan, A.R.; Siuly, S.; and Zhang, Y. (2016). Epileptic seizure detection in eeg signals using tunable-q factor wavelet transform and bootstrap aggregating. *Computer Methods and Programs in Biomedicine*, 137, 247-259.
26. Patidar, S.; Pachori, R.B.; Upadhyay, A.; and Acharya, U.R. (2017). An integrated alcoholic index using tunable-Q wavelet transform based features extracted from EEG signals for diagnosis of alcoholism. *Applied Soft Computing*, 50, 71-78.
27. Alçin, Ö.F.; Siuly, S.; Bajaj, V.; Guo, Y.; Şengu, A.; and Zhang, Y. (2016). Multi-category eeg signal classification developing time-frequency texture features based fisher vector encoding method. *Neurocomputing*, 218, 251-258.
28. Siuly, S.; and Li, Y. (2015). Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. *Computer Methods and Programs in biomedicine*, 119(1), 29-42.
29. EEG time series (2005). Retrieved April 5, 2021, http://www.meb.unibonn.de/epileptologie/science/ physik/eegdata.html.
30. Andrzejak, R.G.; Lehnertz, K.; Mormann, F.; Rieke, C.; David, P.; and Elger, C.E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Physical Review E*, 64, 061907.