JOURNAL OF OPTOELECTRONICS LASER
Volume 41 Issue 8, 2022

ISSN: 1005-0086

EEG Epileptic Seizures Clustering Based on K-means and wavelet

Hanan Hassan[1],Raid Luaibi Lafta[1*]
[1]Department of Computer Science , College of Education for pure Sciences, University of Thi-Qar, Iraq.

ABSTRACT
The identification of epileptic episodes using EEG data is essential for making a correct epilepsy diagnosis. Visual inspection is the major method used by experts to spot epileptic convulsions, although it is a laborious task for them. This study suggested creating an automated model to determine whether or not a patient has epilepsy. If successful, this model might enhance clinical research, make specialists' jobs easier, and speed up the detection process. The development of an automated model to identify epileptic seizures from EEG data is the goal of this research. In this study, we present a new method for seizure classification for EEG signals using statistical features coupled with a classifier unsupervised. To achieve this task, each single EEG channel has been divided into segments. In addition, each segmented applied db4 Dwt that result is approximate and detailed then a vector of statistical features are pulled out from each sub-cluster from the final features set. The obtained features set are sent to the K-means classifier. Our developed model has been tested with an epileptic dataset and the proposed model produced an average accuracy, sensitivity, and specificity of 100%, 100%, and 100%, respectively.

Keywords: Epileptic seizures, K-means, Statistical features, EEG signals.

I. INTRODUCTION

EEG is a clinical technique used to track a patient's condition and identify any neurological problems [1]. Epilepsy is a neurological disease that develops when the brain experiences an aberrant discharge. It is linked to sudden alterations in the brain's electrical activity. Epileptic seizures are described by experts as signals connected with a slow-spike wave. The majority of seizure instances have an unknown origin, which may have an impact on the subject's everyday life by causing them to lose consciousness and/or their memory, as well as increase their chance of dying suddenly [2, 3]. According to recent studies, 4% of the population has seizures. [4, 5].

Traditional methods of EEG analysis are deemed unsuitable since they put a significant time and stress on neurologists. Visual examination of the patient's records may result in mistakes and false positives [6]. Additionally, while being recorded, artifacts and background noise are mixed in with the EEG signals. The neurologists must develop automated methods for identifying epileptic seizures in order to get over these problems.

It is said that the brain network is a complicated nonlinear system. That implies using a single-channel EEG to identify seizures is insufficient. It has been demonstrated that processing multi-channel EEG to detect seizures is an essential component of human brain analysis [7, 8].

2. RELATED WORKS

Many previous studies focused on the classification of EEG, including:

Bizopoulos, Paschalis A. Tsalikakis, Dimitrios G. Tzallas, Alexandros T. Koutsouris, Dimitrios D. Fotiadis, Dimitrios I.[9] With an unsupervised approach based on K-means clustering and Ensemble Empirical Decomposition, the current study suggests an automated way to identify epileptic episodes (EEMD). EEG segments are acquired from a publically accessible dataset and divided into two groups: seizure-related and non-seizure-related. The Marginal Spectrum (MS) of each EEG segment is determined using EEMD. The averages of these intervals are then utilized as input features for K-Means clustering once the MS has been partitioned into equal intervals. The evaluation's findings are highly encouraging, showing a 98 percent overall accuracy.

Prabhakar, Sunil Kumar Rajaguru, Harikumar [10] In this study, the classification of epilepsy risk levels from EEG signals uses approximate entropy (ApEn) as a feature extraction technique, followed by K-means clustering and principal component analysis (PCA) as post classifiers. Showing K-means 85% percent overall accuracy and 91% with PCA.

Manjusha, M. Harikumar, R.[11] The performance of the KNN classifier and K-means clustering for the classification of epilepsy risk level from EEG data is examined in this research. Detrend analysis is used to find any non-linearity in the data. Twenty patient's EEG data are examined. It is then utilized to reduce the dimensionality of the power spectral density. K-means clustering and the KNN classifier both attained performance indices of 78.31 and 93.02 percent, respectively.

3. MATERIALS AND METHODS

In this paper, we propose algorithm for classifying EEG signals using statistical features and K-means for classification. The model of the proposed method using the k_means to analyses EEG signals is presented in Fig. 1. More details regarding the proposed method and dataset will be presented in the next sections. 2033–2035.

 ISSN: 1005-0086

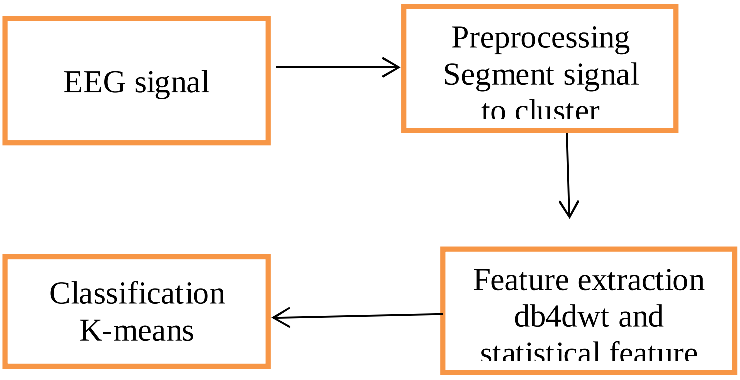

Figure 1. Block diagram of the proposed methodology for EEG signal classification

4. EEG RECORDING AND DATA ACQUISTION

The EEG data set employed in this study are in [12] the data set contains five different sets represented as A, B, C, D, and F with one included 100 single-channel EEG signal. EEG data was recorded at a sampling rate of 173.61 Hz. All EEG signals were filtered using Band-pass filter. Fig.2 displays the 10–20 system used to record EEG signals. Fig.1 displays the 10-20 system used to electrode EEG signals. Sets A and B were acquired from five healthy volunteers while Sets C, D, and F were recorded from five epileptic patients. Sets A and B were recorded activity recordings from seizure free. To make it clear, four sets A, B, C and D were from normal EEG signals while set E was from epileptic EEG signals.

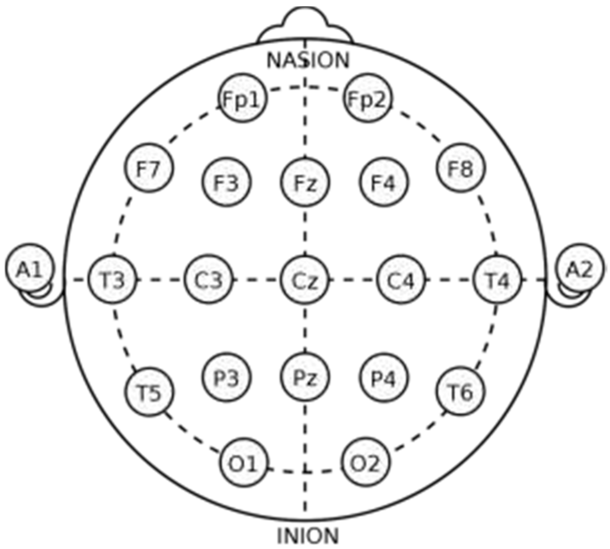

Fig.2 The10-20 system of electrode placement for recording an EEG pattern

 DOI: 10050086.2022.08.68

ISSN: 1005-0086

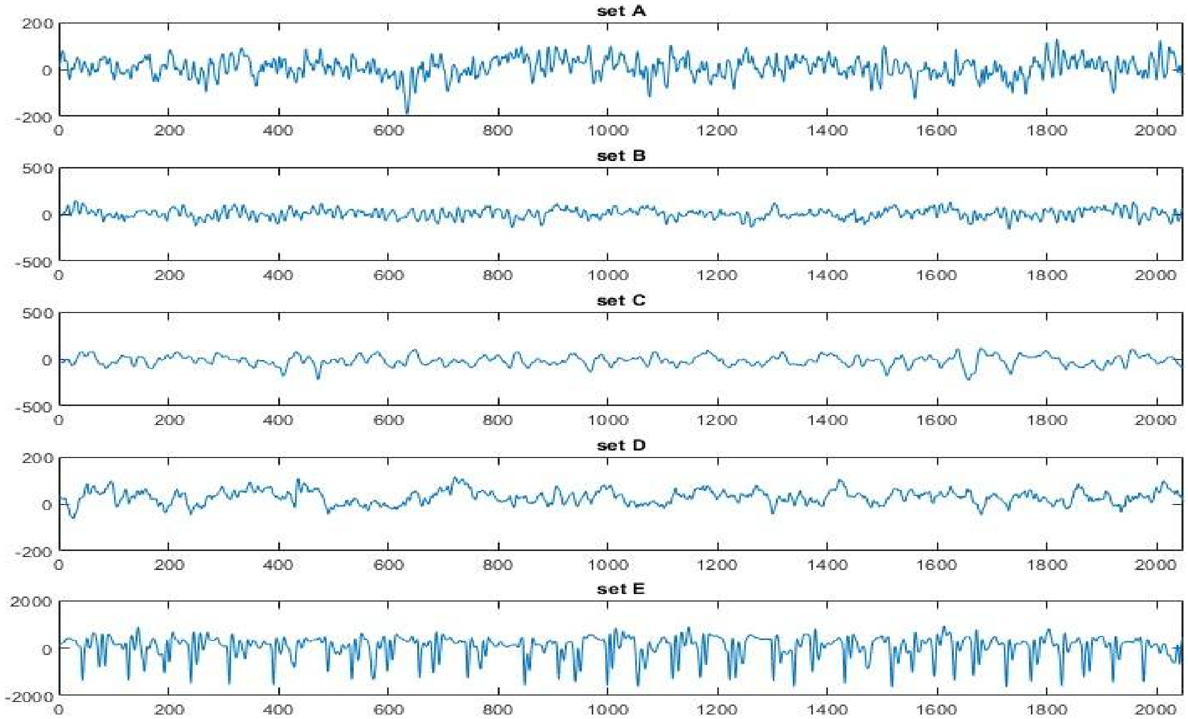

Fig.3EEGsamplesfromsetsA, B, C, D and E

Segmentation Technique

Due to EEG data are non-stationary, it is important to divide EEG signal into intervals. Then, a vector of features is pulled out from each interval. We propose a feature extraction model to represent EEG signals Fig.4 shows the steps of our model to extract the features and the segmentation method. These stages will be discussed in the following subsection.

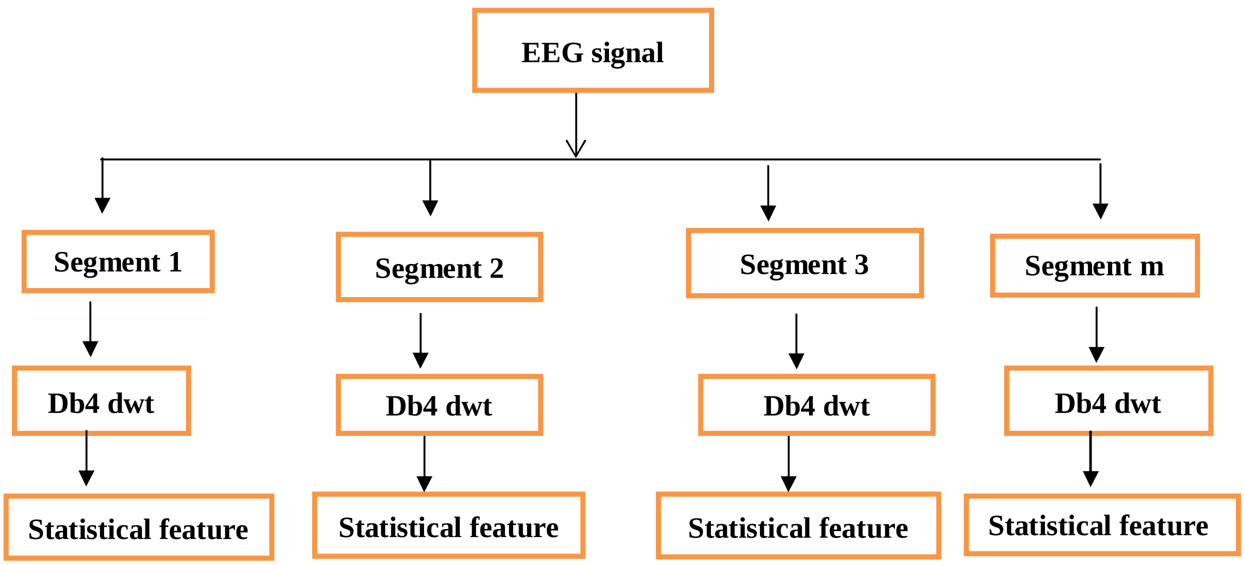

Fig. 4. Segmentation technique diagram for obtaining different segments and statistical features

DOI: 10050086.2022.08.68

ISSN: 1005-0086

Stage1: Determination of segments

Each EEG channel is represented a vector. This vector is segmented into number of intervals based on specific time duration where the number of intervals is m≥1.
In this study we segment signal to 16 segment and size each segment 256.

Stage2: applied wavelet function for each segment

The discrete wavelet transform (DWT) is a spectral analysis method that analyzes non-stationary signals and gives the signals a representation in terms of time and frequency. EEG signals feature non-stationary properties, hence, DWT has been employed extensively for EEG signal analysis. In order to achieve good time-frequency localization, DWT employs long time windows at low frequencies and short time windows at high frequencies. With the use of sequential high-pass and low-pass filtering of the time domain signal, DWT divides a signal into a number of sub-bands. Initial level approximation (A1) and detail coefficients (D1) are the terms used to describe the down sampled signals after passing through the first filters. Then, using the approximation coefficient of the preceding level, the approximation and detail coefficients of the subsequent level are derived. The major frequency components of the signals define the number of breakdown levels. Discrete type Wavelets like the Daubechies Wavelet are quite popular. Multi-resolution decomposition is the name given to this form of decomposition. An orthogonal and asymmetrical Wavelet is this one. It has the most disappearing moments of any phenomenon. This Wavelet comprises coefficients that range from being very localized to being extremely smooth. There are db1 through db10 variations of the Daubechies Wavelet.[12].

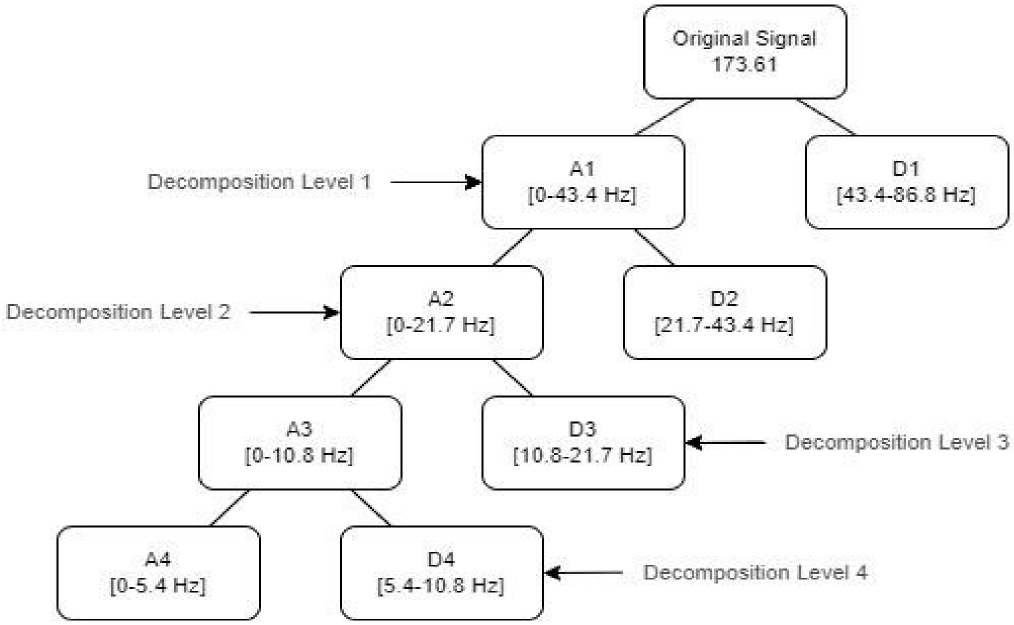

Fig.5. Sub-band decomposition of a signal by using DWT

Stage3: Statistical feature extraction

Because feature extraction removes undesirable information from the source data, choosing strong features enhances classification outcomes. But you must choose the most important elements from the EEG sample. In this study, Standard deviation, mean, min and max statistical characteristics that are collected from each cluster are used to represent EEG data. EEG data can be categorized as symmetric or skewed [13].We discovered that while certain data sets are skewed, EEG data are distributed symmetrically. For these reasons, we examined the symmetric and skewed behavior of EEG data using these characteristics.

DOI: 10050086.2022.08.68

ISSN: 1005-0086

Table 1: The used statistical measures.

1. Minimum value: x=min/**x(n)/**
2. Maximum value: x=max/x(n)/
3. . Standard deviation: x std= $\sqrt{\frac{\sum_{n=1}^{N} x(n)-xm}{N-1}}$
4. Mean xm = $\sum_{n=1}^{N} x(n)N$

Implementation of the Proposed Method

Our proposed method is tested using the epileptic EEG data which contains five sets. In this method, each EEG channel is segmented into intervals and each interval is divided into clusters. With each interval contains 256 data points in 1.468s. The feature are pulled out from each cluster to form a matrix of [16×500]. That represents any two class signals. The matrix is fed to the K-means algorithm.

K-means clustering

One of the most used partitional clustering techniques. It is an unsupervised, nondeterministic partitioning technique. This program divides the supplied data set into various subgroups in accordance with specific criteria. The algorithm's goal is to increase similarity inside a cluster of patterns while decreasing similarity across clusters. It is a quick and reliable clustering technique. The number of clusters and starting centroids selected have a significant impact on how well the K-means clustering performs. For best results, the K value has to be changed throughout a range [11].

The outline of basic K-means algorithm is given below.

a) As the initial centroids, pick K spots.

b) Measure the separation between each data point and the centroid of the cluster.

c) Reassign the data point with the shortest distance to the new cluster.

d) Recomputed the centroids for the new cluster.

e) Continue until the convergence requirements are met

The sum of squared error is the objective function employed by K-means method. The mathematical representation of objective function is given below.

$$S=\sum\sum(\| xi-yj \|)^2$$

|| Xi – Yj ||' is the Euclidean distance between Xi and Yj

Ci' is the total number of pattern points in ith cluster.

'C' is the number of cluster centroids.

Performance Measures

The following metric were used to test the proposed model including accuracy, sensitivity, specificity

$$\text{Accuracy}=\frac{TP+\text{TN}}{\text{TP +FN+TN+FP}};\ \text{Sensitivity}=\frac{TP}{\text{TP+FN}};\ \text{Specificity}=\frac{\text{TN}}{\text{TN+FP}}$$

5. RESULTS AND DISCUSSION

The developed model to EEG signals was tested using epileptic dataset in six cases A–E . Six experiments or cases were designed to detect epileptic EEG seizures. Five experiments were employed to classify a pair of EEG classes (compare just two class) and one experiment for all classes (compare all classes). The EEG cases are sorted out as follows:

Case1: Set A vs. Set E
Case2: Set B vs. Set E
Case3: Set C vs. Set E
Case4: Set D vs. Set E
Case5: Set A vs. Set D
Case6: All classes vs. all classes

DOI: 10050086.2022.08.68

ISSN: 1005-0086

Table2. Experimental results of the proposed method

	Performance metrics		
EEG Case	**Accuracy**	**Sensitivity**	**Specificity**
Case 1	100%	100%	100%
Case 2	100%	100%	100%
Case 3	100%	100%	100%
Case 4	100%	100%	100%
Case 5	100%	100%	100%
Case 6	100%	100%	100%

From the results that has been obtain can see that the proposed method overcome on most of the methods of epileptic diagnosis that existing in this time.

The confusion matrix for the Experiment 1

class	A	B
A	16	0
E	0	16

The confusion matrix for the Experiment 2

class	A	B	C
A	16	0	0
E	0	16	0
C	0	0	16

The confusion matrix for the Experiment 3

class	A	B	C	D	E
A	16	0	0	0	0
B	0	16	0	0	0
C	0	0	16	0	0
D	0	0	0	16	0
E	0	0	0	0	16

Note. The 16 is the vector in every class of the dataset

DOI: 10050086.2022.08.68

Table 3. comparison between our method and other method in the literature for two class.

Research	method	Training data/testing	Accuracy
Reference[9]	**EnsembleEmpirical Mode Decomposition &K-Means Clustering**	**0/500[a]**	**98%**
The Proposed Method	Dwt based combined Statistical feature &**K- Means Clustering**	**0/500[a]**	100%
[a] No training data are used.			

6. CONCLUSION

ThisstudyproposesamodelforEEGdataclassificationbasedonsegmentationtechniquecoupledwithstatistical features and k_ means) classifier. We developed the proposed system to support experts by which the presence or absence of the epilepsy can be detected. Several experiments are designed using publicly EEG dataset. From the results of tests or experiments, and when compared to the results of existing methods, it can be concluded that the proposed method is very effective in determining clustering of data.

REFERENCES

[1] Dabye AS, Issaka M, Gueye L (2015) Localization of epileptic seizure with an approach based on the PSD with an auto-regressive model. arXiv:1506.00947.
[2] Kamath C(2013) A new approach to detect epileptic seizures in Electroencephalograms using Teagerenergy. ISRN Biomedical Engineering, Hindawi. https://doi.org/10.1155/2013/358108.
[3] Kamath C (2013) Teagerenergy-based filter-bank Cepstra in EEG classification for seizure detection using radial basis function neural network. ISRN Biomedical Engineering, Hindawi. https://doi.org/10.1155/2013/498754.
[4] Abbasi R, Esmaeilpour M (2017) Selecting statistical characteristics of brain signals to detect epileptic seizures using discrete wavelet transform and perceptron neural network. IJIMAI4 (5):33–38.
[5] S.Pati, A.V. Alexopoulos, Pharmaco resistant epilepsy: from pathogenesis to current and emerging therapies, Cleve.Clin.J.Med. 77(7)(2010)457–467.
[6] Zhou W, Liu Y, Yuan Q, Li X (2013) Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG. IEEE Trans Biomed Eng 60(12):3375–3381.
[7] JiZ, Sugi T, Goto S, Wang X, Ikeda A (2015) An automatic spike detection system based on elimination of false positives using the large-area context in the scalp EEG. IEEE Trans Biomed Eng 58:2478–2488.
[8] Xia YS, Leung H (2006) Nonlinear spatial-temporal prediction based on optimal fusion. IEEE Trans Neural Netw 17:975–988.
[9] Bizopoulos, P. A., Tsalikakis, D. G., Tzallas, A. T., Koutsouris, D. D., & Fotiadis, D. I. (2013). EEG epileptic seizure detection using K-means clustering and marginal spectrum based on ensemble empirical mode decomposition. 13th IEEE International Conference on Bio Informatics and Bio Engineering, IEEE BIBE 2013, 13–16. https://doi.org/10.1109/BIBE.2013.6701528
[10] Prabhakar, S. K., & Rajaguru, H. (2016). PCA and K-means clustering for classification of epilepsy risk levels from EEG signals - A comparitive study between them. ICIIBMS 2015 - International Conference on Intelligent Informatics and Biomedical Sciences, 83–86. https://doi.org/10.1109/ICIIBMS.2015.7439467
[11] Manjusha, M., &Harikumar, R. (2016). Performance analysis of KNN classifier and K-means clustering for robust classification of epilepsy from EEG signals. Proceedings of the 2016 IEEE International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2016, 2412–2416. https://doi.org/10.1109/WiSPNET.2016.7566575
[12] EEG time series data, Department of Epileptology University of Bonn, http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html.
[13] R.D. De Veaux, P.F. Velleman, D.E. Bock, Stats Data and Models, 3rd ed., Pearson Addison Wesley, Boston, 2008.