Epileptic Seizure Diagnosis Using EEG Signals

Shahad Saad Alwain[1]; Naji Mutar Sahib[2]

[1]Department of Computer Sciences, College of Science, Diyala University, Iraq
[2]Department of Computer Sciences, College of Science, Diyala University, Iraq
[1] wesaladeeb2@gmail.com; [2] al.sahib@sciences.uodiyala.edu.iq

Abstract— To measure and track human performance and physiological state, a great number of biomedical data such as the electroencephalographic (EEG) signal is collected from the human body every day. For research and for medical diagnosis and treatment, understanding these signals is critical. There are many research and development projects Within this field, but there are weaknesses point, including their reliance on specific registration technologies, low accuracy and timing of implementations. In This paper a computer-aided system are used for diagnosing epileptic seizures more accurately using electronic imaging data. The framework relies on identifying (EEG) signals for use in linear and non-linear applications. This study was conducted using six algorithms of machine learning algorithms which are logistic regression (LR), support vector machine (SVM), k-Nearest Neighbors (KNN), Gaussian Naive Bayes, Artificial Neural Networks (ANN), Principal Component Analysis(PCA). Also, improve surrogate data technique are used for feature extractions that depend on Fourier transforms (FT). It was concluded that the classifier (SVM) works better and showed high accuracy in the classification of data taken for all epilepsy cases.
Keywords— Electroencephalograms (EEG), epilepsy, machine learning techniques (MLT).

I. INTRODUCTION

Because of the increase in diseases in recent years as a result of wars and their effects, it has contributed to the increase and emergence of many new and complex physical and psychological diseases. One of the most common diseases is epilepsy[1]. In 1875, the English physician, Richard Canton, discovered electrical currents in the brain responsible for elliptic seizure [2]. According to the latest study, more than 2% of the population worldwide is affected from epilepsy where 85% of those live in developing countries and has adverse effects on their daily life and productivity[3].Epilepsy is a critical neurological brain disorder originating from temporary abnormal discharges of the brain electrical activity, leading to uncontrollable movements and trembling[4]. resulting in altered behaviors , such as losing consciousness, jerky movements, temporary loss of breath and memory loss[5].There are many injuries that can lead to epilepsy, the most important of which are (Genetic injury, head injury, brain tumor, birth trauma, ext …). The general detection of epileptic activity requires manual scanning of EEG recordings, which commonly takes several days to complete[2]. Visual inspection for discriminating EEG signals is time-consuming, imprecise, and costly, especially in the case of long-term recordings [4]. CAD (computer-aided diagnosis) play an important role to develop robust and reliable techniques for epileptic activity detection in EEG signals [6]. EEG is collected by placing electrodes on the scalp of the patient. One can then, record the electro-activity that the brain produce along the scalp [16]. Computer-aided diagnosis (CAD) systems in the medical field can be viewed as cutting-edge expert and intelligence systems in

the interface of medicine and computer science. CAD systems in medicine may use diagnostic rules to emulate the way a skilled human expert makes diagnostic decisions[7]. Nowadays, many researchers have used different techniques and methods such as using time-frequency analysis [8], naive bayes and K-NN classifiers [7]**,** multi-domain feature extraction and nonlinear analysis [3], a multi-context learning approach[12], logistic model trees [9], Innovative genetic programming [6], EEG signals [13], biomedical signals [4], deep learning approach [15], a computer aided analysis scheme [7], by which people with epilepsy are detected. These method are completely different from the way that has been used in the past, where epileptic seizures are diagnosed by clinical observation and recent changes recorded[8]. In the clinical contexts, the main diagnosis of EEG is to discover abnormalities of brain activities referred to the epileptic seizure. A seizure occurs when the neurons generate uncoordinated electrical discharges that spread throughout the brain and epilepsy is a recurrent seizure disorder caused by abnormal electrical discharges from brain cells, often in the cerebral cortex [9]. In 1929, Hans Berg discovered the possibility of recording electrical impulses in the human brain[5],these record are called (EEG) this test is commonly used to diagnose epilepsy[10]. There are more than one classifications techniques are used such as(logistic regression, SVM, K-NN, Gaussian Naïve Bayes, ANN, PCA) .the Classification is the process of finding a model that describes and distinguished data classes.

II. RELATED WORKS

Zakarey aLasefr, et al, 2017 [19]: suggested Efficient Automated Technique for Epilepsy Seizure Detection Using EEG signals .They suggest using three classification algorithms (KNN, ANN, and SVM) to identify epilepsy by relying on EEG signals. They also process and apply filters to these signals to remove unwanted noise. Ye Yuan**,** ext al, 2017 [24]: proposed a multi-context learning approach for EEG epileptic seizure detection .They generated EEG sequences from EEG records by using wave transduction to describe the frequency content over time. The experimental results indicate that the representative context features can be obtained from the suggested model from a number of viewpoints and thus enhance the efficiency of EEG seizure detection. Lina Wang , ext al, 2017[2] proposed Automatic Epileptic Seizure Detection in EEG Signals Using Multi-Domain Feature Extraction and Nonlinear Analysis .In there paper, a multi-domain feature extraction method and nonlinear analysis of EEG signals were used. The wavelet threshold method was also used to extract the nonlinear pickling features. Relying on this method. The experimental findings revealed that the suggested epileptic seizure detection approaches achieved a high point, suggesting that this approach is successful in the diagnosis of epilepsy. Yash Paul, et al, 2018 [3] proposed various epileptic seizure detection techniques using biomedical signals presents a novel analysis system for detecting epileptic seizure from EEG signals .The OAT method and the LMT method were used to detect epileptic seizures, based on EEG signals. The OAT method are used for identifying the analogue EEG signals and a LMT method are used for extracting statistical features from the EEG signals was used to detect epilepsy. These methods showed a good percentage compared with modern methods of detecting epileptic seizures.

III.MATERIALIZED VIEWS

The key challenge of any automated epileptic seizure detection method is the extraction of the distinguishing features from EEG signals as it significantly affects the performance of the classifier. In this paper we will use the EEG signals to detect epileptic seizures, this test is one of the most important modern tests to obtain high results in diagnosing people with epilepsy. Also, the database used in this research is available to the public and is divided into groups(A-E) Each group contains a number of different people It also contains people with epilepsy in various cases. there are many techniques have been developed for EEG classifications in different fields such as (logistic regression ,K-NN, ANN,SVM, PCA, Gaussian naïve bayes, ext…), an overview of classification methods in these six methods is provided., samples of EEG signals are shown in Figure 1.
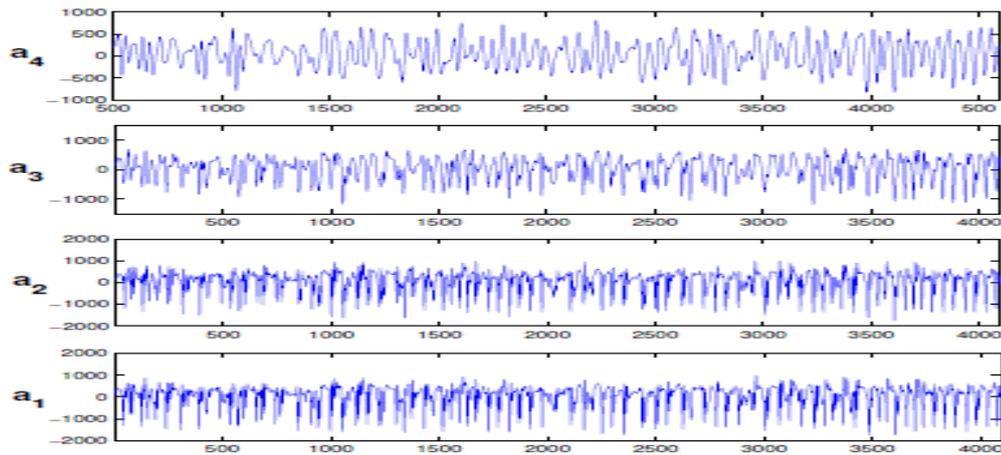
Fig1 : Sample of EEG signal.

IV.PROPOSED METHODS

The proposal model can be divided into six parts (EEG Dataset, Feature Extraction, Feature Matrix, Preprocessing, Classification Technique).as show in figure 2.
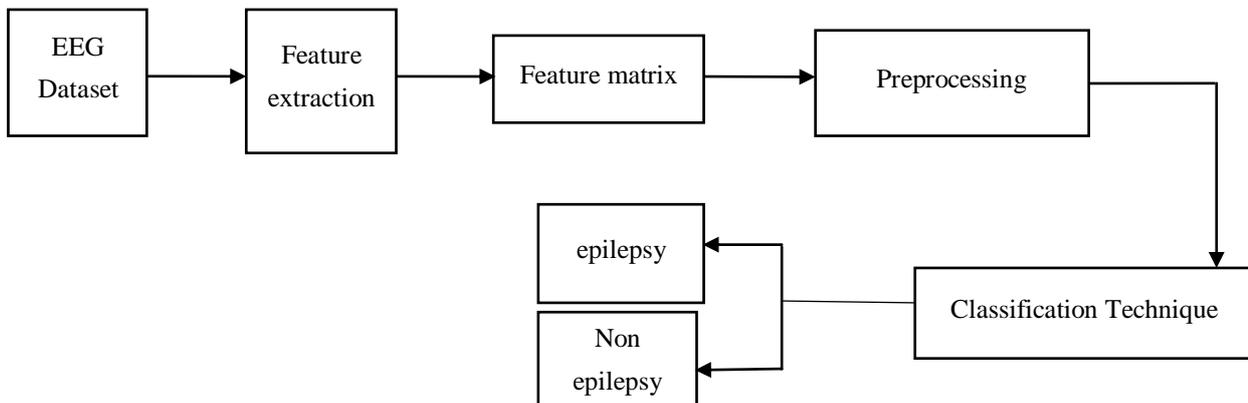

Fig.2: Proposed method.

4.1 EEG Dataset

The dataset used in this study is open to the public. The research dataset consisted of 5 separate directories, each of 100 items, each item is a single subject / person file. The brain activity is recorded for 23.6 seconds for each item. The data collection is brain EEG data for 500 people, where 100 with epileptics and 400 are healthy. PCI-MIO 16E Data Acquisition card system are used for the recording of EEG signals this system provide a real-time process, also it required a Lab View software, and a personal computer. The EEG signals transition of the electrode positioning method from Analog to Digital using International 10-20electrode placement system.

4.2 Feature Extraction

Feature extraction of the signal are used to remove unwanted signal components. The surrogate data technique are used for EEG feature extraction. Surrogate analysis of data is a method to test the existence of non - linearity in a time sequence. It have the same linear statistical properties as the original data, such as amplitude distribution, mean , variance and power spectrum. Each column is a feature for a specific model, where feature are 4096 samples of surrogate data. Surrogate data aims to produce new data that have the same power spectrum and distribution as a given time series. In this case, the time series is the brain EEG captured signal.

4.3 Feature Matrix

The process of feature extracting that extracted from the previous process is a matrix of the rank 500 * 178 where the 500 represents the total number of people used as dataset and the 178 is the point file value. These values were calculated for each person within one second and since the data recording process was for 23.6 seconds per person This means that for each person there are 23 identical matrices.

4.4 Preprocessing

After performing the feature extraction using the surrogate method and transferring the data to the values in this stage, we will perform the preprocessing on these values, the preprocessing include a (checking missing data, feature scaling , feature analysis). After the data extracted the preprocessing operation applied on the data the first step in the preprocessing is:

a- Checking missing data: to converts the data into a clean dataset that can be easily handled and performed many operations on it. The linear time series analysis(Linear Interpolation) will be use to deal with missing data, it is characterized by ease and speed of implementation. This method will used Mean (X_{mean}) ,medium (X_{medium}) ,mod (X_{mod}) the mean is a very basic imputation method, it is the only tested function that takes no advantage of the time series characteristics or relationship between the variables, It is very fast.

b- Feature scaling : are preserved on the data, it used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing steps. Scaling the data distribution to normal standard Gaussian was applied to the samples before they were passed to Training samples to optimize the functions for Classified.

c- Feature analysis : for analysis the data the (EDA) will be used exploratory data analysis (EDA) is a technique for the analysis of collecting data, to describe their basic characteristics .EDA is mainly concerned with seeing what the data may teach us outside the standard of modeling or hypothesis testing process.

4.5 Classification Techniques

After the completion of the preprocessing stage and obtaining the clean data, we will classify these data, by using six methods which are (LR,SVM ,PVC, ANN,K-NN, and Gaussian naïve beyas). We will classify the data.

4.5.1 Logistic Regression Algorithm

Logistic Regression is the method for binary classification. It gives a discrete binary outcome between 0 and 1. In this thesis we will use a Logistic Regression algorithm for detection epileptic seizure, it takes data set signals as an input and should tell if a patient has epilepsy (1) or not (0). LR consist on the flowing components as show in figure (3).
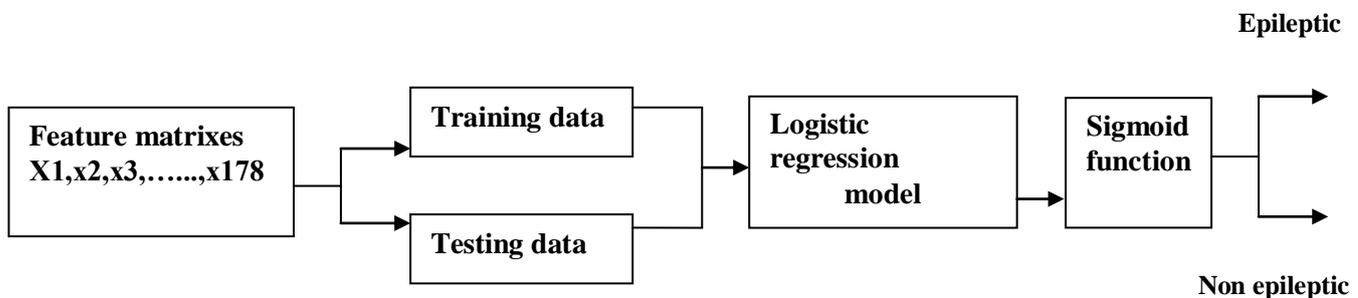

Fig 3:logistic regression component.

4.5.2 Support Vector Machine (SVM)

SVM is the second classification technique in the proposed system .A statistical learning algorithm that classifies the samples using a subset of training samples called support vectors. Considered as a powerful tool of real data classification. Considered a powerful tool for classifying real data. This is known to be a biased classifier that locates a dividing hyperplane between the two categorized classes, which defines the largest minimum distance between such classes.

4.5.3 k-Nearest Neighbors

K-nearest neighbor (KNN) classifier is in the first group, where whole data is divided into training data and testing data point . Distance from all training data to testing data is evaluated, candidate the point with the lowest distance is called the nearest neighbor. The K-NN algorithm structure are show in figure (4).
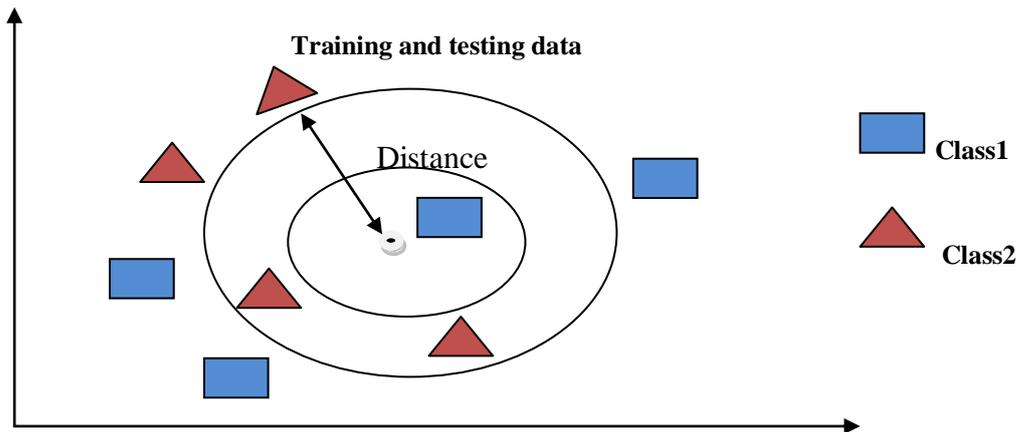
Fig 4 K-NN algorithm.

4.5.4 Gaussian Naive Bayes

It is based on Bayes theory and assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. . In figure (5) the Gaussian Naive Bayes dell with the data was clarify .
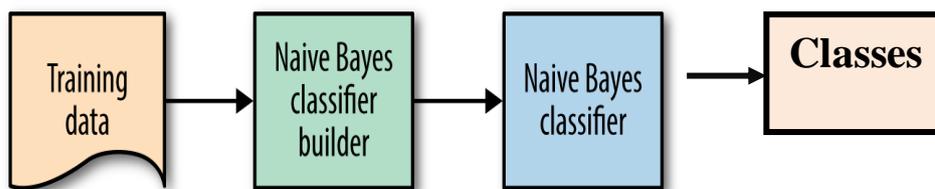

Fig 5: Gaussian Naive Bayes.

4.5.5 Artificial Neural Networks(ANN)

 It is the five classification technique in the proposed system Neural networks are typically organized in layers. A number of nodes, which include a activation function are built in layers. Patterns are shown to the network via the 'input layer' which passes to one or more external(hidden) layer , where the definite action is made through a weighted connections method. The hidden layers are then connected to an exit layer where the answer is executed.

4.5.6 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a method to minimize the measurements (dimension) to a small number that still includes much details in a large collection of variables.

V. The Result

In this section we will performs the preprocessing results of the selected dataset

1. Checking Missing Data Results

After entering the dataset we will check for any missing data in the dataset and ensuring that there is no data are lost in table (1)it represent the checking of the missing data where (X) is the feature and (0) mean that there is no value are missing .

Table(1) Checking Missing Data

unnamed	0
X1	0
X2	0
X3	0
X4	0
X5	0
X6	0
X7	0
X8	0
X9	0
X10	0
:	
X175	0
X176	0
X177	0
X178	0
Y	0

2. Feature Analysis Using (EDA) Results

We use exploratory data analysis to analyzing datasets to summarize their main characteristics, often with visual methods. After analyzing the data was represented as a sample of these data as show in figure (6).
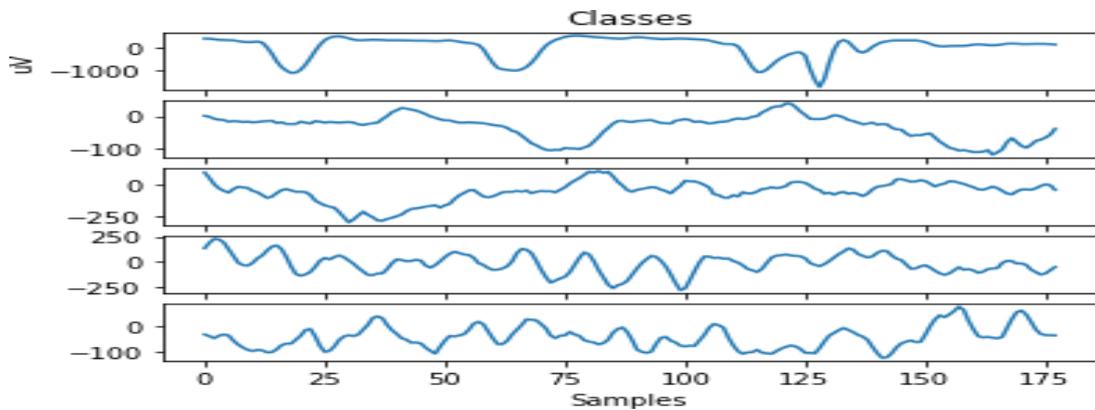

Fig 6: Representing the selected samples from the dataset .

3. Training And Testing Data

After we split the data into training and testing, we randomly chose the sample the go to training data and testing data. The data has been divided to training and testing as show in the table 2.

Table (2): Training And Testing Data.

	Training data	Testing data	total
Epileptic	80	20	100
Non Epileptic	320	80	400
Total	400	100	500

VI. Classification Results

In our work we are provide statistical analysis of the six Machine Learning classification methods used to identify the epilepsy seizure.

Table (3) algorithm results

Classification Methods	ACC	Execution Time
Logistic regression	82.23%	0.427051s
SVM	98.18%	2.062404s
KNN	82.23%	4.286700s
Gaussian naïve bayes	95.47%	0.032489s
ANN	95.47%	87.616128s
PCA	88.0%	4.286700s

Comparison with research and previous studies

As we can see from Table 4, we have compared the performance and the results of our proposed method to a number of the existing methods that were carefully studied in the literature review. It is clear from the table that our method has outperformed the existing methods in the literature.

	Author	Title	Methods	ACC
a	Zakareya Lasefr,[19]	An Efficient Automated Technique for Epilepsy Seizure Detection Using EEG signals	SVM	95%
	Proposed methods		SVM	98.18%
b	R.Panda and P.S. Khobragade [20]	Classification of EEG Signal Using Wavelet Transform and Support Vector Machine for Epileptic Seizure Diction	SVM	91%
	Proposed methods		SVM	98.18%
c	Zakareya Lasefr et al [13]	Epilepsy Seizure Detection Using EEG signals	SVM	96.2%
	Proposed methods		SVM	98.18%
d	E. Juarez-Guerra V. Alarcon-Aquino, P. Gomez-	Epilepsy Seizure Detection in EEG Signals Using Wavelet Transforms and Neural Networks	ANN	93.23%

	Gil [21]			
	Proposed method		**ANN**	**95.47%**
e	**Mohammad Fani Ghasem Azemi[23]**	**Automatic Epilepsy Detection Using the Instantaneous Frequency and Sub-Band Energies of the EEG Signals**	**ANN**	**94%**
	Proposed methods		**ANN**	**95.47%**
f	**G.R.Kiranmayi V.Udayashankara [22]**	**Neural Network Classifier for the Detection of Epilepsy.**	**ANN**	**81.71%**
	Proposed methods		**ANN**	**95.47%**

VII.   Conclusions

Several conclusions have been deduced from the obtained test results. In the following some of these conclusions are listed: The SVM algorithm has proven to be the appropriate algorithm to work on classification of large data for comprehensive EEG signals for all registrations and all conditions of people with epilepsy. Use an improve surrogate data method to analyze data and features extractions from EEG signals, that have proven their efficiency by obtaining high accuracy and less time compared to research and other studies that have been compared with them. They use the same dataset and the same classifier (SVM). Where it requires recording large amounts of data and to deal with data we need to record and classify it. It was concluded that epilepsy varies from one person to another, as it is diagnosed with accustomed to identifying the affected areas in the brain, and accordingly. For the purpose of reaching high-accuracy results by classification, the base of the pant should be very large and includes records for different times and conditions. When using linear and non-linear algorithms it was concluded that non-linear algorithms gave better results in dealing with big data whereby when standard SVM was used a classification accuracy of 98.3%. However, the linear SVM gave a 82.14% classification accuracy. To obtain a clean and easy data, the linear time series method presented an effective and complete data with zero loss rate. The using six algorithms from the machine learning algorithms were used to classify the dataset, It was found that three of them achieved high ratios (SVM, ANN, GNB), and the rest (KNN,LR,PCA) achieved lower results compared to the first three algorithms.

References

[1]. S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egypt. Informatics J.*, vol. 16, no. 2, pp. 213–230, 2015.

[2]. U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: a review," *Knowledge-Based Syst.*, vol. 45, pp. 147–165, 2013.

[3]. A. Alkan, E. Koklukaya, and A. Subasi, "Automatic seizure detection in EEG using logistic regression and artificial neural network," *J. Neurosci. Methods*, vol. 148, no. 2, pp. 167–176, 2005.

[4]. H. Bhardwaj, A. Sakalle, A. Bhardwaj, and A. Tiwari, "Classification of electroencephalogram signal for the detection of epilepsy using Innovative Genetic Programming," *Expert Syst.*, vol. 36, no. 1, p. e12338, 2019.

[5]. P. A. Dekker and W. H. Organization, "Epilepsy: A manual for medical and clinical officers in Africa," World Health Organization, 2002.

[6]. M. Fani and G. Azemi, "Automatic epilepsy detection using the instantaneous frequency and sub-band energies of the EEG signals," in *2011 19th Iranian Conference on Electrical Engineering*, 2011, pp. 1–5.

[7]. R. Hussein, H. Palangi, R. Ward, and Z. J. Wang, "Epileptic seizure detection: A deep learning approach," *arXiv Prepr. arXiv1803.09848*, 2018.

[8]. E. Juarez-Guerra, V. Alarcon-Aquino, and P. Gomez-Gil, "Epilepsy seizure detection in EEG signals using wavelet transforms and neural networks," in *New trends in networking, computing, E-learning, systems sciences, and*

engineering, Springer, 2015, pp. 261–269.

[9]. E. Kabir, S. Siuly, J. Cao, and H. Wang, "A computer aided analysis scheme for detecting epileptic seizure from EEG data," *Int. J. Comput. Intell. Syst.*, vol. 11, no. 1, pp. 663–671, 2018.

[10]. E. Kabir and Y. Zhang, "Epileptic seizure detection from EEG signals using logistic model trees," *Brain informatics*, vol. 3, no. 2, pp. 93–100, 2016.

[11]. G. R. Kiranmayi and V. Udayashankara, "Neural network classifier for the detection of epilepsy," in *2013 International conference on Circuits, Controls and Communications (CCUBE)*, 2013, pp. 1–4.

[12]. Z. Lasefr, S. S. V. N. R. Ayyalasomayajula, and K. Elleithy, "Epilepsy seizure detection using EEG signals," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, 2017, pp. 162–167.

[13]. Z. Lasefr, S. S. V. N. R. Ayyalasomayajula, and K. Elleithy, "An efficient automated technique for epilepsy seizure detection using EEG signals," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, 2017, pp. 76–82.

[14]. Q. Lei, H. Lv, H. Zhang, H. Sun, and L. Tang, "Logistic regression-based device-free localization in changeable environments," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, 2016, pp. 1062–1066.

[15]. M. E. L. Menshawy, A. Benharref, and M. Serhani, "An automatic mobile-health based approach for EEG epileptic seizures detection," *Expert Syst. Appl.*, vol. 42, no. 20, pp. 7157–7174, 2015.

[16]. H. Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2027–2036, 2009.

[17]. R. Panda, P. S. Khobragade, P. D. Jambhule, S. N. Jengthe, P. R. Pal, and T. K. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction," in *2010 International conference on systems in medicine and biology*, 2010, pp. 405–408.

[18]. M. Z. Parvez and M. Paul, "Epileptic seizure detection by analyzing EEG signals using different transformation techniques," *Neurocomputing*, vol. 145, pp. 190–200, 2014.

[19]. Y. Paul, "Various epileptic seizure detection techniques using biomedical signals: a review," *Brain informatics*, vol. 5, no. 2, p. 6, 2018.

[20]. A. Sharmila and P. Geethanjali, "DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers," *Ieee Access*, vol. 4, pp. 7716–7727, 2016.

[21]. C.-P. Shen *et al.*, "A physiology-based seizure detection system for multichannel EEG," *PLoS One*, vol. 8, no. 6, p. e65862, 2013.

[22]. L. Wang *et al.*, "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, p. 222, 2017.

[23]. S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, 1987.

[24]. J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Syst. Appl.*, vol. 138, p. 112821, 2019.

[25]. Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-context learning approach for EEG epileptic seizure detection," *BMC Syst. Biol.*, vol. 12, no. 6, pp. 47–57, 2018.