Hjorth Parameter based Seizure Diagnosis using Cluster Analysis

Siddhartha Kumar Arjaria¹, Gyanendra Chaubey¹, Nishtha Shukla²

¹Department of Information Technology, Rajkiya Engineering College, Atarra, Banda (India) – 210201
²Department of Computer Science and Engineering, Institute of Engineering and Technology (IET), Lucknow (India) - 226027

arjarias@gmail.com

Abstract. The health-related issues have been increased with a wide range in few years. Hence the need for effective and advanced health care systems or aids is expanding. New methodologies and instruments must be developed to aid the doctors in intelligent health caring of patients. Biomedical signals are a rich source of information, and it is not easy to understand by the normal human beings. To provide ease, extraction and analysis of biomedical signals can help get the correct information to everyone. The signals generated by the brain control the status of the mind and control the action of the whole body. Epilepsy is a disease by which around 50 million people are affected worldwide. Abnormal synchronisation of the neural activity with symptoms like convulsion is the phenomenon of epileptic seizures. An advanced seizure diagnosis system will help in the detection and diagnosis of epileptic seizures. In this paper, clustering algorithms are applied to Electroencephalogram (EEG) data to classify it in normal and epileptic seizures using the Hjorth parameters. After extracting the Hjorth parameters from EEG signals and k-means, basic sequential algorithmic scheme (BSAS), partitioning around medoids (PAM), fuzzy c-means (FCM), and Valley-Seeking clustering algorithms are applied to group it into normal and seizure. With the used dataset, the Valley Seeking clustering algorithm gives the best performance with an accuracy of about 87%.

1. Introduction and related works

Neurological disorders and brain function can be studied widely with rich information as EEG [1]. A tremendous amount of data is generated by EEG systems, and its regular analysis and monitoring are not possible manually [2]. Hence, determining the epileptic and non-epileptic seizures is a tedious task for manually analysing the long-range electroencephalogram signals by doctors. Therefore, many seizure diagnosis systems came into existence. In this work, a method is developed to detect the seizure. First, features are found from EEG signals by calculating the Hjorth parameter, and after that, different clustering algorithms are being applied to the diagnosis of seizure activity in the brain [3]. The clustering algorithm is compared based on classification accuracy and execution time. The application of digital signal processing and soft computing techniques makes doctors’ task easy to diagnose the seizure activity accurately.

Ongoing through several studies, it is found that few works have been done on the clustering with Hjorth parameters based on clustering algorithms for seizure diagnosis using EEG signals. Therefore, few actual results of recent years are discussed below:
In 2016, J. Mohylova et al. used Fuzzy c-means to classify the epileptic pattern and creating a homogenous compact of class segments of EEG [4]. A study on seizure detection by analysing EEG signals using power spectral density, entropy, and Teager energy by N. Sriram et al., published in 2018, was conducted. Descriptive Analytics and Wilcoxon rank-sum test and descriptive analysis is used to confirm the features' suitability. Finally, multilayer perceptron neural networks are used to classify the signals [5]. In 2018, Hemant Choubey and Alpana Pandey introduced a Masking and Checkin-based feature extraction technique to corkscrew the attributes from EEG signals and then used K-means and KNN as classification algorithms for segregation of normal and abnormal signals. They have compared their results with some previous works to show outcome enhancement [6].

In 2019, Ashok Sharmila and Purshotaman Geetanjali surveyed pattern discovery methods for epilepsy seizure diagnosis from EEG signals. They have found that seizure detection methods vary from different EEG signals to other conditions. They have also found that EEGs taken in various states have different characteristics [7]. A study on Epileptic Seizure detection using Hidden Markov Model using a probability-based method was presented by Deba Prasad Das and Mahesh Kumar H. Kolekar. Discriminant correlation analysis is used to differentiate the signals with Fuzzy c-means clustering. They have achieved the highest accuracy of 98.57% [8].

In 2020, Qianyi Zhan and Weihu presented epilepsy detection using multi-view clustering using in-depth features. Deep CNNs are used to extract the features. They have used FCM to do the clustering with the three-step method. This work showed a satisfactory result in an analysis of seizures [9]. Ozlem Karabiber Cura et al., in 2020, present a classification method for epileptic seizures using the technique and its derivatives of empirical mode decomposition. They have used four different features as energy, power spectral distance, statistical significance, and correlation measures on various supervised learning algorithms to classify the signals and got a highest of 97% accuracy [10].

Based on this study, it is observed that fewer works have been done on the Hjorth parameters-based clustering of EEG signals. Therefore, four prominent unsupervised clustering algorithms as k-means, FCM, PAM, and BSAS, are used to analyse the results in EEG signal classification. The rest of the paper is divided into five different parts. Section 2 contains the dataset description in detail. Section 3 is all about the research methodology used to classify signals. In section 4, results have been discussed, and conclusions are drawn in section 5 with the references given at the end.

2. Dataset Description:
The data described in Andrzejak et al. (2001) [11] that is published is used. The dataset link is given in the reference section below [12]. This dataset has five separate sets (A - E), each containing 100 single-channel electroencephalogram signals with a duration of 23.6s. The single-channel data was extracted from the multichannel signals by visual interaction of eyes or muscle movements. Five healthy and fit volunteers were carried out to take surface EEG recordings of set A and B using a standardised electrode placement scheme. Using the pre-surgical diagnosis, sets C, D, and E are achieved. The EEG of five healthy volunteers who have gained complete seizure control is considered an epileptogenic zone after the resection of the hippocampal formation. Therefore, signals C is determined from the Hippocampal formation, while signals D were recorded with an epileptogenic zone. Seizure-free intervals are recorded in sets C and D, while seizure activity is recorded in set E. Only set A and set E are considered in this study, as shown in figure 1.
Figure 1(a). Normal EEG (set A)

Figure 1(b). Seizure EEG (set E)

Figure 1. Signal of Normal and Seizure EEG

3. Methodology:
In the given approach of figure 2, EEG signals (data) are first collected. Only two sets of signal A and E, corresponding to the standard and seizure dataset consisting of 100 samples of both groups, are taken. Next, the Hjorth parameters of both sets are extracted. It has three parameters activity, complexity, and mobility, where the activity and mobility are chosen for clustering purposes. Hjorth parameters are calculated in the time domain. The parameters are extracted and then collected parameters data into a matrix. Now the data matrix is of the size 2*200 containing the activity and complexity features of the EEG signal. After collecting the data, clustering algorithms K-means, FCM, PAM, BSAS, Valleyseeking clustering approach is applied. The achievement of the clustering algorithm is analyzed by accuracy and time [13].

The Hjorth parameter analysis and clustering algorithms are explained in detail in the below subheadings:

3.1. Hjorth Parameters
It is one of the accustomed statistical indicators of those properties implemented in signal processing within a time realm [14]. Following are the types of Hjorths parameters:
3.1.1. Activity. The activity parameter is the parameter that indicates the signal strength and the variance of a time function. It can show the surface of the power spectrum in the frequency domain. The following equation (1) represents the activity:

$$\text{Activity} = \text{var}(x(t))$$

Where $x(t)$ reflects the signal.

3.1.2. Mobility. The mean frequency or the standard deviation proportion of spectrum power is represented by the mobility parameter [15]. Mobility is defined as the ratio of the square root of the variance of the first derivative of signal $x(t)$ to the variance of signal $x(t)$ and is presented in equation (2):

$$\text{Mobility} = \sqrt{\text{var}(dx(t)/dt)} \cdot \text{var}(x(t))^{-1}$$

3.1.3. Complexity. The complexity parameter represents the variation in frequency. The ratio of mobility of the first derivative of signal to signal mobility is called complexity. Given in equation (3):

$$\text{Complexity} = (\sigma_1^2 / \sigma_2^2)^{-1} (\sigma_1^2 / \sigma_2^2)^{-1}$$

So these are the normalised slope descriptor (NDS) parameters that have their application in EEG.

3.2. K-means clustering algorithm

The most often used vector quantisation approach for signal processing clustering algorithms is K-Means. Its goal is to divide $n$ instances into $k$ clusters, with each instance belonging to the cluster with the closest mean (cluster centroid) [16].

K-means algorithm can be given in mathematical equations (4):

Given a set of $m$ observations, $D = \{x_1, x_2, \ldots, x_i, \ldots, x_m\}$, where each observation is an $n$-dimensional vector as $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$,

$$C_j = \text{Cluster}(X) = \arg \min_{\mu_j} \sum_{i} (x_i - \mu_j)^2$$

(within cluster sum of squares)

$$J = \sum_{i=1}^{m} \sum_{j=1}^{k} v_{ij} \| x_i - \theta_j \|^2$$

(4)

K-means stops itself when there is movement in the cluster centres, and for this, it minimises intra-cluster distance, not the Euclidean distance. Hence this can be considered as the most difficult Weber problem.

K-means algorithm minimises the cost function as in equation (5)

$$f(\theta, U) = \sum_{i=1}^{N} \sum_{j=1}^{m} v_{ij} \| x_i - \theta_j \|^2$$

(5)

Where $\theta = [0, \ldots, 0 m^T]$, $\| \|_2$ is Euclidean distance, and $V_{ij} = 1$ if $x_i$ lies nearest to $\theta_j$ or 0 otherwise. In other words, this results in minimising the intra-cluster distance or the distance between each data vector [17]. Therefore, when the data vectors of $X$ form $m$ compact groups (with no substantial differences in size), if $m$ is known, and each $\theta_j$s (nearly) placed in the centre of each cluster, it is expected that $J$ is minimised. It is not necessarily followed in the cases when:

(a) Compact clusters are not formed by data vectors.

(b) Sizes will not differ drastically.

(c) Cluster number ($m$) has not been given correctly.

K-means is an algorithm that works on the principle of moving the points to the nearest cluster centre. Convergence of K-means can be faster with the use of some other distance calculation methods. Spherical K-means, PAM, are the modification of the K-means algorithms.

3.3. Partitioning Around Medoids

The Partitioning Around Medoids (PAM) algorithm is also called the k-medoid algorithm, is a classical machine learning method of dividing the data into instances into $k$ clusters, where the $k$ clusters
should be prior known. Like the k-means algorithm, it is also a clustering algorithm. The objective of \( k \)-Means is to reduce the total squared error. In contrast, the goal of \( k \)-medoids is to reduce the sum of dissimilarities between the observations labelled to be in a cluster and a point designated as the centre of that cluster. The PAM algorithm, unlike the k-means algorithm, selects the data points as centres (medoids). The set of the \( X \) vectors that describe the clustering body (called medoids) is obtained by minimising cost function \( J(\theta) \). It can be calculated as overall data vectors, distance from each vector, and its nearest medoid [18].

The realization of the K-medoid algorithms can be done with the Partitioning Around Medoid (PAM) algorithm and can be understood from the following steps:

1. Step1 (Initialisation): K-medoids are initialised randomly from \( n \) data points.
2. Step2 (Assignment): The nearest medoids are allocated to each of the data points.
3. Step 3 (Updating): For each data point with medoid and each medoid, do the swapping between medoid and each associated data point and calculate the cost that is average dissimilarity in between all the data points and its associated medoids. Choose the medoid with the lowest price.
4. Repeat steps2 and 3 until there is no change.

So the algorithm is iterative. The runtime complexity of the original PAM algorithm per iteration of (3) is \( O(k(n-k)(n-k)) \), by only computing the change in cost [19].

3.4. Fuzzy \( c \)-means Algorithms

A parameter vector \( \theta_j \) represents each (compact) cluster in the Fuzzy \( c \)-means (FCM) algorithm, where \( j \) ranges from \( 1 \) to \( m \). Furthermore, it is assumed that a vector \( x_i \) belonging to the data set \( X \) does not essentially belong to a single cluster \( C_j \). Instead, it can, to some extent, belong to numerous clusters at the same time. The variable \( x_i \) evaluates \( x_i \)'s "grade of membership" in the cluster \( C_j \) such that \( u_{ij} \in [0,1] \), and

\[
\sum_{j=1}^{m} u_{ij} = 1
\]

for all \( x_i \).

Again, the number of clusters, \( m \), is believed to be known. FCM seeks to direct each of the \( m \) available dimensional parameter vectors (representatives) \( \theta_j, j = 1, \ldots, m \) to the data-intensive part of the data space. Finally, the algorithm includes an additional parameter \( q (>1) \), called the fuzzifier [20]. FCM is a well-known algorithm. It is iterative, starting estimations, \( \theta_1(0), \ldots, \theta_m(0) \), for \( \theta_1, \ldots, \theta_m \), respectively, and at each iteration \( t \):

- The (squared Euclidean) distances between \( x_i \) from all \( \theta_j \)'s, \( j = 1, \ldots, m \) are used to compute the grade of membership, \( u_{ij}(t-1) \), of the data vector \( x_i \) in the group \( C_j \), \( i = 1, \ldots, N, j = 1, \ldots, m \).
- The weighted means of all data vectors are modified in the representative's \( \theta_j \)'s (each data vector \( x_i \) is weighted by \( u_{ij}^0 (t - 1) \)).

3.5. Basic Sequential Algorithmic Scheme

On the given data set, the BSAS algorithm operates a single pass. Furthermore, the mean of the vectors associated with each cluster represents it [21]. The BSAS working procedure is outlined below. For each new vector \( x \) supplied to the algorithm, the distance from the already created clusters is determined first. If the following criteria are not met, a new cluster with \( x \) is created:

1. If the distances are more significant than a (user-defined) threshold of dissimilarity, \( \Theta \).
2. If the maximum number of clusters allowable, \( q \) has not been reached. Otherwise, \( x \) is assigned to its cluster nearest to it, and the representative for that cluster is updated.

When all data vectors are considered once, the algorithm ends. Therefore, BSAS is suitable for sorting out compact clusters [22]. Furthermore, BSAS is fast, and the reason behind this is that it requires one pass through the data set, \( X \); thus, it is a strong option for processing large data sets. However, the clustering that results in several cases may be of poor quality. Improvements can be achieved in a refinement step.
3.6. Valley-Seeking Clustering Algorithm

Clusters are defined as the probability distribution function \( p(x) \) peaks, denoting \( X \), separated by valleys in this method (also called VS). This method does not use representation, but the local area \( V(x) \) around each data vector \( x \in X \). The latter is interpreted as the set of vectors in \( X \) (excluding \( x \)) at a distance less than \( 'a' \) given distance from \( 'x' \) where \( 'a' \) is a user-defined Parameter. Euclidean squares can be used (other distances can also be applicable) for distance measurement. In addition, VS requires an (overestimated) number of clusters, \( m \). This algorithm is sensitive to the initial allocation of the data vectors to clusters and the choice of \( 'a' \).

Iterative in nature, the algorithm begins with an initial assignment of \( X \) vectors to \( m \) clusters. Then, all data vectors are presented only once at each epoch (\( N \) successive iterations). The region \( V(x_i) \) is determined during the \( i^{th} \) epoch and the cluster to which major of the data vectors in \( V(x_i) \) belong is recognised and stored for each \( x_i \) in \( X \), \( i=1,\ldots,N \). When all the data vectors have been presented (during the \( t_{th} \) epoch), re-clustering happens, and each \( x_i \) is now assigned to the cluster with the most significant number of points in \( V(xi) \). The algorithm finishes once no re-clustering occurs between two successive epochs [23,24].

4. Experimental Results

The methodology starts with collecting EEG data, which is then used to determine the Hjorth parameters, activity, complexity, and mobility of sets A and E. Table 1 shows the sample of feature one from each class (A & E).

| Feature | dataset  | Set-A(normal) | Set-E(seizure) |
|---------|----------|---------------|----------------|
| Activity |          | 43.133        | 480.8          |
| Mobility |          | 0.65248       | 0.48943        |

Figure 3 shows the distribution of data in 2d space indicates the Hjorth parameters activity and mobility of class A & E data having 100 samples of each class.

![Figure 3. Data set of 200 elements](image)

After collecting the features from set A and set E, we apply the clustering algorithms one by one and see the classification accuracy and time. Table 2 indicates the classification accuracy of all the clustering algorithms used in the papers.
The result in table 2 shows the performance of different clustering algorithms. For example, K-means and PAM clustering algorithms have the same clustering capabilities regarding accuracy and execution time in both these algorithms. Only one mismatch was found in the set A cluster, and 43 mismatches were found in the set E cluster. Therefore, the overall accuracy of both the algorithm is 78%, and the time required is the same 0.0470 seconds.

FCM clustering algorithm is better in terms of accuracy from K-means. For example, PAM and BSAS having an accuracy of 80%, one mismatch in class A and 39 mismatches in class E, but it takes 0.1041 sec execution time.

BSAS clustering algorithm has one mismatch in class A and 41 mismatches in class E. The overall accuracy is about 79% and takes an execution time of 0.1892 sec.

The valley-seeking clustering algorithm gives the best performance in classification accuracy when applied to the data set. This algorithm indicates only one mismatch in class A and 25 mismatches in class E. so overall accuracy of 87%. Time execution time is 0.3060 sec.

Figure 4 shows the cluster structure after executing different clustering algorithms on the same data set.

| S.No | Clustering Algorithm | Mismatch in set A (out of 100) | Mismatch in set E (out of 100) | Total accuracy | Execution time |
|------|----------------------|---------------------------------|-------------------------------|----------------|---------------|
| 1    | K-means              | 1                               | 43                            | 78%            | 0.0470        |
| 2    | PAM                  | 1                               | 43                            | 78%            | 0.0470        |
| 3    | FCM                  | 1                               | 39                            | 80%            | 0.1041        |
| 4    | BSAS                 | 1                               | 41                            | 79%            | 0.1892        |
| 5    | Valley seeking       | 1                               | 25                            | 87%            | 0.3060        |

Figure 4(a). The signal output of K-means and PAM algorithm.

Figure 4(b). Clustering output of K-means and PAM algorithm.
5. Conclusion
Diagnosis of seizure activity is an integral part of its treatment. Therefore, correct diagnosis of seizure activity without human intervention with the highest degree of accuracy in the least time is the motive of this paper. In this paper, various clustering algorithms are applied to Hjorth parameters of EEG.
signals. The different clustering algorithm gives lower diagnosis accuracy, and the valley seeking clustering algorithm achieves the highest diagnosis accuracy about 87%. The various outcomes of time and clustering accuracy are shown in table 2. In this work, various other algorithms can be used to check the accuracy of the algorithms. The given accuracy can also be increased with balanced data and using the algorithms' hyperparameters tuning.

References
[1] Claude Robert et al., 2002, Clin Neurophysiol, "Electroencephalogram processing using Neural Networks," 113(5):694-701
[2] Elif Derya Ubeli, 2009, Expert Systems, "Probabilistic neural networks combined with wavelet coefficients for the analysis of electroencephalogram signals," Vol. 26, Issue 2, p. 147-159
[3] Athar A. Ein Shoka et al., 2021, Brain Informatics, "Automated Seizure Diagnosis System based on Feature Extraction and channel selection using EEG Signals," 8(1):1.
[4] J. Mohyllova et al., 2016, Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITT'16), Advances in Intelligent Systems and Computing 450, "Fuzzy c-means algorithm in Automatic classification of EEG." Vol. 450
[5] N. Sriram et al., (2018), Brain Informatics, "Automated epileptic seizures detection using multi-features and multilayer perceptron neural networks," Vol. 5, Article No. 10
[6] Hemant Choubey and Alpana Pandey, 2018, Multidimensional Systems and Signal Processing, "A new feature extraction and classification mechanisms for EEG signal processing," 30:1792-1809
[7] Ashok Sharmila and Purshottam Geetanjali, 2019, Biomed. Tech. "A review on the pattern detection methods for epilepsy seizure detection from EEG signals," 64(5):507-517
[8] Deba Prasad Das and Mahesh Kumar H. Kolekar, 2019, SIRS 2018, "Probability-Based Approach for Epileptic Seizure Detection Using Hidden Markov Model," CCIS 968, pp. 268–281
[9] Qianyi Zhan and Wei Hu, 2020, Computational and Mathematical Methods in Medicine, "An Epilepsy Detection Method Using Multiview Clustering Algorithm and Deep Features," Volume 2020, Article ID 5128729, 11 pages
[10] Ozlem Karabiber Cura et al., 2020, BioMedical Engineering Online, "Epileptic Seizure Classification using empirical mode decomposition and its derivative," Vol. 19, Article No. 10
[11] R G Andrzejak et al., 2001, Physical review. E, Statistical, nonlinear and soft matter physics, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state," 64:061970
[12] http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html
[13] Xu, D., Tian, Y. (2015), Ann. Data. Sci, "A Comprehensive Survey of Clustering Algorithms" 2, 165–193
[14] Bhattacharyya, S., Mukul, M.K., (2018), J Ambient Intell Human Comput, "Reactive frequency band based movement imagery classification."
[15] Neelam Rup Prakash et al., 2020, Journal of Healthcare Engineering, "Variations in Electro cortical Activity due to Surgical Incision in Anaesthetized Cardiac Patients: Electroencephalogram-Based Quantitative Analysis," Vol. 2020, Pg. 11
[16] Tsipouras, M.G., 2019, EURASIP J. Adv. Signal Process, "Spectral information of EEG signals concerning epilepsy classification" 10
[17] Chiang, M.M.T., Mirkin, B. (2010). Journal of Classification, Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. 27, 3–40
[18] O. Smart, M. Chen, 2015, In 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), "Semi-automated patient-specific scalp EEG seizure detection with unsupervised machine learning," pp. 1–7.
[19] Jin X., Han J. (2017), Springer, Boston, MA, K-Medoids Clustering. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining.
[20] Md. Abu Bakr Siddiqui et al., 2018, *Preprints,* "Implementation of Fuzzy c-means and probabilistic c-means Clustering Algorithms, Cluster Tendency Analysis and Cluster Validation," 2018110581

[21] Eynard D., Javarone M.A., Matteucci M, (2018), In Alhajj R., Rokne J. (eds) Encyclopedia of Social Network Analysis and Mining. Springer, New York, NY. Clustering Algorithms.

[22] Yeganegi, H., Salami, P. &Daliri, M.R. 2020, Cognitive Computation, A Template-Based Sequential Algorithm for Online Clustering of Spikes in Extracellular Recordings. 12, 542–552.

[23] Mayank Singh and Viranjay M. Srivastava (2019). Implementing Architecture of Fog Computing for Healthcare Systems based on IoT. International Journal of Engineering and Advanced Technology (IJEAT), vol. 8, no. 4c, pp. 23-27.

[24] Zhang, C., Zhang, X., Zhang, M. Q., & Li, Y. (2007). Pattern Recognition Letters, Neighbor number, valley seeking and clustering., 28(2), 173–180.