International Journal of Academic Health and Medical Research (IJAHMR)
ISSN: 2643-9824
Vol. 6 Issue 12, December - 2022, Pages: 41-53

Early Prediction of Epileptic Seizure Using EEG Spectral Features and Machine Learning Approaches

Rokeya Akter[1], Fahima Hossain[2,*]

[1,2]Department of Computer Science & Engineering, Hamdard University Bangladesh, Munshiganj, Dhaka, Bangladesh

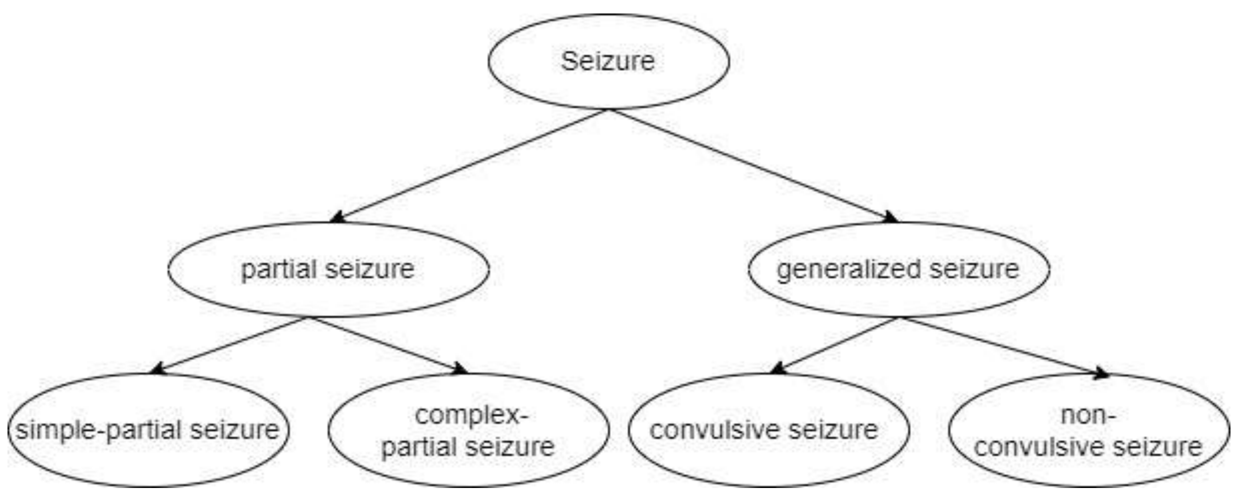
[1]al.mowtushi@gmail.com, [2]minda.fahima@gmail.com

Abstract: *This paper proposes a system for predicting epileptic seizures from EEG signals using Machine Learning approaches in order to prevent seizures through medication. A significant chronic neurological illness called epilepsy can be identified by examining the brain signals that brain neurons produce. Electrocorticography (ECoG) and electroencephalography (EEG) media are frequently used to detect these brain impulses. These signals generate a large amount of data and are complicated, noisy, non-linear, and non-stationary. Therefore, identifying seizures and learning about the brain's functions is a difficult undertaking. Without sacrificing performance, machine learning classifiers can classify EEG data, detect seizures, and highlight pertinent, meaningful patterns. In this study, the epileptic seizure dataset was classified using a variety of classifiers. Support vector machines performed better than Naive Bayes, K-Nearest Neighbors, Random Forest classifier, Logistic Regression, Bagging classifier, AdaBoost classifier, Gradient Boosting classifier, Stochastic Gradient Descent (SGD) classifier, Multi-layer Perceptron (MLP) classifier, XGBoost classifier, and Decision Tree classifier, as demonstrated. In this study, we employed the CHBMIT dataset of scalp EEG signals and tested our suggested methodology on the dataset's 22 participants. With superior performance and higher prediction accuracy, our suggested seizure prediction approach is able to reach 95.88% accuracy, 86.91% recall, 1% precision, and 1% sensitivity.*

Keywords: EEG signals, epileptic seizure, prevalence, scaling, machine learning algorithms

1. Introduction

Epilepsy is a neurological brain disorder identified by the frequent occurrence of seizures. It is estimated that 50 million people globally have epilepsy [1]. Anti-epileptic medication (AED) advancements for seizure control have given individuals with the chronic illness hope. However, early seizure prediction is a crucial component of effective epilepsy treatment and management [2]. The Latin and Greek words "epilepsia" (which mean "seizure" or "to seize upon") are the source of the word epilepsy [3]. An epileptic seizure is a rapid anomaly in the brain's electrical activity that affects the entire body and is characterized by excessive discharges of neuronal networks in the cerebral cortex [4]. A loss of awareness may also be present during epileptic seizures, which are brief episodes of partial or complete aberrant, unintended movements of the body [5] . They are quite prevalent and exhibit a wide range of symptoms, including bewilderment, odd behavior, and loss of awareness. Patients with epilepsy experience unanticipated, sudden seizures that leave them defenseless and vulnerable to suffocation, drowning, or injury from falls or car accidents [6]. These symptoms frequently result in injuries from falls or tongue-biting [7]. If seizures have already occurred, it has been found that in more than 30% of cases, the patient's following seizures cannot be controlled by the available medical or surgical treatments [8]. Because the attack can be prevented by medication, early prediction of epileptic seizures ensures adequate time before it actually occurs [9]. Young children are susceptible to epilepsy, birth defects, inherited conditions, common diseases like meningitis, and even uncontrollable fevers. Accidents may result in seizures in adults by damaging brain cells. In older persons, heart attacks and trauma are the leading causes of seizures [10]. An epileptic patient may feel amnesia, amnesia, mild melancholy, and persistent headaches as a result of their frequent seizures. It results in abnormal body movements and possibly death. About 70% of those who suffer from epileptic seizures are adults, and 30% are children .Therefore, in order to fulfill the aims of personalized medicine, it is required to automate the detection of epilepsy by identifying the aberrant EEG condition. Precision medicine, another name for customized medicine, is a medical idea in which groups of people are separated and medical decisions, treatments, and/or medications are tailored to each patient based on their expected response or risk of disease [11]. Patients with epilepsy experience seizures, which are events brought on by excessive electrical impulses that are sent out by the brain's nerve cells, an epileptic seizure is "a transitory sign of excessive or synchronized neuronal activity in the brain." Two or more of these spontaneous seizures constitute epilepsy [12]. As indicated in figure 1, there are two categories of epileptic seizures: partial seizures and focal and generalized seizures. There are two types of partial seizures: simple-partial and complex-partial. In the simple-partial, a patient retains awareness but struggles to speak clearly. The term "focal impaired awareness seizure" refers to a condition in which a person experiencing a complex-partial becomes disoriented about their surroundings and begins acting erratically, such as chewing and mumbling. On the other hand, generalized seizures swiftly disrupt entire brain networks and cause damage to all brain areas. There are many different types of generalized seizures, although they can be broadly categorized into convulsive and non-convulsive seizures [3, 24, and 25]. An electroencephalogram (EEG) is a recording of the electrical signals produced by the brain, with the help of electrodes placed on the scalp. Currently, epilepsy is widely diagnosed by monitoring the occurrence of seizures. However, the seizure events are infrequent [13]. EEG (Electroencephalogram) signals can be used to record abnormal brain activity that begins prior to the onset of a seizure [14]. To identify epileptic seizures, a number of screening methods have been developed, including MRI, EEG, magnetoencephalography (MEG), and positron emission tomography (PET). EEG signals are popular because they are affordable, portable, and exhibit distinct patterns in the frequency domain. It takes a lot of time and effort to diagnose epilepsy using EEG data since a neurologist or epileptologist must carefully examine the signals. Creating a computer-based diagnostic may help to solve these issues [15]. In order to diagnose epilepsy, clinicians commonly analyze EEG data, and wearable EEG devices have long been investigated for seizure prediction [16]. In hospitals, EEG is the gold standard method to identify all seizure types, but there is currently no effective wearable EEG that can transfer this technology for extremely long-term monitoring at home [17]. Traditionally, a standard electroencephalogram (EEG), which is often a 20-minute recording of the patient's brain waves, is used to assess suspected seizures. It is doubtful that genuine events are recorded during a regular EEG because of its brief duration. Regular EEGs may capture spikes, sharp waves, or spike-and-wave complexes, which are interictal epilepsy features [18]. An EEG signal is band-limited in frequency (0.1-60 Hz) and is modeled and classified into these five rhythmic waves. The five rhythmic waves that make up an EEG signal are delta (0.1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-60 Hz) waves that capture various brain activities [19]. The four typical brain states of an epileptic seizure occurrence are shown in figure 2. the preictal state, which is a state that appears prior to the onset of the seizure, the ictal state, which starts with the onset of the seizure and ends with an attack, the postictal state, which begins after the ictal state, and the interictal state, which begins after the postictal state of the first seizure and ends before the start of the preictal state of a subsequent seizure. Additionally, the beginning of the preictal state can be used to predict seizures [9, 24, and 26]. In order to diagnose epilepsy in clinical settings, neurophysiologists with experience must visually analyze electroencepaholographic (EEG) recordings to spot recognizable interictal and ictal activity patterns [20]. It takes a lot of time and effort to visually comb through a patient's EEG data to find epileptic seizures, which are often recorded over a few days. Additionally, in order to identify epileptic activity, a professional must thoroughly examine the EEG recordings. A trustworthy automatic categorization and detection system would guarantee objectivity, facilitate treatment, and considerably enhance epilepsy diagnosis, long-term patient monitoring, and therapy [21]. The time-consuming manual seizure identification process and potential for human mistake prompted the researchers to look into novel ways to predict seizures using epileptic EEG and artificial intelligence. Deep learning (DL) and machine learning (ML) are becoming more popular in the analysis of EEG signals [22]. Machine learning is a modern method for seizure prediction [23]. One of two categories applies to the seizure detection process: single-channel or multichannel. In a single-channel approach, a strong channel or signal that is around the seizure origin is chosen based on specific metrics, such as local variance. Better outcomes in the seizure detection procedure can be obtained by combining data from many channels using certain data fusion techniques [24]. For the analysis of EEG signals, the wavelet transform techniques to time-frequency estimation are frequently appealing. For instance, the discrete wavelet transform (DWT) technique has been used to extract characteristics from EEG data. It is a traditional time-frequency analysis technique akin to the short time Fourier transforms [25]. To successfully handle the patient-independent seizure prediction challenge, research into deep learning systems that can learn from data from several subjects is important [26]. Machine learning-based methods have been created to detect abnormal patterns in the EEG data during seizures thanks to developments in IoT-based data collection [27]. PCA with neural network has been proposed for seizure detection. For the categorization of EEG signals, Wavelet transform, PCA, ICA, and linear discriminant analysis (LDA) using SVM have all been reported [28]. In order to identify epileptic seizures, the machine learning system examines the aberrant time-frequency domain properties of the EEG data of individuals with epileptic seizures. The advancement of seizure prediction and location has been aided recently by seizure detection [29]. EEG data patterns would perform very poorly when tried to be classified directly by feeding sampled waveforms into classifiers. However, the curse of dimensionality, or sparsely dispersed data over the high dimensional feature space, causes a significant decline in classifier performance [30].



International Journal of Academic Health and Medical Research (IJAHMR)
ISSN: 2643-9824
Vol. 6 Issue 12, December - 2022, Pages: 41-53

Fig. 1: Types of Seizure and its sub-types.

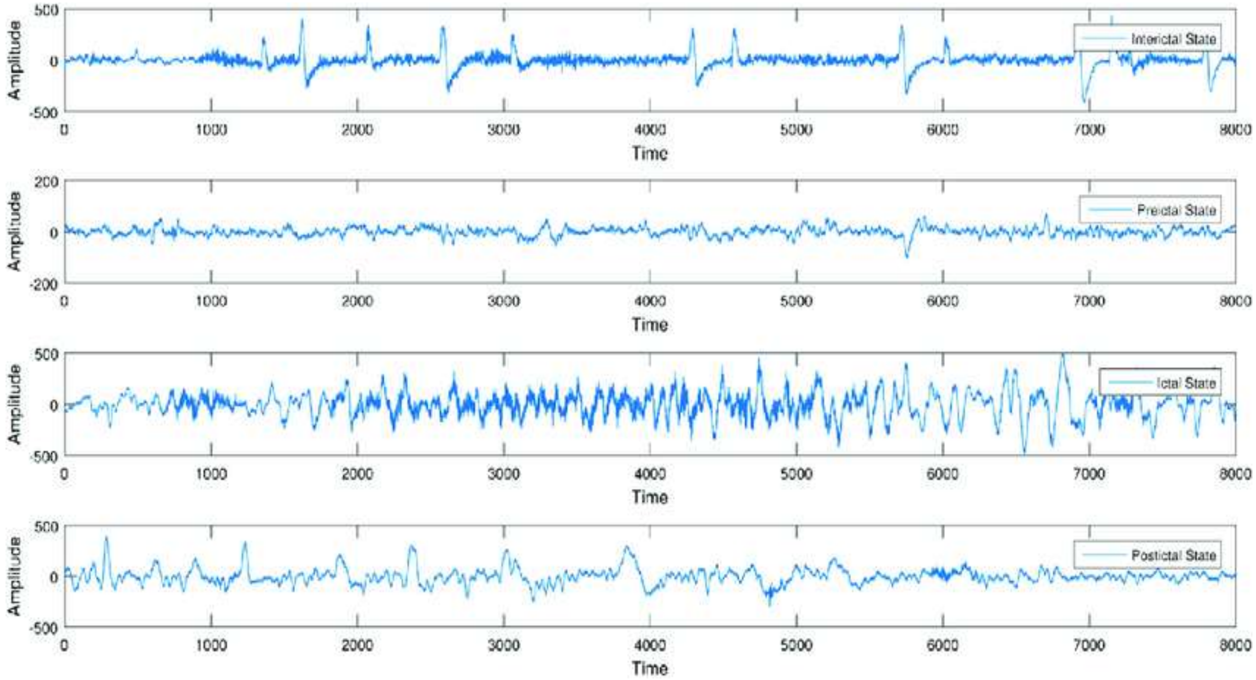


Fig. 2: States of epileptic seizure.

In this paper, preprocessing is done to find missing values, check and remove duplicate values, checking prevalence of target class etc., then features scaling is performed using StandardScaler and finally splitting the dataset such as training data, testing data and validation data. Here 13 different classification algorithms are used to build the system, like- Gaussian Naïve Bayes, Bernoulli Naïve Bayes, Random Forest, K-Nearest Neighbors, Support vector machine, Logistic Regression, Bagging Classifier, AdaBoost Classifier, Gradient Boosting Classifier, stochastic gradient descent (SGD) Classifier, Multi-layer Perceptron (MLP) Classifier, XGBoost Classifier, Decision Tree Classifier are used for increasing overall accuracy.

The rest of the paper is organized as follows: Section 2 provides an overview of related work in seizure prediction. Section 3 describes the dataset used, proposed methodology. Section 3 provides comparison of results our proposed epileptic seizure prediction methods. Section 4 contains the conclusion of the research work and the future work.

2. Related Work

The ongoing study of artificial intelligence in recent years has accelerated the growth of smart health. As a hotspot of smart health, seizure prediction can lessen patients' suffering and ensure their safety. Seizure prediction methods, which are crucial for enhancing recognition outcomes, have a strong emphasis on feature extraction and classification [23].

Sanguk Ryu et al. [1], has showed that the CHB-MIT dataset used in the proposed method consists mostly of pediatric patients. Using scalp EEG data, we have developed a novel deep learning hybrid model called DenseNetLSTM for predicting patient-specific epileptic episodes. The DenseNet technique, which extends the current CNN problem suggested in this paper, increases computational efficiency, and improves network information flow. The long-term temporal properties of the EEG data are also learned by the network using LSTM.it needs to be extensively tested with more EEG data.

Syed Muhammad Usman et al. [8], provided a technique to use the CHB-MIT dataset. It is a dataset of 24 scalp EEG recordings from patients ranging in age from 2 to 22. Using the "edf-read" function, these signals are first transformed into mat files. EEG signals are subjected to a Butterworth bandpass filter to eliminate power line and baseline noise. Short Time Fourier Transform (STFT) is used after noise removal to boost signal to noise ratio and transfer signals from time domain to frequency domain. A non-overlapping window of 30 seconds is chosen. Then they utilized convolutional neural networks to extract features (CNN). As these traits are derived with the aid of class information taken into account, they provide superior interclass variance. After feature extraction from CNN, fully linked layers were swapped out for SVM. CNN is utilized to extract the features, and Support Vector Machine is used to classify the interictal and preictal segments (SVM). Nevertheless, there is still room for development in several areas. If preprocessing is improved in the future to improve signal to noise ratio. Many parameters need to be learned when utilizing deep learning techniques for feature extraction and/or classification.

Syed Muhammad Usman et al. [9], made use of a dataset from CHB-MIT that was freely accessible online. By putting 23 electrodes on the scalps of 22 participants, the dataset was collected. To increase the SNR, they first preprocessed the data in two rounds, converting the 23 channels of EEG signals into a single signal known as a surrogate channel. Empirical mode decomposition (EMD) has been applied to the surrogate channel as part of the second preprocessing stage to further boost SNR. Both time- and frequency-domain features have been derived. In contrast to spectral features, which have been extracted in a frequency domain, statistical features have been extracted in a time domain. Therefore, Support Vector Machines are used for classification to separate testing data into preictal and interictal stages. To boost the sensitivity of seizure prediction, the preprocessing of the EEG signal can be further improved in the future. Other preprocessing techniques, such as hybrid preprocessing techniques and those with adaptive window sizes, can be tested.

John Thomas et al. [13], At Massachusetts General Hospital (MGH), Boston, the data was collected using the International 10-20 electrode system for this investigation. They describe the initial findings of an automated EEG classification system designed to differentiate between EEGs with and without IEDs. There are three primary modules in it: The pre-processing module transforms the data from the EEG recording equipment into the input format for the EEG classification system. The following tasks are completed by this module: 1) Reducing the data's sample rate to 128 Hz. 2) Filtering to eliminate baseline drifts and power-line interference 3) Setting up the montage (here we apply CAR montage). EEG-level classification and waveform-level classification (CNN) (SVM). They took 20 characteristics, or the percentage of CNN outputs in 20 different intervals of equal size, from the output of CNN. Based on p-values, we chose the best attributes. To examine each feature's capacity to distinguish between EEGs with and without IEDs, a two-sample t-test was run. Selected intervals for the CNN output portion were [0.45-0.5, [0.55-0.6, [0.6-0.65, [0.65-0.7, [0.7-0.75, [0.75-0.8, [0.8-0.85, and [0.9-0.95). The EEG-level classification SVM is used with this 8-dimensional feature vector. We chose 0.4 as the SVM output threshold for recognizing EEGs with IEDs. We want to reduce the false detections in their future work by using artifact rejection. Additionally, we want to customize the preprocessing module to work with other montages and EEG recording devices.

Syed Muhammad Usman et al. [14], tested their suggested strategy on 22 participants from the CHBMIT dataset of scalp EEG signals. To address the issue of class imbalance, they used generative adversarial networks to generate preictal samples in the preprocessing section. Automated features were extracted using three-layer convolutional neural networks, and classification between preictal and interictal states was carried out using long short-term memory units. Their suggested approach is adaptable to diverse datasets. In this work, they have concentrated on deep learning approaches, but in the future, it may be possible to increase the average anticipation time by combining deep learning and machine learning techniques. Their suggested technique can be used on real-time EEG signal recordings for seizure prediction in the future.

A. T. Tzallas et al. [18], It makes use of an online-accessible EEG dataset that contains recordings from both epileptic and healthy participants. The dataset consists of five subsets (labeled Z, O, N, F, and S), each of which has 100 single-channel EEG segments with a duration of 23.6 seconds. This study proposes a time-frequency analysis-based approach for analyzing EEG signals. The energy distribution in the time-frequency plane is first represented by several features that are extracted for each segment of the EEG signals after they have undergone time-frequency analysis. These features are then sent into an artificial neural network (ANN), which determines whether to classify the EEG segments as seizure-related or not. However, after visual investigation, a few other artefact kinds were eliminated from this database. Due to this constraint in the evaluation of their technology, additional testing in actual clinical settings is necessary to fully realize its potential. Another drawback is that high-frequency components (those with frequencies greater than 40 Hz) were not measured or considered in the current study; the use of high-frequency components, such as gamma activity, and their significance for epileptic seizure detection will be covered in a future communication.

Dragoljub Gajic et al. [21], The EEG data that were used were a portion of the EEG data that Dr. Ralph Andrzejak from the University of Bonn's Epilepsy Center made available for both epileptic and normal participants. The automatic classification of EEG signals for the identification of epileptic seizures using wavelet transform and statistical pattern recognition is described in this work. Three basic phases make up the decision-making process: (a) feature extraction based on the wavelet transform; (b) feature space dimension reduction using scatter matrices; and (c) classification using quadratic classifiers. On EEG data sets from three subject groups—healthy participants, epileptic subjects during a seizure-free period, and epileptic subjects experiencing a seizure—the proposed methodology was used.

Biao Sun et al. [22], Using the Boston Children's Hospital-MIT scalp EEG public datasets, they assess the proposed approach. Single-domain information input (time domain, frequency domain, etc.) in seizure prediction research neglects some signal information. To properly extract the signal's relevant information, we present in this research a novel deep learning framework called channel attention dual-input convolutional neural network (CADCNN). Short-time Fourier transform (STFT)-extracted spatial-temporal characteristics are given to the CADCNN along with the raw EEG signals for additional feature extraction. CADCNN can learn accurate and distinct representations of EEG signals and improve the temporal, spectral, and spatial information usage capabilities by fusing two inputs from separate domains and combining channel attention. To achieve high accuracy seizure prediction, we anticipate continued refinement of our technique in further seizure prediction research and its application to more datasets.

Yuan Zhang et al. [23], The CHB-MIT EEG dataset contains scalp EEG (sEEG) recordings of 23 individuals with medically untreatable focal epilepsy, and this dataset served as the source for the EEG data used in this study. In this paper, a novel approach to seizure prediction is put forth using convolutional neural networks and common spatial patterns (CSP) (CNN). First, to address the issue of trial imbalance between the two states, segmented pre-ictal signals are combined to create artificial pre-ictal EEG signals based on the actual ones. The distinguishing features can be extracted in both the time domain and the frequency domain using a feature extractor that uses CSP and wavelet packet decomposition. It can shorten training time while increasing overall accuracy. To distinguish between the pre-ictal state and the inter-ictal state, a shallow CNN is then applied.

Lina Wang et al. [25], using a public EEG database at the University Hospital Bonn in Germany, the proposed approach is examined. In this work, they present a paradigm for an automatic epilepsy diagnosis that combines multi-domain feature extraction with nonlinear EEG data processing. First, the wavelet threshold approach is used to pre-process EEG signals and remove artifacts. Using information theory, they extract representative features from the time, frequency, time-frequency, and nonlinear analysis domains. Based on the clinical interest, these features are then extracted in five frequency sub-bands, and the original feature space's dimension is then decreased using both a principal component analysis and an analysis of variance. For the purpose of detecting epileptic seizures in EEG signals, various classifiers are used to identify and assess the best combination of the retrieved features. In order to assess their suggested method and further establish the significance of the discriminative features found in our work, they intend to study larger databases.

Umar Asif et al. [27], the largest publicly accessible dataset for seizure type categorization in the world, TUH EEG Seizure Corpus (TUH-EEGSC) was made available by Temple University for epilepsy research. SeizureNet, a deep learning system that trains multi-spectral feature embeddings using ensemble architecture for cross-patient seizure type classification, is introduced in this paper. They assessed SeizureNet's performance using the recently released TUH EEG Seizure Corpus (V1.4.0 and V1.5.2). Additionally, they demonstrate that for applications with memory constraints, the high-level feature embeddings learned by SeizureNet significantly increase the accuracy of smaller networks through knowledge distillation. Future research will focus on the merging of data from wearable sensors and videos for multi-modal seizure type classification in epilepsy monitoring equipment used in real-world settings.

3. Proposed Methodology

This section provides a detailed description about the proposed methodology of this model. The proposed model is illustrated and it consists of five main parts.

3.1 Dataset and Preprocessing

3.1.1 Dataset Collection

The CHB-MIT dataset used in the paper is a scalp EEG recording measured from 23 pediatric patients at Children's Hospital Boston, which is a public dataset and is available with open access at PhysioNet.org.

3.1.2 Preprocessing

The original dataset used in the paper is divided into five folders, each containing 100 ".edf" files, each representing a single subject or person. Each file contains a 23.6-second recording of brain activity. The dataset is created by extracting data points from the file ".edf" with Python's mne module. The corresponding time-series dataset has 4097 preictal and ictal data points. Each data point represents the value of the EEG recording at a specific point in time. So we have a total of 500 people, each with 4097 data points for 23.5 seconds.

Each chunk contains 178 data points for 1 second, and each data point is the value of the EEG recording at a different point in time. So we now have 11,500 pieces of information (row), each of which contains 178 data points for 1 second (column), and the last column represents the label y 1,2,3,4,5. Column 179 contains the response variable, y, and the explanatory variables, X1, X2,..., X178. The dataset is then converted to ".csv" format for subsequent processing.

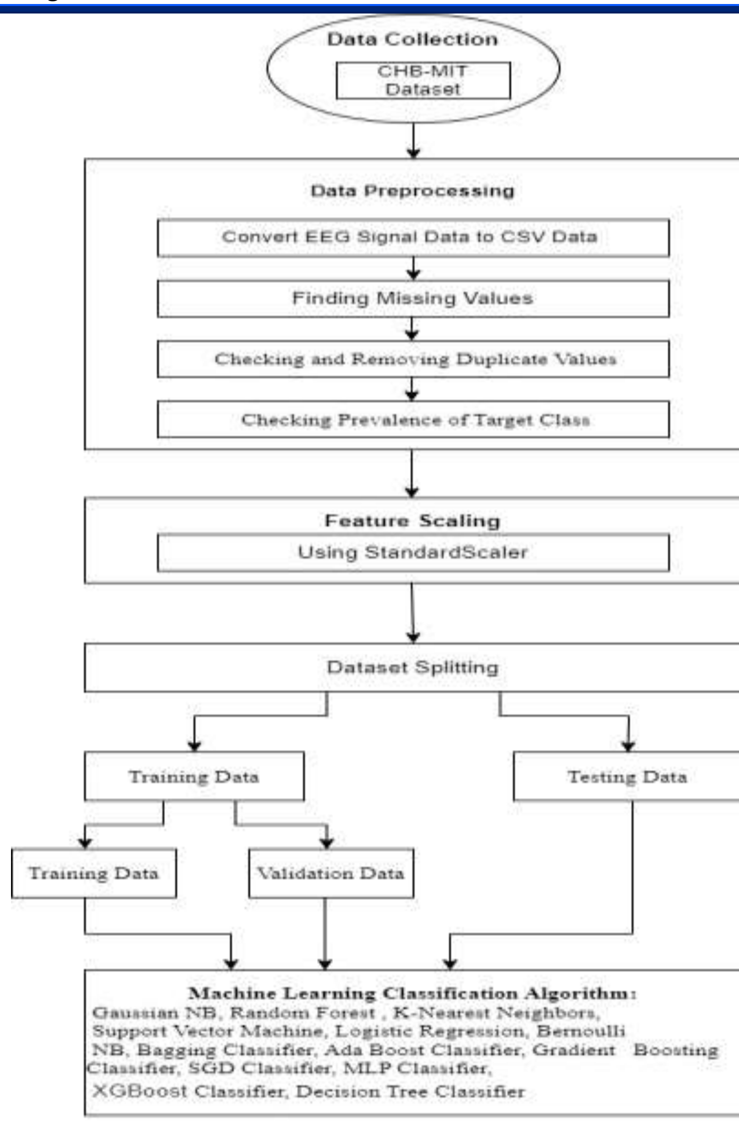


Fig. 3: Proposed Architecture.

Table 1: Different classes in the target attribute and their meaning.

Class	Meaning
1	recording of seizure activity
2	the EEG signal are taken from the area where the tumor was located
3	the EEG signal are taken from the healthy brain area
4	the EEG signal was taken when the patient had their eyes closed
5	the EEG signal was taken when the patient had their eyes open

Table 1 depicts the target attribute classes and their meaning in the dataset. All subjects in classes 2, 3, 4, and 5 did not experience an epileptic seizure. Only class 1 subjects have epileptic seizures. Data points are loaded and preprocessed as data frames using Python's pandas library.

3.2 Exploration of Data

The prevalence rate must then be calculated, which is done by dividing the percentage of samples that are positive overall by the percentage of patients in our dataset who are experiencing a seizure. We have a 20% prevalence rate. Knowing this is helpful when balancing classes and assessing our model using the "lift" metric.

International Journal of Academic Health and Medical Research (IJAHMR)
ISSN: 2643-9824
Vol. 6 Issue 12, December - 2022, Pages: 41-53

3.3 Splitting Data and Building Training/Validation/Test Sets:

Since all of our features are numerical values of EEG readings, there is no need to perform any feature engineering in order to simply dump our dataset into our machine learning model. It is best to keep the dataset's predictor and response variables separate. Our dataset should now be divided into training, validation, and testing sets. The training sets typically range from 50% to 90% of the core dataset. Depending on the amount of samples in the dataset, and the validation and testing sets is typically the same size. A dataset's sample size determines how many samples we can afford to add to our training set. To ensure that there is no order connected with our samples, we must first shuffle our dataset. So let's split our dataset using the 70/15/15 split that was chosen. Because we want the distributions of our validation and testing sets to be comparable, we will choose to separate them from our training set first. In order to avoid building a model that mistakenly classifies samples as belonging to the majority class, we then want to balance our dataset. Then, train, valid, and test sets are saved as.csv files. We must scale our variables in order to make our models to function properly. We have scaled our variables using the StandardScaler from sklearn library.

3.4 Classification

Finally, the system is built using 13 different classification algorithms, including Gaussian Nave Bayes, Bernoulli Nave Bayes, Random Forest, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Bagging Classifier, AdaBoost Classifier, Gradient Boosting Classifier, Stochastic gradient descent (SGD) Classifier, Multi-layer Perceptron (MLP) Classifier, XGBoost Classifier, and Decision Tree Classifier. There is no single classifier that works best with all datasets. As a result, we trained our model with various classifiers to assess its predictive performance.

4. Result Analysis

The classification of the dataset of epileptic seizures using a variety of classifiers is provided in this section, along with maybe some adjusted classifier parameters. One of the difficulties with the implementation is working with a large dataset that has numerous qualities (features), such as 178. As mentioned in the feature extraction (selection component) section, feature reduction can be applied to the near prediction of elliptic seizure cases using some selected features.

The effectiveness of this machine learning model is assessed using a performance matrix. To assess the performance requirements, Confusion Matrix must be aware of certain criteria, including TP, FP, TN, FN, and TPR, among others. The accuracy, recall, precision, and sensitivity of our suggested approach were 95.88%, 86.91%, 1%, and 1%, respectively.

Based on the following metrics, the effectiveness of the various feature selection techniques, including the classifier, has been assessed:

True positive (TP): The ANN recognizes a seizure segment that the expert identified as a seizure.

True negative (TN): Both the expert and the ANN concur that the EEG pattern does not indicate a seizure.

False positive (FP): the identification of a seizure segment that the expert misdiagnosed as a non-seizure.

False negatives (FN): occur when an expert-identified seizure segment is missed by the ANN.

So, Confusion matrix $=\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$

The following parameters are typically used to assess the performance of classifiers based on the aforementioned metrics:

Area Under the ROC Curve (AUC): AUC measures the complete two-dimensional region beneath the entire ROC curve from (0,0) to (1,1).

Average Accuracy, AA: The percentage of how accurately the model is approximated is defined as accuracy.

Accuracy $= \frac{\mathrm{TP + TN}}{\mathrm{TP + TN + FP + FN}}$ x 100%

Recall, r: The total amount of data points that are actually retrieved is referred to as recall.

Recall $= \frac{\mathrm{TP}}{\mathrm{TP + FN}}$ x 100%

Precision (p): Precision is the ratio of the number of relevant data points to the total number of relevant and irrelevant ones.

Precision = $\frac{TP}{TP + FP}$ x 100%

Specificity (SPE): Specificity is the indicator of how well a classifier can identify non-seizure activity.

Specificity = $\frac{TN}{TN + FP}$ x 100%

Figure 4 demonstrates a comparison of AUC values for various classifiers. The result clearly indicates that Random Forest, Gradient Boosting, and XGboost achieved a comparatively higher AUC value for both the training and validation datasets. The AUC value of a bagging classifier, MLP, AdaBoost, and Decision Tree is higher in the training dataset than in the validation dataset.

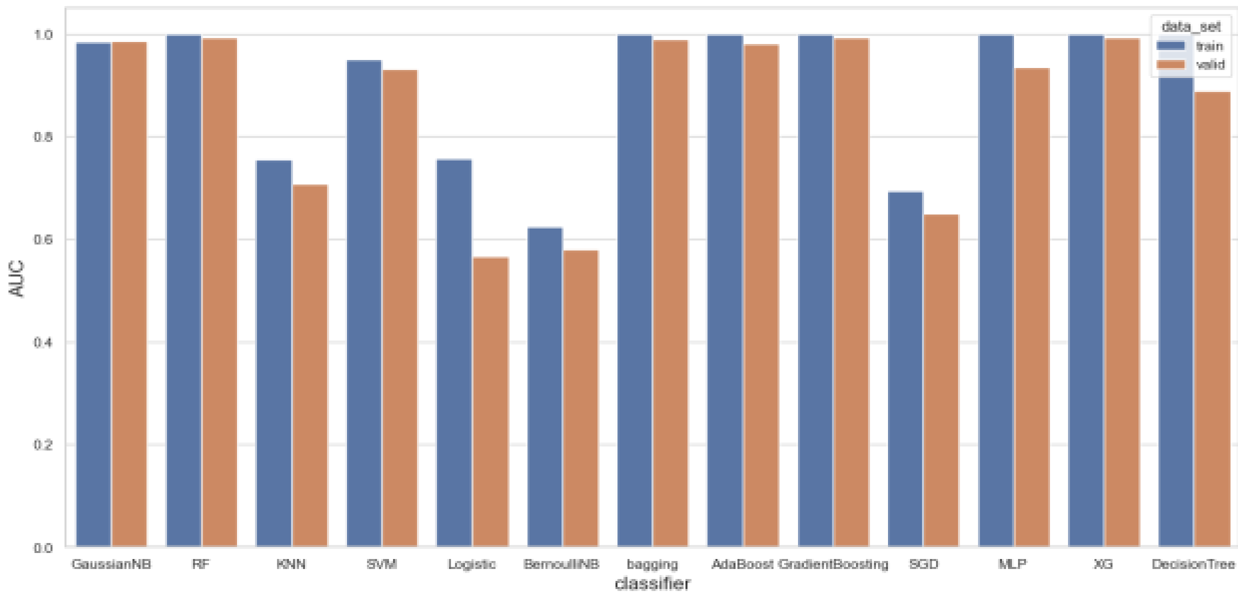


Fig.4: AUC for all Classification Algorithms.

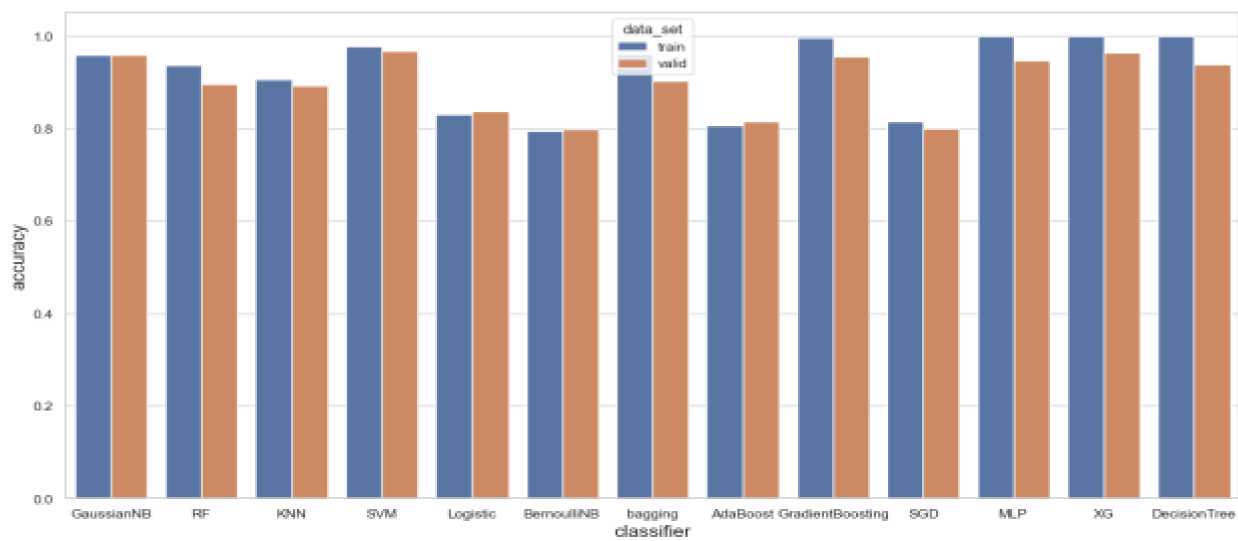


Fig.5: Accuracy for all Classification Algorithms.

Various machine learning techniques are used to evaluate the performance of the model. The most effective classifiers are the Support Vector Machine and Gaussian Naive Bayes. According to figure 5, the classification accuracy of the Support Vector Machine is 95.88%, that of the Gaussian Naive Bayes is 95.30%, and that of the SGDClassifier with the poorest performance is 79.07%.

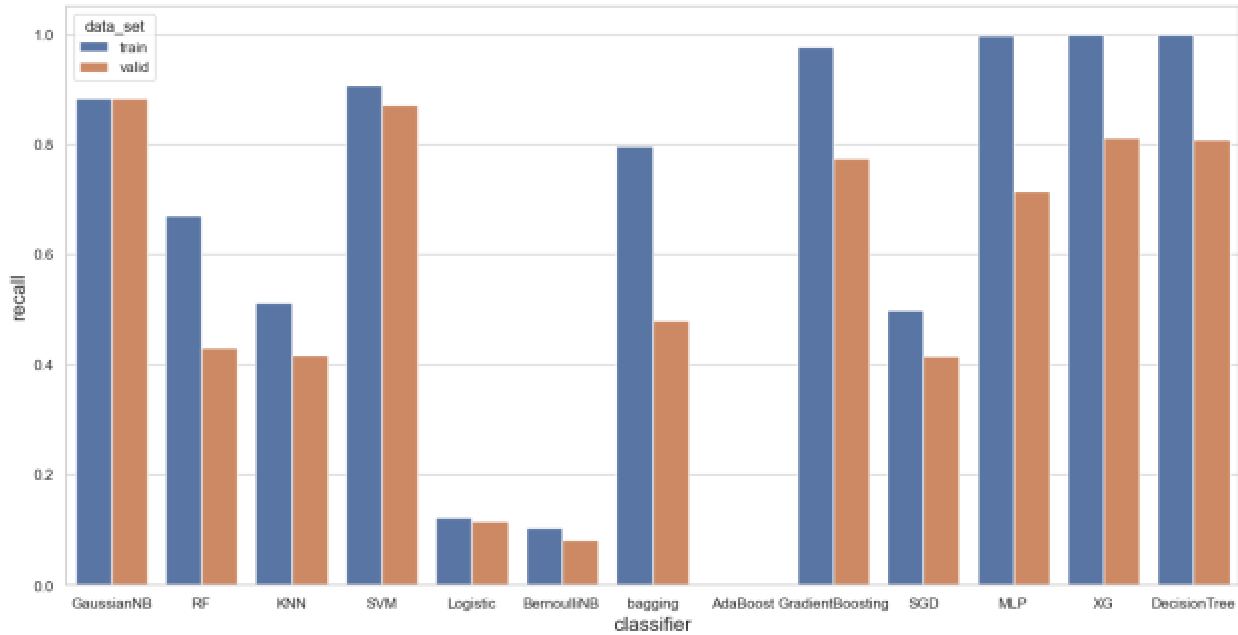


Fig.6: Recall for all Classification Algorithms.

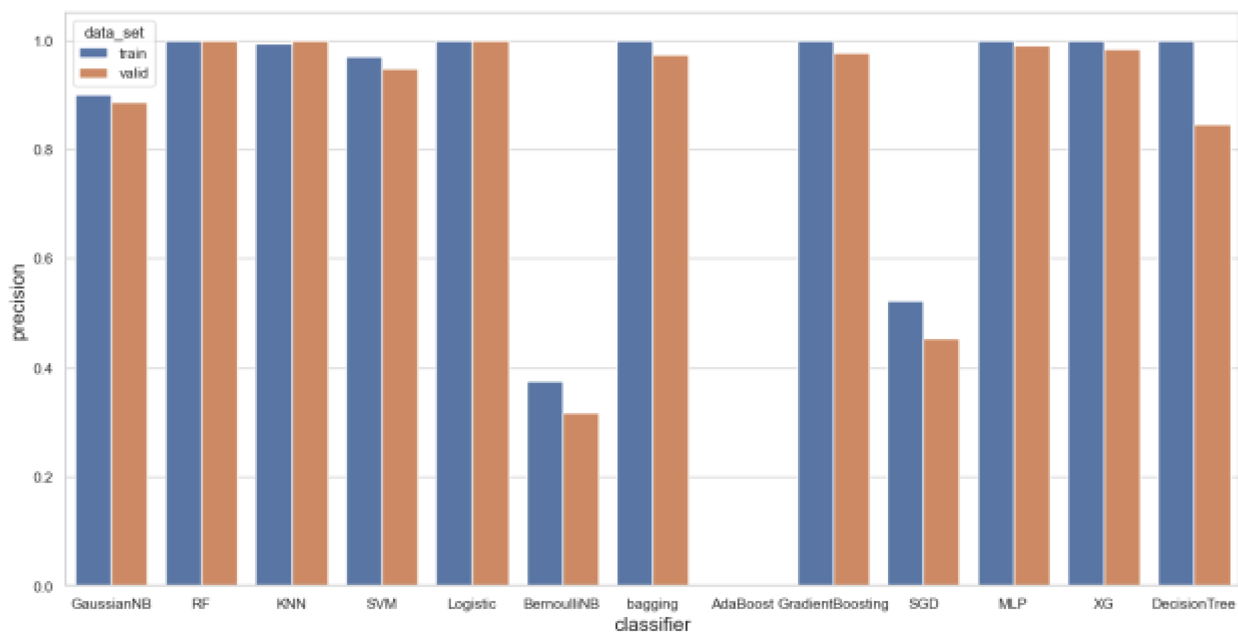


Fig.7: Precision for all Classification Algorithms.

International Journal of Academic Health and Medical Research (IJAHMR)
ISSN: 2643-9824
Vol. 6 Issue 12, December - 2022, Pages: 41-53

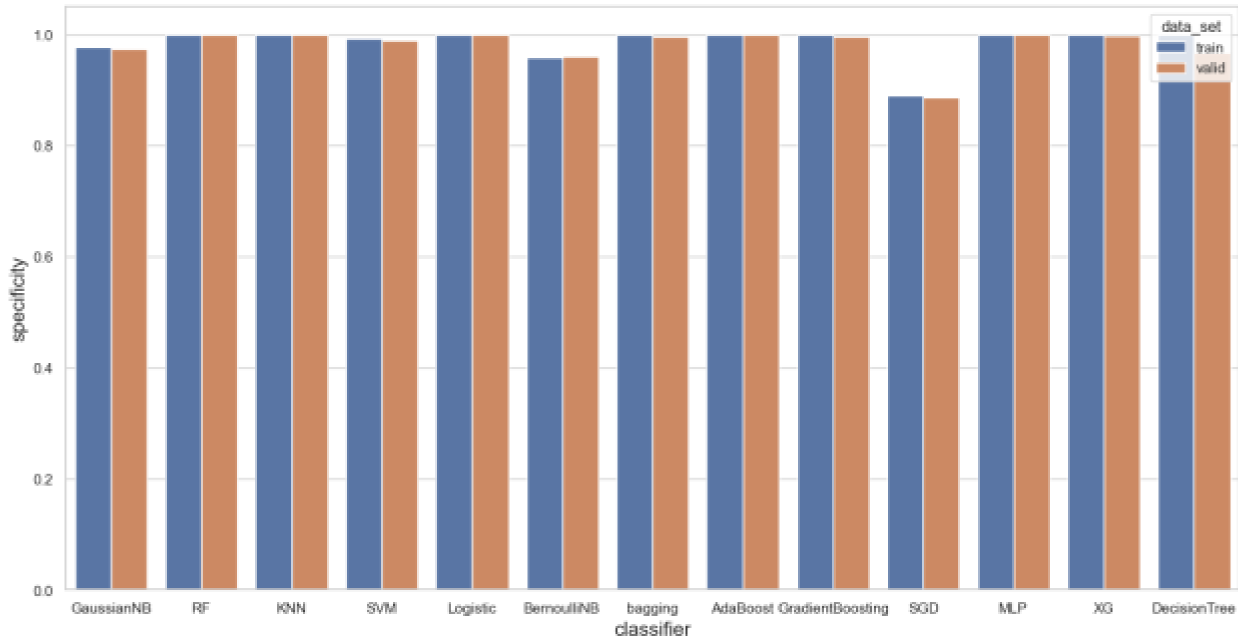


Fig.8: Specificity for all Classification Algorithms.

Table 2. Performance of several classifiers in terms of score, accuracy, recall, precision and specificity.

	Model	Score	Accuracy	recall	precision	specificity
2	Support Vector Machines	0.958841	95.884058	0.845930	0.941748	0.986966
0	GaussianNB	0.953043	95.304348	0.869186	0.892537	0.973932
10	XGBClassifier	0.949565	94.956522	0.770349	0.970696	0.994207
7	GradientBoostingClassifier	0.943768	94.376812	0.732558	0.980545	0.996379
11	DecisionTreeClassifier	0.938551	93.855072	0.843023	0.847953	0.962346
9	MLPClassifier	0.932754	93.275362	0.686047	0.967213	0.994207
5	BaggingClassifier	0.894493	89.449275	0.476744	0.987952	0.993483
1	RandomForest	0.878841	87.884058	0.392442	1.000000	1.000000
3	Logistic Regression	0.822609	82.260870	0.110465	1.000000	1.000000
6	AdaBoostClassifier	0.813333	81.333333	0.063953	1.000000	1.000000
8	SGDClassifier	0.812174	81.217391	0.447674	0.534722	0.902969
4	BernoulliNB	0.790725	79.072464	0.081395	0.383562	0.967415

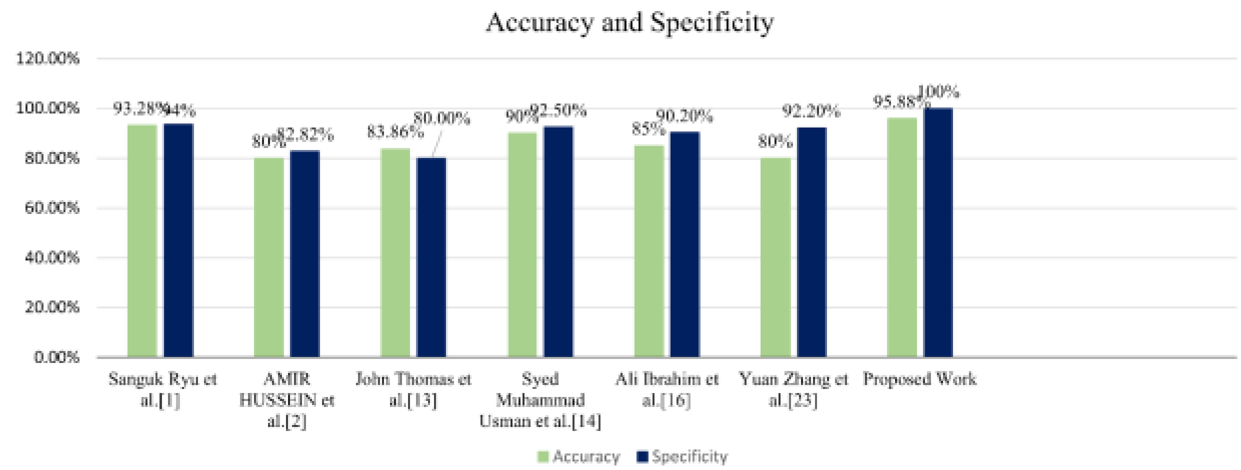


Fig. 9: Comparison of Accuracy and Specificity with other Existing Systems.

Figures 6, 7, and 8 compare the precision, recall, and specificity of various classifiers. The recall values of Gaussian nave Bayes and Support Vector Machine are higher than those of the other classifiers, while AdaBoost has a recall value of nearly zero. Random Forest, Logistic Regression, and AdaBoost have lower precision and specificity values than other classifiers. Table 2 summarizes all of the results shown in figures 5, 6, 7, and 8. Figure 9 compares accuracy and specificity to other existing systems that is represented in a bar chart. The model's accuracy and specificity are higher than those of the other systems in our proposed work.

5. Conclusion and Future Work

The purpose of this work is to detect epileptic seizures in a person with highest accuracy and specificity. For this, we used efficient, well-known supervised learning algorithms and statistical machine learning algorithms. It obtained 95.88% accuracy, 86.91% recall, 1% precision, and 1% sensitivity.

- Here are some key points about the future development trend of EEG-based epilepsy detection methods.

1. Seizure forecasting and localization remain one of the strategies for detecting epilepsy's future development directions. Seizure prediction can significantly enhance the quality of life for epilepsy patients, and non-invasive epileptic focus localization can speed up and reduce the cost of epilepsy diagnosis for physicians.

2. As machine learning advances, more and more new methods will be applied to feature extraction and the Hybrid Ensemble Classification Algorithm of epileptic EEG signals

Reference:

[1] Ryu, S., & Joe, I. (2021). A hybrid DenseNet-LSTM model for epileptic seizure prediction. Applied Sciences, 11(16), 7661. https://doi.org/10.3390/app11167661

[2] A. Hussein, M. Djandji, R. Mahmoud, M. Dhaybi and H. Hajj, "Augmenting DL with Adversarial Training for Robust Prediction of Epilepsy Seizures", ACM Transactions on Computing for Healthcare, vol. 1, no. 3, pp. 1-18, 2020. Available: 10.1145/3386580 [Accessed 14 October 2022].

[3]Siddiqui, M. K., Morales-Menendez, R., Huang, X., & Hussain, N. (2020). A review of epileptic seizure detection using machine learning classifiers. Brain Informatics, 7(1). https://doi.org/10.1186/s40708-020-00105-1

[4] Ahmad, I., Wang, X., Zhu, M., Wang, C., Pi, Y., Khan, J. A., Khan, S., Samuel, O. W., Chen, S., & Li, G. (2022). EEG-based epileptic seizure detection via machine/Deep learning approaches: A systematic review. Computational Intelligence and Neuroscience, 2022, 1–20. https://doi.org/10.1155/2022/6486570

[5] Abdelhameed, A., & Bayoumi, M. (2021). A deep learning approach for automatic seizure detection in children with epilepsy. Frontiers in Computational Neuroscience, 15. https://doi.org/10.3389/fncom.2021.650050

[6] Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., Guo, H., & Xiang, J. (2018). Epileptic seizure detection based on EEG signals and CNN. Frontiers in Neuroinformatics, 12. https://doi.org/10.3389/fninf.2018.00095

[7] Almustafa, K. M. (2020). Classification of epileptic seizure dataset using different machine learning algorithms. Informatics in Medicine Unlocked, 21, 100444. https://doi.org/10.1016/j.imu.2020.100444

[8] Muhammad Usman, S., Khalid, S. and Aslam, M.H. (2020) "Epileptic seizures prediction using Deep Learning Techniques," IEEE Access, 8, pp. 39998–40007. Available at: https://doi.org/10.1109/access.2020.2976866.

[9] Usman, S. M., Usman, M., & Fong, S. (2017). Epileptic seizures prediction using machine learning methods. Computational and Mathematical Methods in Medicine, 2017, 1–10. https://doi.org/10.1155/2017/9074759

[10] Rasekhi, J., Mollaei, M. R. K., Bandarabadi, M., Teixeira, C. A., & Dourado, A. (2015). Epileptic seizure prediction based on ratio and differential linear univariate features. Journal of Medical Signals & Sensors, 5(1), 1. https://doi.org/10.4103/2228-7477.150371

[11] Prasanna, J., Subathra, M. S., Mohammed, M. A., Damaševičius, R., Sairamya, N. J., & George, S. T. (2021). Automated epileptic seizure detection in pediatric subjects of CHB-MIT EEG database—A survey. Journal of Personalized Medicine, 11(10), 1028. https://doi.org/10.3390/jpm11101028

[12] Acharya, U. R., Molinari, F., Vinitha, S. S., and Chattopadhyay, S., Automatic diagnosis of epileptic EEG using entropies. Biomed. Signal Process. Control 7(4):401–408, 2012.

[13] Thomas, J., Comoretto, L., Jin, J., Dauwels, J., Cash, S. S., & Westover, M. B. (2018). EEG classification via Convolutional Neural Network-based interictal epileptiform event detection. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). https://doi.org/10.1109/embc.2018.8512930

[14] Usman, S. M., Khalid, S., & Bashir, Z. (2021). Epileptic seizure prediction using scalp electroencephalogram signals. Biocybernetics and Biomedical Engineering, 41(1), 211–220. https://do i.org/10.1016/j.bbe.2021.01.001

[15] A Shoeibi, M Khodatars, N Ghassemi, M Jafari, P Moridian, R Alizadehsani, M Panahiazar, F Khozeimeh, A Zare, H Hosseini-Nejad, A Khosravi, AF Atiya, D Aminshahidi, S Hussain, M Rouhani, S Nahavandi, and UR Acharya. 2021. "Epileptic seizures detection using deep learning techniques: A review." International Journal of Environmental Research and Public Health, 18, 11.

[16] Ibrahim, A., zhuang, hanqi, Tognoli, E., Ali, A. M., & Erdol, nurgun. (2022). Epileptic seizure prediction based on multiresolution convolutional Neural Networks. https://doi.org/10.21203/rs.3.rs-2121492/v1

[17] Gabeff, V., Teijeiro, T., Zapater, M., Cammoun, L., Rheims, S., Ryvlin, P., & Atienza, D. (2021). Interpreting deep learning models for epileptic seizure detection on EEG signals. Artificial Intelligence in Medicine, 117, 102084. https://doi.org/10.1016/j.artmed.2021.102084

[18] Tzallas AT, Tsipouras MG, Fotiadis DI (2007) Automatic seizure detection based on time-frequency analysis and artificial neural networks. Comput Intell Neurosci 2007:80510

[19] Gupta A, Singh P, Karlekar M (2018) A novel signal modeling approach for classification of seizure and seizure-free EEG signals. IEEE Trans Neural Syst Rehabil Eng 26(5):925–935

[20] Gómez, C., Arbeláez, P., Navarrete, M., Alvarado-Rojas, C., Le Van Quyen, M., & Valderrama, M. (2020). Automatic seizure detection based on imaged-EEG signals through fully convolutional networks. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-78784-3

[21] Gajic D, Djurovic Z, Gennaro SD, Gustafsson F (2014) Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. Biomed Eng Appl Basis Commun 26(2):1450021

[22] Sun, B., Lv, J.-J., Rui, L.-G., Yang, Y.-X., Chen, Y.-G., Ma, C., & Gao, Z.-K. (2021). Seizure prediction in scalp EEG based channel attention dual-input convolutional neural network. Physica A: Statistical Mechanics and Its Applications, 584, 126376. https://doi.org/10.1016/j.physa.2021.126376

[23] Y. Zhang, Y. Guo, P. Yang, W. Chen and B. Lo, "Epilepsy Seizure Prediction on EEG Using Common Spatial Pattern and Convolutional Neural Network", IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 465-474, 2020. Available: 10.1109/jbhi.2019.2933046 [Accessed 15 October 2022].

[24] Paul, Y. (2018). Various epileptic seizure detection techniques using biomedical signals: A Review. Brain Informatics, 5(2). https://doi.org/10.1186/s40708-018-0084-z

[25] Wang, L., Xue, W., Li, Y., Luo, M., Huang, J., Cui, W., & Huang, C. (2017). Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis. Entropy, 19(6), 222. https://doi.org/10.3390/e19060222

[26] Dissanayake, T., Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (2021). Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals. IEEE Sensors Journal, 21(7), 9377–9388. https://doi.org/10.1109/jsen.2021.3057076

[27] Asif, U., Roy, S., Tang, J., & Harrer, S. (2020). SeizureNet: Multi-spectral deep feature learning for seizure type classification. Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-Oncology, 77–87. https://doi.org/10.1007/978-3-030-66843-3_8

[28] Jaiswal, A. K., & Banka, H. (2017). Epileptic seizure detection in EEG signal using machine learning techniques. Australasian Physical & Engineering Sciences in Medicine, 41(1), 81–94. https://doi.org/10.1007/s13246-017-0610-y

[29] Liu, G., Xiao, R., Xu, L., & Cai, J. (2021). Minireview of epilepsy detection techniques based on electroencephalogram signals. Frontiers in Systems Neuroscience, 15. https://doi.org/10.3389/fnsys.2021.685387

[30] Aarabi, A., Wallois, F., and Grebe, R., Automated neonatal seizures detection: A multistage classification system through feature selection based on relevance and redundancy analysis. Clin. Neurophysiol. 117:328–340, 2006.

[31] Gautama T, Mandic DP, Van Hulle MM. Indications of nonlinear structures in brain electrical activity. Phys Rev E 2003;67:046204.