Enhancing Epileptic Seizure Prediction with Machine Learning and EEG Analysis

[1]A Anandaraj and [2]P J A Alphonse
[1]Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore, India.
[2]Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India.
[1]anandaraj.a@sece.ac.in, [2]alphonse@nitt.edu

Correspondence should be addressed to A Anandaraj : anandaraj.a@sece.ac.in.

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202303017
Received 12 November 2022; Revised from 06 February 2023; Accepted 20 March 2023.
Available online 05 July 2023.

Abstract – Prediction of epileptic seizures in accurate manner and on time prediction can help in improving the lifestyle of the affected people. Many computational intelligence methods have been developed for EEG signal analysis. Since they can only handle the algorithm's complexity, new strategies have been developed to obtain the desired outcome. The goal of this work is to create an innovative method that provides the highest classification performance with the least computational expenses. This work concentrates on analyzing various deep learning models and machine learning classifiers like decision tree (C4.5), Naïve Bayes (NB), Support Vector Machine (SVM), logistic regression (LR), k-nearest neighbour (k-NN) and adaboosting model. By considering the results obtained from various classifiers, it is noted that C4.5 works well compared to other approaches. By examining the results obtained from various classifiers, this research provides valuable insights into the ensemble machine learning approaches for enhancing the accuracy and efficiency of epileptic seizure prediction from EEG signals.

Keywords – Electroencephalogram, Brain Seizure Prediction, Machine Learning, Computational Intelligence, Neural Networks.

I. INTRODUCTION

Epilepsy produces spontaneous seizures. Epilepsy has no known cure, although it is controlled through testing and treatment. Recurrent seizures can harm the neurological structure and unintentionally result in injury harm, including accidents, fractures, and even deaths [1]. Epileptic disorders, which affect the brain's Central Nervous System (CNS), are known to induce seizures, which are quite frequent and have a wide range of symptoms, including disorientation, strange behaviour, including loss of consciousness. These concerns often result in injuries from falls or tongue-biting. It might be difficult to predict when someone could have a seizure. Since most seizures occur without warning, many experts have struggled to figure out how to predict when one will happen [2]. Classification algorithms utilized in this work can predict whether someone else has a seizure. Determining the electrical impulses and dynamic characteristics of the brain and recognizing the complex determinism of the dynamics seen in seizures has been the focus of a large body of study [3]. Others used advanced mathematics to explain the brain's dynamics [4]. NTSA has also characterized EEGs based on brain activity [5]. Many investigations have been made on EEGs of Parkinson's, melancholy, and Alzheimer's patients, healthy people, and epilepsy patients. The dynamic mechanism of the neural network has now been understood in a novel way, thanks to these investigations.

Epileptic seizures and the brain activity of healthy individuals are used to classify the findings in articles relevant to epilepsy. These two examples allowed researchers to recognize and simulate the human brain non-linearity function. Additionally, it is discovered that the adjustment to the dynamic system characteristics resulted in various physiological neural impulses [6], which might lead to brain dysfunction or even another problem [7]. Epileptic seizures require efforts for several researchers.

To protect and improve the standard of living for people with epilepsy, it is thus desirable to accurately identify seizures. Many research on electroencephalogram (EEG)-based seizure detection analyse brain neuronal activity [8]. Dynamic motion, perspective fluctuations, and computing complexity make EEG seizure detection difficult. Most seizure detection approaches include segmentation and classification. Conventional seizure detection uses feature extraction before categorization. The separate proposed technique demands more care in image segmentation, is less efficient, and is highly efficient in big medical database analyses [9]. Our research aims to identify the appropriate classification technique to categorize the epileptic seizure database to regulate if the person has the seizure by using various classification approaches, as well as to examine the behaviour of the binary classifier about modifications to the classification factors.

The work is organized as follows: In section 2, various existing approaches are discussed, followed by the analysis of various existing classifiers in section 3. The numerical outcomes are discussed, and the significance of those classifiers is discussed in section 4. The solutions based on the prediction outcomes are given in section 5.

II. BACKGROUND RESEARCH

Prediction of epileptic seizures is a measure in order to prevent patients from epilepsy. The most important works related to this domain are discussed below.

During the analysis, it is predominantly noted that deep learning model has been used in most of the works related to seizure prediction. This is made possible by the use of neural networks which helps in handling complex data and thereby identifies the inherent features.

Deep Learning Techniques has been used to predict the Epileptic Seizures. It uses CHB-MIT dataset. It uses Short Time Fourier Transform for noise removal. Feature extraction has been carried out using Convolutional Neural Network. Support Vector Machine (SVM) has been for classification purpose. In [28] analyses EEG signals of scalp data and acquired 92.7% sensitivity rate and specificity of 90.8%.

To predict the seizure on the signals, Convolutional Neural Network and Common Spatial Pattern has been used. The [34] uses dataset belongs to Boston Children's Hospital-MIT for processing. Band pass filter has been used to remove noises. It uses SVM classifier and analyses scalp EEG dataset and acquires 92.2% as rate of sensitivity and False Prediction Rate (FPR) of 0.12/h.

Independent components are classified based on novel deep learning neural network combining time-series to improve classification, power spectrum densities and topoplots. Transfer learning approaches has been considered by [33] while producing new deep learning models. BASE dataset and EPILEPSIAE data sets are used for analysis.

Ensemble learning method based on deep learning was proposed by [35] for predicting the epileptic seizures. It uses empirical mode of decomposition for pre-processing and for noise removal, it uses bandpass filtering. During the generation of preictal segments, the class imbalance problem had been raised. Features had been extracted automatically from pre-processed data using a three-layered convolutional neural network. The output of SVM, CNN and LSTM models has been combined with the help of ensemble classifier using Model agnostic meta learning. It uses CHB-MIT dataset and an average rate of sensitivity as 96.28% and rate of specificity as 95.65% has been obtained, Kaggle seizure prediction dataset obtained by American epilepsy society has been used and an average sensitivity rate of 94.2% and specificity rate of 95.8% has been achieved on all subjects.

Epileptic seizures are predicted using Deep Learning model. It works on CHB-MIT database uses convolutional neural network. While classifying the preictal and ictal states, an accuracy of 99.47 % has been achieved. In [30] says seizure can be predicted using this method with a rate of sensitivity as 97.83 % and a specificity rate of 92.36 % before 10 min. FPR is calculated as 0.0764.

In [23] uses gated recurrent neural network and convolutional neural network along with cascading deep learning model to predict epileptic seizures. It uses CHB-MIT dataset and acquired an accuracy of 71.91% and a sensitivity rate of 74%. The model predicts seizures based on a patient specific approach.

Semi-dilated convolutional network (SDCN) has been proposed by Ramy Hussein et al., to predict the seizures and it uses continuous wavelet transform for pre-processing. Sensitivity rate of 98.80% was obtained as Seizure prediction when EEG signals of scalp data were analysed and 88.45–89.52% was obtained while analysing EEG signals of invasive data. It uses CHB-MIT dataset, challenge dataset of The American Epilepsy Society (AES) Seizure Prediction and challenge dataset of Melbourne University AES/MathWorks/NIH Seizure Prediction.

Numerous academics have studied various aspects of epileptic convulsions. A strategy to permit the thorough characterization of EEG time series signals was put out by the investigators in [10] to enhance signal categorization. The [11] proposed a neural network-based system for epilepsy identification was suggested to identify epileptic episodes from EEG data. The categorization of EEG signals using wavelet decomposition was addressed in [12]. They coupled a neural network's adaptation strategies with fuzzy logic and consistent transformations of the distribution function; assessed EEG sub-bands in terms of δ, θ, alpha, beta, and γ; and used multi-resolution decomposed and a multi-layer perceptron.

Recurrent neural networks of a certain kind were suggested for automatically detecting epileptic seizures. In [13] suggested EEG correlation measures using real data. Several entropy estimation techniques were used to analyze the EEG signals of epilepsy and control patients to identify epileptic seizures [14],[15]. The author implemented an eigenvector feature extraction approach employing computer algorithms for EEG signal detection. The [16] provided a novel hybrid computerized authentication protocol for EEG signal categorization. EEG signal categorization in [17] was accomplished using a feed-forward approach that made use of a back-propagation neural network (NN), while epileptic fit detection in [18] was performed with principal component (P-C.A.). The authors suggested a multi-layer perceptron neural network-based classification technique for epilepsy therapies for the identification of seizure-like activity in EEG data, and then basis-based discrete wavelet entropy was published to extract features from EEGs for a seizure-detection method by the authors.

The [19] employed an AI approach for automated epileptic seizure identification using EEG signals, and the author offered sub-band non-linear parameters for seizure detection using EEGs. The DWT and K-NN classifiers [20] were recently suggested for use to identify epilepsy. The author conducted computerized epileptic fit recognition & forecast using the wavelet transformation and packet segmentation, while [21] suggested a technique for classifying EEG data to identify epileptic seizures. Hybrid machine learning approaches to identify epileptic seizures were given by [22], robust data augmentation classification models using diverse extraction of features strategies were provided, and convolutional layers for epileptic fit forecasting were examined.

Machine learning utilizes mathematical and statistical techniques to analyze data samples, enabling computers to learn and make inferences without the need for explicit scripting or programming instructions. In [23] introduced computer vision in games and analytical thinking techniques to learn which the first time this substantial improvement was acknowledged was. The fundamental idea of machine learning is to gain insight from big to foresee or make judgments based on a given goal [24]. Machine learning technology makes many time-consuming processes easier and faster. More computational capabilities and data make it easier to train machine learning models to predict events with near-perfect reliability. Various machine learning techniques are presented in several works [25]. Machine learning algorithms are often divided into supervised, unsupervised, and adequate institutional. Learning techniques like SVM may separate machine learning algorithms, Random Forest algorithm (RF), and naves Bayes (NB) are a few of the well-known machine learning) methods [25].

Table 1. Related Works on Deep Learning Model with Different Dataset

Research Work	Pre-processing /Feature Extraction	Patient Specific	Data Set	Classifier	Channel	Accuracy	Sensitivity
Abir Affes a et al	Neural network with attention mechanism	Yes	CHB-MIT	CNN, GRNN	23 * 2	71.91	74
Syed Muhammad Usman a et al	EMD, Bandpass filter, GAN, STFT	NO	CHB-MIT, American epilepsy	Ensemble SVM, CNN,LSTM	6	96	95
Mamli et al.	Bandpass Filter	NO	CHB-MIT	KNN, SVM	23	90	97
Ramy Hussein a et. al	continuous wavelet transform	Yes	CHB-MIT EEG dataset.	SDCN	23	99	99
Ramy Hussein a et. al	continuous wavelet transform	Yes	American Epilepsy	SDCN	16	0.928/0.856	88.45
Mingkan Shen a et. al	discrete wavelet transform	NO	CHB-MIT	SVM	16	97	96.15
Mingkan Shen a et. al	discrete wavelet transform	NO	Bonn Data Set	RUSBoosted tree Ensemble models	-	96	96
Fábio Lopes et. al	**Not Mentioned**	NO	EPILEPSIAE	ensemble deep neural networks	-	93.48	
Yuan Zhang et. al	Butterworth band-pass,PCA, CSP	NO	CHB-MIT	CNN,SVM,MLP	16	92.2	90
Yan et. al	Medium Power Spectrogram (MPS)	NO	CHB-MIT	CNN	16	98	98
Tsiouris et al	Minimizing Channel	NO	CHB-MIT	LSTM	23	99	
Hussein et al.,	Band-Pass Filte	NO	Bonn Data Set	RNN ,LSTM	-	97	94

Table 1. depict the outcomes based on the evaluation with various deep learning and models upon comparing the techniques, a common disadvantage identified while using such model is of there is an increase in computational

complexity. Because of this, real time implementation of most of the models is not possible. Patient specific analysis is a bit complex as it requires high level of computing resources.

Based on the analysis conducted, it has been identified that feature extraction and EEG signal classification pose significant challenges due to the variations in brain signals. These variations are influenced by factors such as the number of channels, brain location, and the diverse signal patterns observed across different samples. These challenges complicate the accurate extraction of relevant features and the classification of EEG signals.

Another prominent challenge faced by researchers in this field is the prediction and detection of EEG signals to achieve reliable and effective epileptic seizure outcomes using learning approaches, particularly in real-time applications. Real-time prediction requires quick and accurate identification of seizure patterns in the EEG signals, which can be challenging due to the complexities and dynamics of the brain's electrical activity during a seizure event.

Addressing these challenges is crucial for future advancements in epileptic seizure prediction. Researchers need to develop robust techniques that can handle the variations in EEG signals, optimize feature extraction methods, and design efficient algorithms for real-time prediction and detection. Additionally, collaboration and data sharing among researchers can contribute to the development of more accurate and reliable models for epileptic seizure prediction and improve the overall understanding of this complex condition.

III. DATA ACQUISITION

The Epileptic Seizure Recognition Data Set utilized in this research work [18] consists of 11,500 samples, with each sample containing 178 characteristics. The dataset is well-structured, and the samples are evenly distributed. These samples are categorized into five groups, denoted as Y = 1, 2, 3, and 4, based on predefined standards.

- **Class 5 -** eyes open during recompilation of EEG data; this study is called EYEO.
- **Class 4 -** eyelids closed while receiving EEG signal; in this work, this condition is referred to as E. YEC.
- **Class 3 -** Brain tumour was confirmed after EEG from the healthy brain region was recorded; it is referred to as an HSTU. MOR.
- **Class 2 -** EEG signal obtained where brain tumour existed, labelled TUMOR.
- **Class 1 -** Seizure event monitoring, ES.

Every sample has 178 characteristics representing brainwave measurements/second for the various instances. **Fig 1** displays the epileptic seizure database sample and each class's waveform. Since only samples belonging to class 1 had an ES, our work will be divided into ES and non-ES instances, including classes 2, 3, and 4 in **Table 2**.

Table 2. Number of Cases/Class

Class	Description	Abbreviation	Total cases	Binary case
1	Seizure activity recording	ES	2300	2300
2	Tumour location	TUMOR	2300	9200
3	EEG activity recorded from the healthy brain area	HUMOUR	2300	-
4	Eyes closed	EYES	2300	-
5	Eyes open	EYEO	2300	-

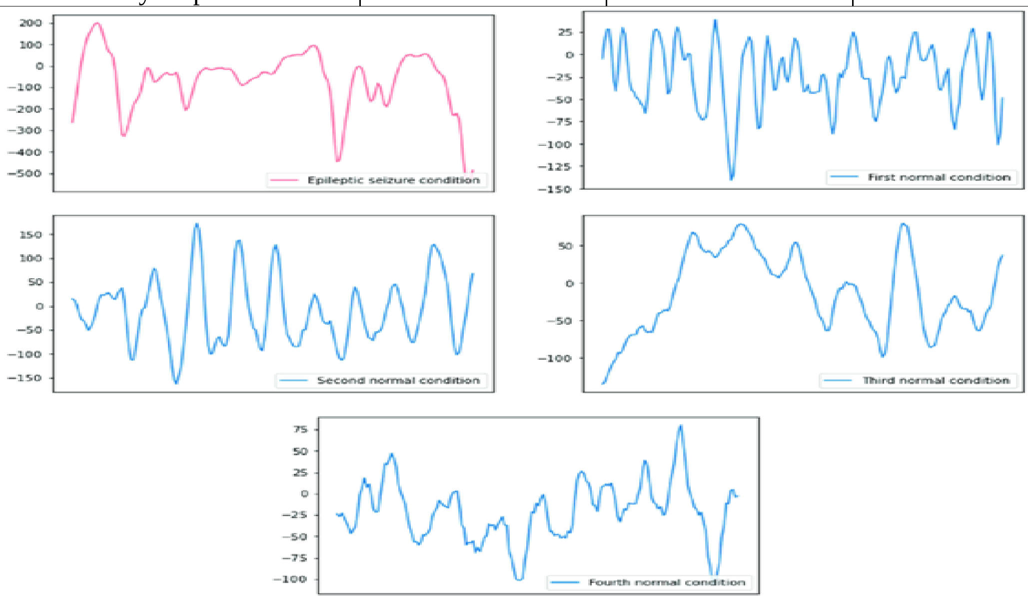

Fig 1. Epileptic Seizure Dataset-Based Sample Waveforms

IV. MACHINE LEARNING ALGORITHMS

The most popular machine learning models for diagnosing diseases are thoroughly described in this article.

Decision Tree (C4.5)

In [9] proposed the ID3 algorithm who uses information gain to construct the DTs. The extended form of the ID3 algorithm is C4.5, a classifier where the information gain ratio is adopted for classification. It is designed using divide and conquers concept and expressed as in Eq. (1) & Eq. (2)

$$\mathrm{Gain}\ (\mathrm{S},\mathrm{A}) = \mathrm{Entropy}\ (\mathrm{S}) - \sum_{\mathrm{j}=1}^{\mathrm{v}} \frac{|\mathrm{D_j}|}{|\mathrm{D}|} * \mathrm{Info}\ (\mathrm{D_j}) \tag{1}$$

$$\mathrm{Entropy}\ (\mathrm{S}) = \sum_{\mathrm{i}=1}^{\mathrm{c}} - \mathrm{P_i} \log_2 \mathrm{P_i} \tag{2}$$

Here, $'D'$ is dataset partitioning, $Entropy\ (s)$ is sub-set before splitting, $'v'$ dataset values and p_i is the i^{th} target value proportion. The fitness function is evaluated and the flag bit is allocated for feature characterization. From the $'n'$ population, the attribute with 1 is a feature and 0 is unselected feature as in Eq. (3):

$$\mathrm{x_{id}^{new}} = \begin{cases} 1 & \text{if sigmoid } (\mathrm{V_{id}^{new}}) > U\ (0,1) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The error prediction rate is utilized as fitness function for feature selection as depicted in Eq. (4) & Eq. (5):

$$\text{fitness function} = \varepsilon \mathrm{A} + \varepsilon' \frac{\mathrm{N} - \mathrm{LS}}{\mathrm{N}} \tag{4}$$

$$\mathrm{A} = \frac{\text{number of properly predicted samples}}{\text{total samples}} \tag{5}$$

Here,$'\mathrm{A}'$ represents accuracy, LS specifies feature subset length, N is total features ε and ε' represents weighted parameter and feature selection quality, ε = [0, 1] and$\varepsilon' = 1 - \varepsilon$. The model provides better results while handling uni-modal and multi-modal dimensionality functions. The target is to offer the optimal solutions.

Support Vector Machine (SVM)

It is a supervised learning approach for classifying data into positive and negative classes as in Eq. (6) & (7):

$$x_i.w + b \geq +1 \forall y_i = \pm 1 \tag{6}$$

$$x_i.w + b \leq -1 \forall y_i = -1 \tag{7}$$

Here, $w\ and\ x$ represents vectors, and $'b'$ represents bias, the hyperplane is expressed as $w^T.x = 0$ which is given based on the distance among margins to improve the margin among hyperplanes. It is reduced with $||w||$ as in Eq. (8):

$$\min_{w,b} = \frac{1}{2}||w||^2 \tag{8}$$

The non-linear data is not classified properly. To validate the error rate, the stacking variable has to be included ζ_i where $i = 1,2,3,..,n$, and it is related to $'C'$ as expressed in Eq. (9):

$$\min_{\mathrm{b,w,}\zeta} = \frac{1}{2}||w||^2 + C \sum_{i=0}^{n} \zeta_i \tag{9}$$

During kernel computation, SVM needs to determine marginal width among hyperplanes. There are diverse methods to enhance SVM algorithm-based prediction accuracy.

k-Nearest Neighbor (k-NN)

It is a supervised learning algorithm adopted for prediction problems. It measures the feature similarity to predict the new data points based on how nearer the points fit with the training set. It works superiorly for non-linear data when there is no consideration towards the data.

Algorithm 1: k-NN
1: Provide testing and training data
2: Initialize the nearest data point, i.e. k can be any integer.
3: For all test data
4: Evaluate distance among the training and testing data with Euclidean distance.
5: Sort data points; ' //ascending
6: Select 'k' rows from the array
7: Allocate class to test points.

It is used to categorize new objects with distances to labelled samples where the values represent the number of neighbourhood. It is known as the deciding factor as is decide how many neighbours can stimulate the process. When the value is set as 1, new data objects are allocated to the class on the nearest neighbour. Neighbourhoods are considered from a set of training data objects where correct classification is known already.

Naïve Bayes

NB is the statistical model that relies on the record or data point, it predicts class membership. The class with the best odds has been the most likely. The NB algorithm projects probability rather than making predictions. It is based on the Bayes theorem, where all the features are independent conditionally. NB is appropriate for categorizing high-dimensional datasets. The model considers an attribute value independent of others (class). Consider, D as training data corresponding class labels. The tuple is specified as $X = \{A_1, A_2, \dots, A_n\}$ and m classes are represented as $C_1, C_2, \dots, and\ C_m$. The classifier predicts X from the class with a high probability for the given tuple X and it predicts the tuple X to C_i iff

$$P(C_i|X) > P\left(C_j|X\right)\ for\ 1 \le j \le m, \qquad j \ne i \tag{10}$$

Therefore, $P(C_i|X)$ is improved, class C_i as $P(C_i|X)$ is improved, and it is known as the maximal posteriori hypothesis.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{11}$$

The attribute values are independent conditionally on one another, then.

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \tag{12}$$

Where x_k specifies attribute value A_k for tuple X. When A_k is classified, then $P(x_k|c_i)$ represents several tuples with C_i in D with x_k value for A_k partitioned by $|C_{i,D}|$, total tuples of class C_i in D. The classifier identifies the X class labels with class C_i iff:

$$P(X|C_i)\ P(C_i) > P\left(X|C_j\right)P\left(C_j\right)\ \ for\ 1 \le j \le m, \qquad j \ne i \tag{13}$$

The classifier is efficient that may show minimal error for classification.

Logistic Regression

Logistic regression (LR) is a machine learning technique used for classification, specifically in predicting the probability of disease occurrence by considering risk factors and categorizing each classification accordingly. It is adapted generally to categorize and specifies the probability occurrence of every classification event as in Eq. (14):

$$Prob\ (Y = 1) = \frac{e^z}{1 + e^z} \tag{14}$$

Here, Y specifies binary dependent variables ($Y = 1\ when\ event\ occurs\ and\ Y = 0\ else$), $'e'$ specifies the natural logarithms and $Z -$ specifies $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$ with coefficients β_j, constant β_0, and predictors X_j for predictors $'p'(j = 1,2,3 \dots., p)$.

AdaBoost

A classifier called AdaBoost combines many weak classifiers into a single classifier. AdaBoost works by providing more weight to information that is challenging to identify and less emphasis on those who are already correctly classified. Both regression analysis and classification are possible uses [20]. Adaboost is also a machine learning-based Meta algorithm that sequentially integrates independent individual hypotheses to improve accuracy. Also, it is known adaptive boosting. The target is to convert weak learners to stronger and helps to resolve the bias issues.

The algorithm that corresponds to Adaboost is provided below

Algorithm 2. AdaBoosting

Input: Training input data $(x_1, y_1) \ldots . (x_n, y_n)$ with $y_n \in \{-1, +1\} \forall\, n;$ set k training objects

Output: Boosting hypothesis

Step 1: Sample initialization $(x_n, y_n): D_1(n) = \frac{1}{N}, \forall\, n;$ $//t = 1: T;$

Step 2: Learning rate $h_t(x) -. \{-1, +1\}$ with training data; //weaker

Step 3: Measure error rate with $\in_t = \sum_{n-1}^{N} D_t(n) \prod [h_t(x_n) \neq y_n]$

Step 4: Hypothesis analysis $h_t : \propto_t = \frac{1}{2} \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$

Step 5: Upate sample with $D_{t+1}(n) \propto \begin{cases} D_t(n) * \exp(\propto_t) & if\ h_t(x_n) = y_n \\ D_t(n) * \exp(\propto_t) & if\ h_t(x_n) \neq y_n \end{cases}$

Step 6: Compute $D_{t+1}(n) \propto = D_t(n) * \exp\left(\propto_t y_n h_t(x_n)\right)$

Step 7: Normalize $D_{t+1};$

Step 8: Normalize $1: D_{t+1}(n) - \frac{D_{t+1}(n)}{\sum_{m-1}^{N} D_{t+1}(m)}$

Step 9: Normalize hypothesis with $H(x) = sign\ (\sum_{t-1}^{T} \propto_t h_t\ (x))$

Step 10: Generate class labels for (x_n, y_n)

Step 11: Return ensemble hypothesis $f(x_n, y_n)$

V. PERFORMANCE EVALUATION

A unified performance evaluation methodology for seizure detection is absent, and NSD system results are reported using different measures. Consequently, it is difficult to compare the suggested methodologies. Main seizure detection system measures are epoch-based and event-based.

The fragmentation of the signals into distinct periods, or "epochs," is the foundation of epoch-based metrics. This method is a common pre-processing phase. The collected epochs are split into two classes: historical seizure periods, often referred to as positive and non-seizure epochs typically refers to negative. As a result, seizure diagnosis is the binary issue. The classifiers designed for automatic detection typically indicate that a certain epoch falls into the favourable or unfavourable category. The effectiveness of the systems is determined by comparing the classification choices made by the classification to the manual choices made with each session through one or more EEG specialists. The classifier's conclusion is captured by the binary classification with four types: true positives (TP), epochs properly identified as seizures; false positives (FP), epochs wrongly tagged as seizures; true negatives (TN), successfully identified non-seizure epochs; and false negatives (FN).

Accuracy (Acc). It is the number of occurrences accurately identified. The formulae given below are to determine accuracy:

$$Accuracy = \frac{TP + TN}{(TP + FN + TN + FP)} \tag{15}$$

Precision (Pn). Precision is calculated as the ratio of accurately forecasted to all anticipated positive observations.

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

Recall (Rc). The percentage of total useful content that the good stuff identifies is known as recall.

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

Sensitivity (Sn). Sensitivities is the only positive metric considering all situations.

$$Sensitivity = \frac{TP}{TP + FN} * 100 \tag{18}$$

Specificity (Sp). It measures the number of correctly detected true negatives and is computed as follows:

$$Specificity = \frac{TN}{TN+FP} *100 \tag{19}$$

F-measure. The F1 score is a harmonic average of memory and accuracy. The highest possible F grade is 1, which denotes faultless accuracy and recall.

$$F - measure = \frac{2 * recall * precision}{recall + precision} \tag{20}$$

As determined by professionals, a seizure's "event" is often defined as the period between its onset and termination. Some of the metrics are.

- *Good Detection Rate (GDR)* measures the system's power accuracy in identifying seizure occurrences. If the system recognizes at least one epoch during seizures, it has successfully detected the event.
- *False Discovery Rate (FDR)* gives the fraction of falsely recognized seizure occurrences.
- *False Detection per Hour (FDH)* expresses how many seizure episodes the system detects in one hour that does not coincide with the events assigned labels by the expert.

Table 3. Classifier Outcomes

Method	Accuracy (%)	ROC	MAE	RMSE	Time (s)
k-NN	95.2	88.3	4.7	21.8	0.02
NB	95.7	95.9	4.2	20.6	0.30
SVM	94.8	90.1	5.4	22.6	10.36
C4.5	97.09	99.6	6.67	15.27	17.04
LR	81.94	52.9	29.64	38.84	3.68
Adaboost	91.98	92.20	12.8	25.5	22.05

Table 4. Performance Metrics Comparison

Method	Accuracy (%)	Recall	F-measure	Precision
k-NN	95.2	82	82	82
NB	95.7	85	85	85
SVM	94.8	89	89	89
C4.5	97.09	87	87	87
LR	81.94	86	85	86
Adaboost	91.98	84	84	84

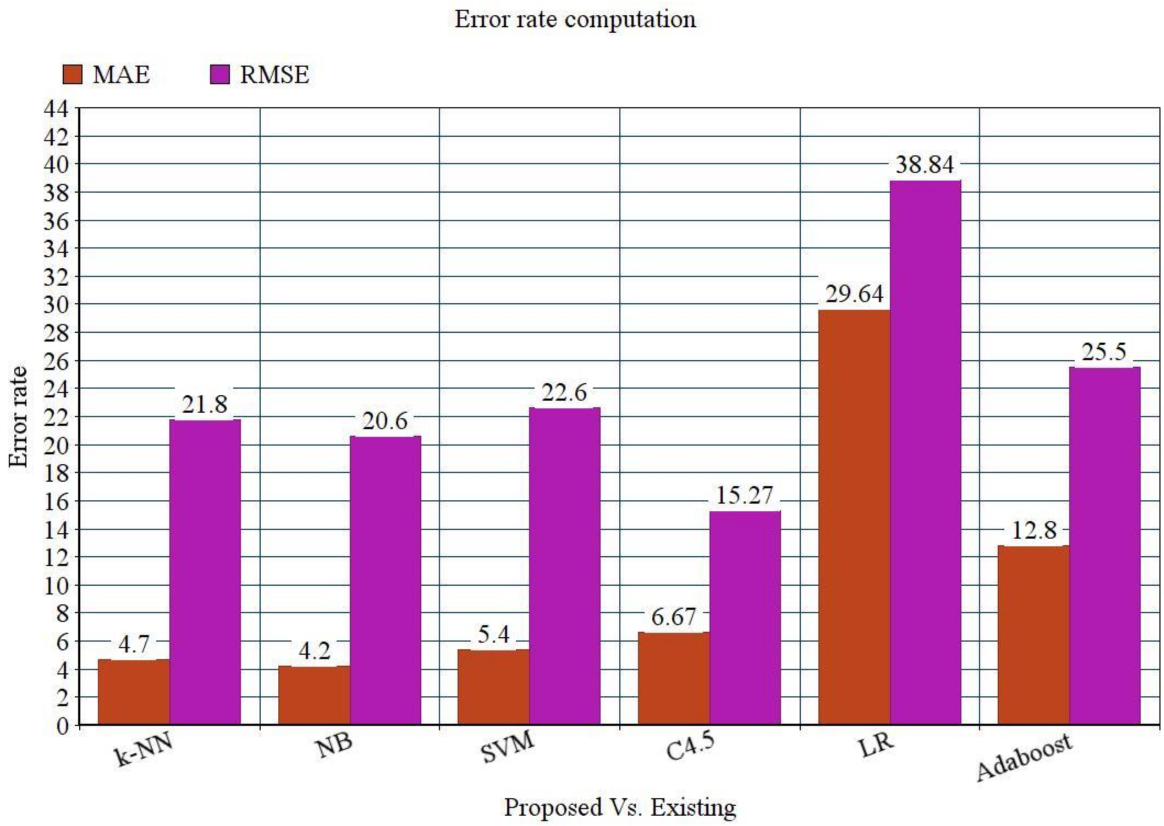

Fig 2. Error Rate Computation

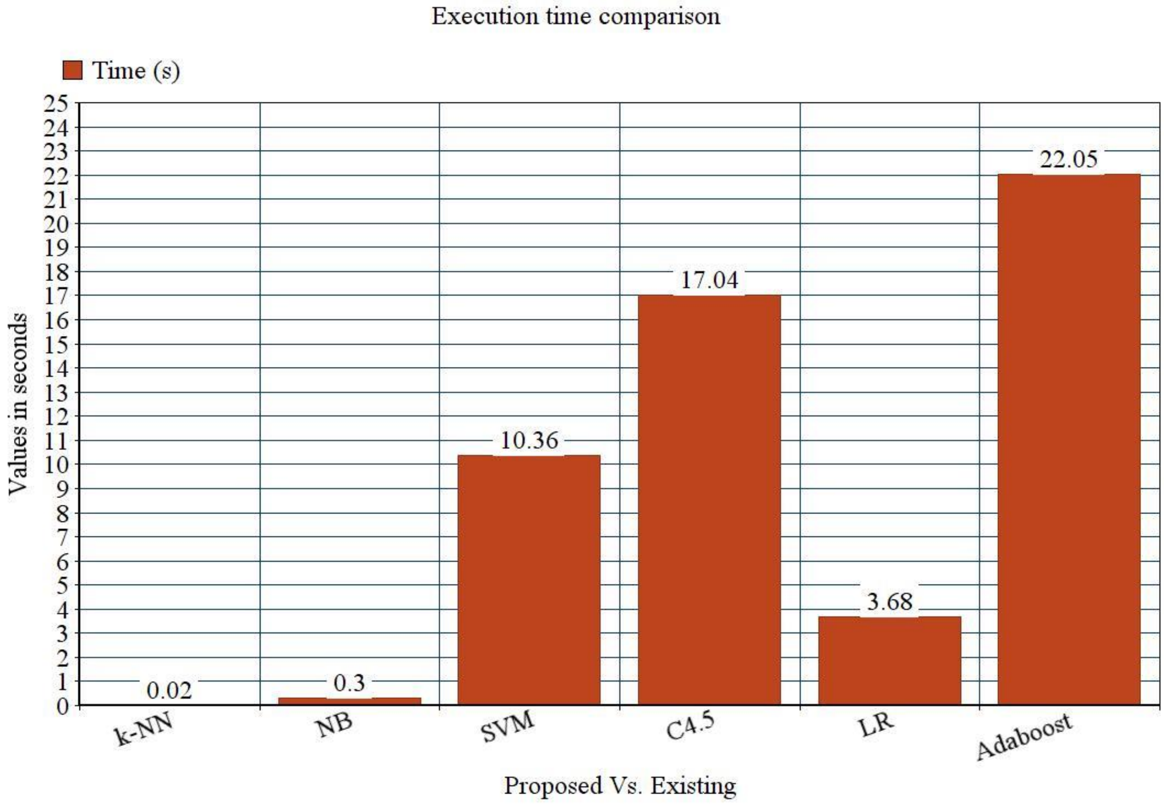

Fig 3. Execution Time Comparison

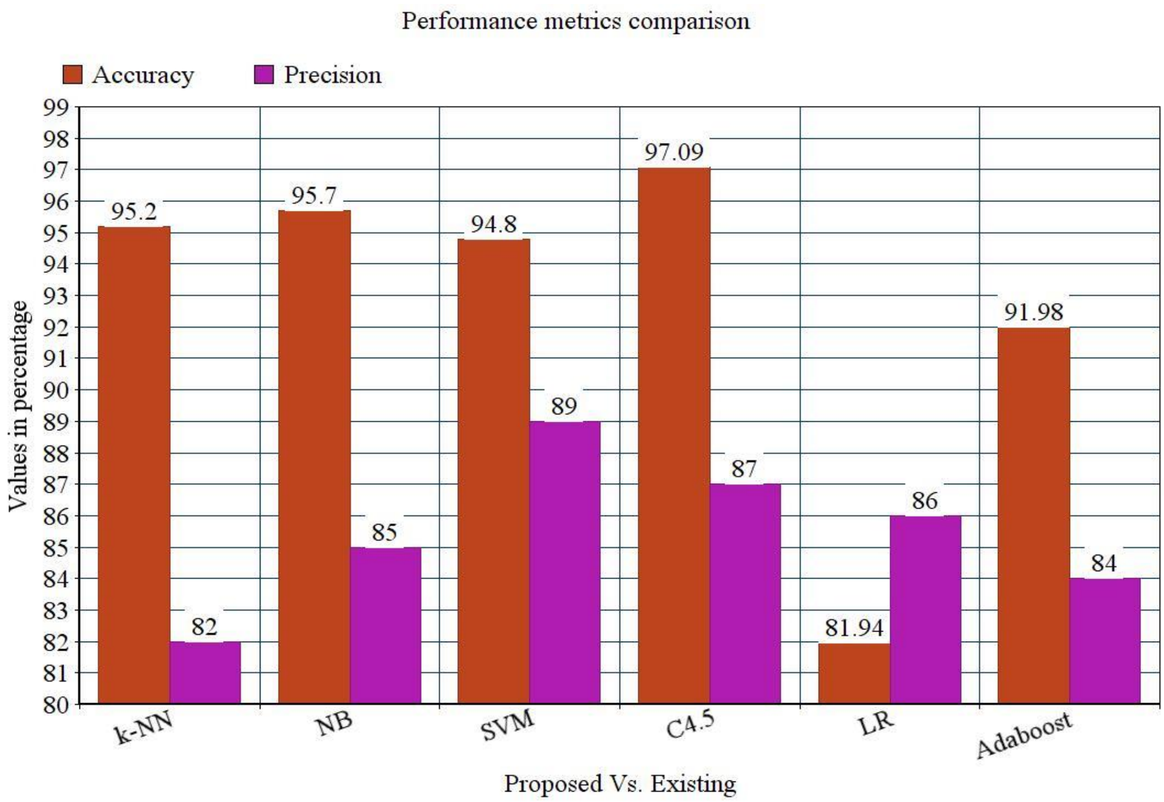

Fig 4. Accuracy and Precision Comparison

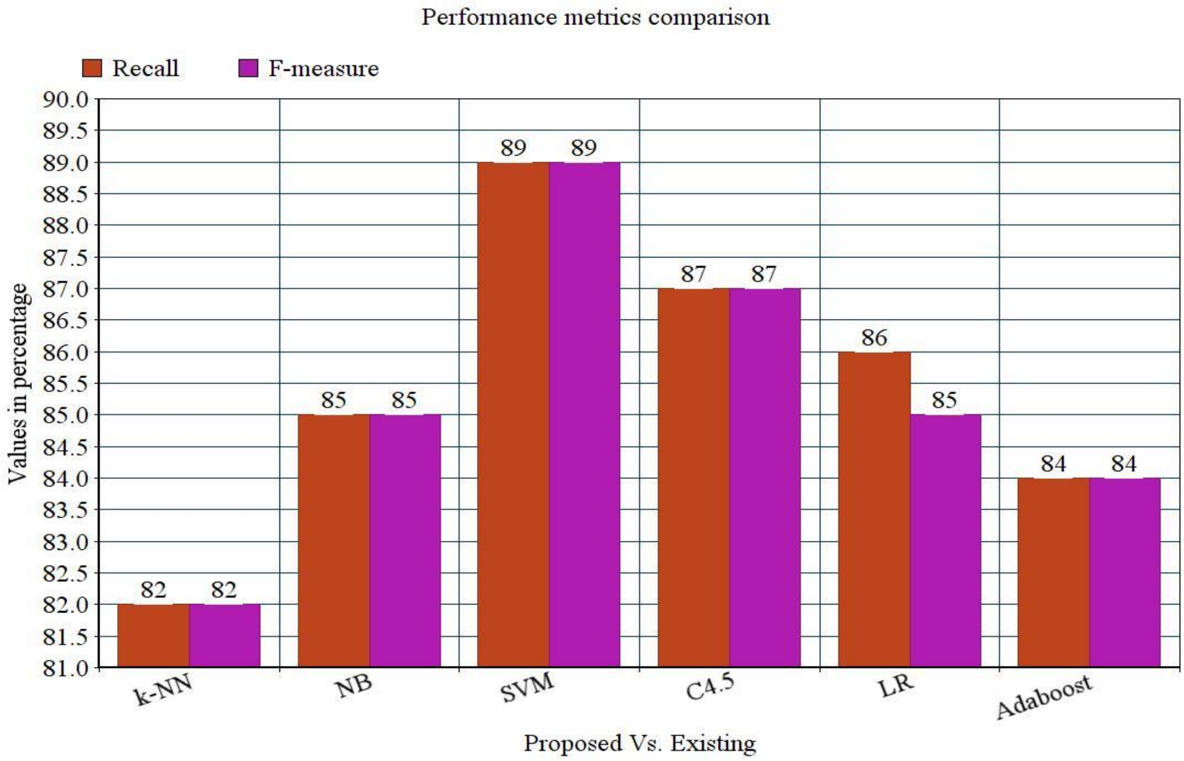

Fig 5. Recall and F-Measure Comparison

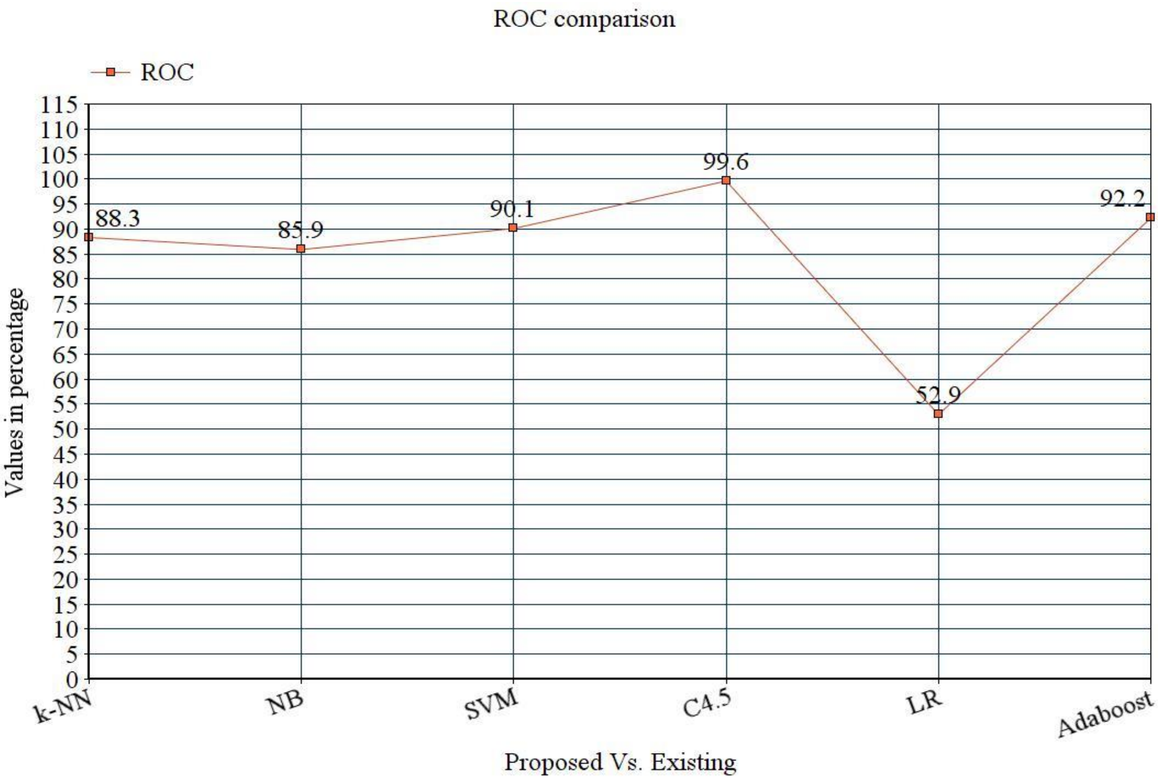

Fig 6 . ROC Curve

Table 3 and **Table 4** depict the outcomes based on the evaluation with various classifier models. The accuracy of C4.5 is 97.09% which is 1.89%, 1.39%, 2.29%, 15.15% and 5.11% higher than k-NN, NB, SVM, LR and Adaboosting. The recall of SVM is 89% which is 7%, 4%, 2%, 3% and 5% higher than k-NN, NB, C4.5, LR and AdaBoost. The F1 measure of SVM is 89% that is 7%, 4%, 2%, 3% and 5% superior to k-NN, NB, C4.5, LR and AdaBoost. The precision of SVM is 96% which is 2% higher than k-NN and NB, 1.8%, 17.09% and 6% higher than LR and AdaBoost. The ROC of C4.5 is 99.6% which is 11.3%, 3.7%, 9.5%, 46.7% and 7.4% higher than other approaches. The MAE of NB is 4.2, which is comparatively lesser than k-NN, SVM, C4.5, LR and adaboosting. The execution time of k-NN is 0.02s which is extremely lesser than NB, SVM, C4.5, LR and AdaBoost. Based on these analyses, it is proven that the provided classifier models show potential in their way. In this analysis, it is known that C4.5 shows a better prediction rate compared to other approaches. Apart from C4.5, NB shows a prediction accuracy of 95.7%. Similarly, k-NN shows lesser execution time, and NB shows a lesser error rate (See **Fig 2** to **Fig 6**).

The major advantages of C4.5 are its functionality for both discrete and continuous data. Also, it deals the issues with incomplete data. The C4.5 naturally adopts a pruning process to deal with over-fitting issues. However, this DT model is relatively expensive because of its computational complexity and time. While dealing with k-NN, the model is extremely easier to implement. Also, new data can be included seamlessly before the prediction process. The major disadvantage encountered during the epileptic seizure prediction is its inability to handle high-dimensional data as it is sensitive toward outliers, missing values and noisy data. In the case of NB, the model is best suited for handling multi-class prediction with independent feature consideration. It works well compared to other features and needs lesser training data. The drawback is its independency towards the traits, which is not best suited for real-time application. While in the case of SVM, the model comparatively works well even in the case of understandable dissociation margin among the classes. The accuracy is more productive with high-dimensional data.

The model is efficient when dimensionality exceeds the number of samples. However, the model is inappropriate for larger datasets and fails to work well in noisy data. When the number of features at every point exceeds the total training samples, then SVM may fail. Logistic regression is a learning model which is easier to interpret, implement and efficient for training purposes. When the number of observations is lesser than the number of extracted features, LR should not be used for experimentation as it leads to over-fitting issues. Also, it does not have a common assumption about class feature space. The accuracy of some weak classifiers can be enhanced with Adaboost. It works well for binary classification and needs a quality dataset. With these practical implications, it is seen that C4.5 gives a better outcome than the provided epileptic seizure dataset. But the computational complexity encountered in these abovementioned models paves the way to adopt a novel ensemble approach or hybridized approach in the future.

VI. CONCLUSION

This work provides an extensive work on epileptic seizure prediction using learning approaches with the online available EEG dataset. Here, various existing classifiers like DT (C4.5), SVM, k-NN, NB, LR and adaboosting classifiers are used for prediction. The accuracy of C4.5 is 97.09%, the recall of SVM is 89%, F1 measure of SVM is 89%, precision of SVM is 96%, ROC of C4.5 is 99.6%. The MAE of NB is 4.2, execution time of k-NN is 0.02s. Based on the analysis with the provided learning classifiers, it is observed that there are some research issues which need to be addressed. In the future, an ensemble classifier will be proposed to classify EEG signals for epileptic seizure prediction. Ensemble classifiers have the ability to provide higher prediction accuracy compared to various individual models. It is extremely useful when both non-linear and linear data are considered for analysis. Also, it can reduce the prediction dispersion and model performance.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding was received to assist with the preparation of this manuscript.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1]. Sazgar M, Young MG,"Seizures and Epilepsy. In: Absolute Epilepsy and EEG rotation review", pp. 9–46. Springer, New York,2019.
[2]. Fisher RS, "The new classification of seizures by the international league against epilepsy 2017", Curr Neurol Neurosci Rep 17(6):48,2017.
[3]. Aljumah AA, Ahamad MG, Siddiqui MK," Application of data mining: diabetes health care in young and old patients". J King Saud University Computations Information Science 25(2):127–136,2013.
[4]. Singh GA, Gupta PK, "Performance analysis of various machine learning-based approaches for detecting and classifying lung cancer in humans", Neural Computer Application 31(10):6863–6877,2019.
[5]. Paul Y,"Various epileptic seizure detection techniques using biomedical signals: a review". Brain Inform 5(2):6,2018.

[6]. Boonyakitanont P, Lek-uthai A, Chomtho K, Songsiri J, “A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. Biomed Signal Process Control 57:101702,2020.
[7]. Van Westrhenen A, De Cooman T, Lazeron RH, Van Huffel S and Thijs RD , “Ictal autonomic changes as a tool for seizure detection: a systematic review”. Clin Auton Res 29(2):161–181
[8]. Sharmila A, Geethanjali P, “A review on the pattern detection methods for epilepsy seizure detection from EEG signals”, Biomed Eng /Biomedizinische Technik. 64(5):507–17,2019
[9]. A. K. Sangaiah, S. Rezaei, A. Javadpour, and W. Zhang, “Explainable AI in big data intelligence of community detection for digitalization e-healthcare services,” Applied Soft Computing, vol. 136, p. 110119, Mar. 2023, doi: 10.1016/j.asoc.2023.110119.
[10]. Lahmiri S, “An accurate system to distinguish between normal and abnormal electroencephalogram records with epileptic seizure-free intervals”. Biomed Sign Process Control 40:312–317,2018
[11]. Donos C, Dümpelmann M, Schulze-Bonhage A, “Early seizure detection algorithm based on intracranial EEG and random forest classification. International journal of Neural System 25(05):1550023,2015.
[12]. Al Ghayab HR, Li Y, Abdulla S, Diykh M, Wan X , “Classification of epileptic EEG signals based on simple random sampling and sequential feature selection”,Brain Inform 3(2):85–91,2016.
[13]. Adnan MN, Islam MZ ,“Forest CERN: A new decision forest building technique. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining”, pp. 304–315,2016,
[14]. Quinlan JR , “C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc.”, San Francisco, 1993
[15]. Arlot S, Celisse A, “A survey of cross-validation procedures for model selection. Statistics surveys” 4:40–79, 2010
[16]. Siddiqui MK, Islam MZ, Kabir MA,”A novel quick seizure detection and localization through brain data mining on ECoG dataset”. Neural Computing and Applications 1–14, 2018
[17]. Lee H, Kim S “Black-box classifier interpretation using decision tree and fuzzy logic-based classifier implementation”. Fuzzy Logic Intelligent System 16(1):27–35, 2016
[18]. Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE, “Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state”, Phys. Rev. E, 64, 061907,2001.
[19]. Zhang Y, Zhang Y, Wang J, Zheng X, “Comparison of classification methods on EEG signals based on wavelet packet decomposition”. Neural Comput Appl 26(5):1217–1225, 2014.
[20]. Gill AF, Fatima SA, Akram MU, Khawaja SG, Awan SE, “Analysis of EEG signals for detection of epileptic seizure using hybrid feature set”,Theory and Applications of Applied Electromagnetics, pp. 49–57, 2015.
[21]. Orellana MP, Cerqueira F, “Personalized epilepsy seizure detection using random forest classification over one-dimension transformed EEG data”. bioRxiv, 070300, 2016.
[22]. Yan A, Zhou W, Yuan Q, Yuan S, Wu Q, Zhao X, Wang J, “Automatic seizure detection using Stockwell transform and boosting algorithm for long-term EEG”. Epilepsy Behav 45:8–14, 2015.
[23]. Amin HU, Malik AS, Ahmad RF, Badruddin N, Kamel N, Hussain M, Chooi W-T, “Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques” . Austr Phys Eng Sci Med 38(1):139–149, 2015.
[24]. Mursalin M, Zhang Y, Chen Y, Chawla NV, “Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier”, Neurocomputing 241:204–214, 2017
[25]. Fergus P, Hussain A, Hignett D, Al-Jumeily D, Abdel-Aziz K, Hamdan H, “A machine learning system for automated whole-brain seizure detection”. Applied Computer Information 12(1):70–89, 2016.
[26]. R. Subha, M. Haritha, B. Nithishna and S. G. Monisha, "Coma Patient Health Monitoring System Using IOT," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 1454-1457, 2020, doi: 10.1109/ICACCS48705.2020.9074174.
[27]. S. Yuvaraj, J. V. Franklin, V. S. Prakash, A. Anandaraj and R. Subha, "A Survey on Human Face Emotion Recognition using Machine Learning Models," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, pp. 1993-1996, 2023, doi: 10.1109/ICACCS57279.2023.10112955.
[28]. Syed Muhammad Usman, Shehzad Khalid, Sadaf Bashir, “A deep learning based ensemble learning method for epileptic seizure prediction”, Computers in Biology and Medicine, Vol 136, 2021,104710.
[29]. Mingkan Shen, Peng Wen, Bo Song, Yan Li,An, “EEG based real-time epilepsy seizure detection approach using discrete wavelet transform and machine learning methods”, Biomedical Signal Processing and Control, Vol77,2022,10382.
[30]. Ranjan Jana, Imon Mukherjee, “Deep learning based efficient epileptic seizure prediction with EEG channel optimization”, Biomedical Signal Processing and Control, Volume 68,2021,102767.
[31]. Abir Affes, Afef Mdhaffar, Chahnez Triki, Mohamed Jmaiel, Bernd Freisleben, “Personalized attention-based EEG channel selection for epileptic seizure prediction”, Expert Systems with Applications, Volume 206,2022,117733.
[32]. A. K. Sangaiah, A. Javadpour, F. Ja’fari, H. Zavieh, and S. M. Khaniabadi, “SALA-IoT: Self-Reduced Internet of Things with Learning Automaton Sleep Scheduling Algorithm,” IEEE Sensors Journal, pp. 1–1, 2023, doi: 10.1109/jsen.2023.3242759.
[33]. Fábio Lopes , Adriana Leal, “Ensemble Deep Neural Network for Automatic Classification of EEG Independent Components”, IEEE Transactions On Neural Systems And Rehabilitation Engineering, Vol. 30, 2022.
[34]. Yuan Zhang, “Epilepsy Seizure Prediction on EEG Using Common Spatial Pattern and Convolutional Neural Network”, IEEE Journal Of Biomedical And Health Informatics, VOL. 24, NO. 2, Feb 2020.
[35]. S.Muhammad Usman, “Epileptic Seizures Prediction Using Deep Learning Techniques”, IEEE Journal publication February 27, 2020.