EpiNet: A Hybrid Machine Learning Model for Epileptic Seizure Prediction using EEG Signals from a 500 Patient Dataset

Oishika Khair Esha, Nasima Begum*, Shaila Rahman
Department of Computer Science and Engineering
University of Asia Pacific, Dhaka, Bangladesh

***Abstract*—The accurate prognosis of epileptic seizures has great significance in enhancing the management of epilepsy, necessitating the creation of robust and precise predictive models. EpiNet, our hybrid machine learning model for EEG signal analysis, incorporates key elements of computer vision and machine learning , positioning it within this advancing technological domain for enhanced seizure prediction accuracy. Hence, this research aims to provide a thorough investigation using the Bonn Electroencephalogram (EEG) signals dataset as an alternative method. The methodology used in this study encompasses the training of five machine learning models, such as Support Vector Machines (SVM), Gaussian Naive Bayes, Gradient Boosting, XGBoost, and LightGBM. Performance criteria, including accuracy, sensitivity, specificity, precision, recall, and F1-score, are extensively used to assess the efficacy of each model. A unique contribution is the development of a hybrid model, integrating predictions from individual models to enhance the overall accuracy of epilepsy identification. Experimental results demonstrate notable success, with the hybrid model achieving an accuracy of 99.81%. Performance matrices for both classes demonstrate the hybrid model's epileptic seizure prediction reliability. Visualizations, including ROC-AUC curves and accuracy curves, provide a nuanced understanding of the models' discriminative abilities and performance improvement with increasing sample size. A comparative analysis with existing studies reaffirms the advancement of our research, positioning it at the forefront of epileptic seizure prediction. This study not only highlights the promising integration of machine learning in medical diagnostics but also emphasises areas for future refinement. The achieved results open avenues for proactive healthcare management and improved patient outcomes.**

Keywords*—*Epilepsy; seizure prediction; computer vision; hybrid model; electroencephalography; bonn dataset; proactive healthcare

I. Introduction

In this study, we introduced EpiNet, a novel hybrid machine learning model, designed to significantly advance epileptic seizure prediction using EEG signals. EpiNet uniquely combines the strengths of various advanced machine learning techniques, resulting in a model that not only outperforms existing single-model systems in accuracy but also addresses critical challenges in seizure prediction such as high variability in EEG signals and the need for reducing false positives. Our model stands out in its ability to integrate complex patterns from a large dataset of 500 patients, providing a more robust and reliable prediction mechanism. The introduction of EpiNet represents a pivotal step forward in epilepsy management, promising to enhance patient care through more precise and proactive strategies. Despite the advancements in medical science, epilepsy is a prevalent cerebral disease affecting a substantial portion of the global population [1]. The primary mode of treatment involves the use of medications; however, a considerable proportion of individuals diagnosed with epilepsy have difficulties effectively managing their medication, resulting in a substantial decrease in their overall state of life. For some individuals, the consideration of respecting portions of the brain becomes a drastic option to eliminate the seizure focus, yet this measure does not guarantee freedom from seizures. In response to the limitations of existing treatments, there has been a burgeoning interest in the development of clinically effective seizure prediction systems. A successful prediction method could offer timely warnings to patients, allowing for preventive measures or interventions such as electrical stimulation, medication release, or cooling of the seizure focus area. Despite early attempts to predict seizures, the scientific and clinical communities have faced persistent challenges, partly attributed to the absence of a detailed definition of the preictal stage, the critical period preceding a seizure. The unpredictable nature of seizures adds a layer of complexity and concern for individuals living with epilepsy.

This research endeavour aims to tackle the aforementioned issues by investigating other approaches that might enhance the accuracy of seizure prediction. In our earlier conference paper [2], we conducted an extensive review of machine learning and deep learning approaches for epilepsy diagnosis, identifying critical research gaps such as dataset limitations, preprocessing challenges, and the need for ensemble methods. Building upon these insights, our current paper addresses these challenges head-on. Furthermore, our study delves into the critical aspects of data, feature selection, and model selection, drawing on the recommendations outlined in our previous work. Notably, we emphasise the application of ensemble learning, an idea proposed in our conference paper, demonstrating its effectiveness in enhancing prediction accuracy and mitigating false alarm rates. This seamless connection highlights the evolutionary progression of our research agenda, from identifying challenges to proposing practical solutions. The aim is to predict upcoming seizures before their occurrence, facilitating prompt management and mitigating associated dangers. The subsequent sections delve into a comprehensive methodology and findings, contributing significantly to the advancement of seizure prediction research.

Fig. 1 provides a visual representation of the basic technique used for the recording of epileptic seizures through EEG.

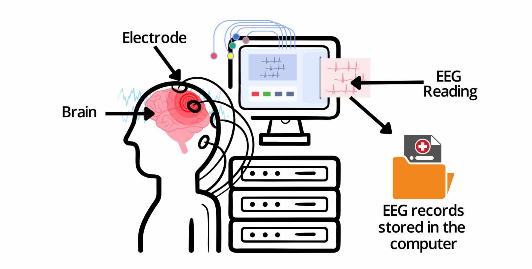

Fig. 1. General methodology for recording epileptic seizures.

During this procedure, electrodes are carefully set on the top of the head in order to assess the electroencephalographic signals, which measure the neural electrical signals produced by the brain. The electrodes are capable of capturing the complex signals produced by neurons in the cerebral cortex.

A. *Key Contributions*

In this study, we introduce an innovative machine learning-based approach to epileptic seizure prediction, labelled EpiNet. While significant strides have been made in the realm of epileptic seizure detection, current methods continue to face challenges, and numerous unresolved issues persist. In light of this, our work aims to address a few pivotal questions, outlined below:

1) Single-Model Limitation in Epilepsy Prediction: Problem: Existing approaches to epilepsy prediction often rely on single machine learning models, limiting the overall accuracy and robustness of the predictions.

Contribution: Proposed a novel hybrid model that amalgamates predictions from diverse models, demonstrating superior performance with heightened accuracy, sensitivity, and specificity compared to individual models. Despite the limited number of existing hybrid models, our approach stands out by consistently outperforming them, underscoring the effectiveness of our tailored combination of models in epileptic seizure prediction.

2) Feature Redundancy and Noise: Problem: The accuracy of prediction models heavily depends on the quality of their features. The presence of redundant or noisy features can compromise the precision of predictions. Identifying and addressing this problem becomes pivotal for enhancing the reliability of epilepsy prediction models.

Contribution: By implementing RFE, the study actively contributes to refining the feature selection process. It goes beyond recognising the issue and presents a strategic solution that decreases noise and improves epilepsy prediction. This work improves prediction model robustness by advancing methodology.

3) Need for High Accuracy and Reliability: Problem: For successful patient treatment, medical diagnostics, notably epilepsy prediction, need great accuracy and dependability.

Contribution: Effectively predicted seizures with 99.81% accuracy using EpiNet.

4) Future Research Directions and Validation: Problem: The application of current research to a variety of situations and real-world healthcare settings often presents difficulties. By pointing out the areas that need attention and development, acknowledging these limitations creates the foundation for a significant contribution.

Contribution: This contribution goes above and beyond just identifying limits by providing a clear path for further study. It takes the lead in resolving the issue by outlining concrete measures to improve the generalizability of the model and to verify its efficacy in actual healthcare settings. This prospective strategy sets the research up to be a driving force behind real improvements in the area of epilepsy prediction.

The remainder of this paper is structured as follows: Section II represents the overview of epilepsy. In Section III we have the literature review. The dataset processing and feature extraction is provided in Section IV. The proposed methodology is described in Section V. Section VI analyzes the results of the conducted experiments. Lastly, Section VII concludes the paper with some future works.

II. Overview of Epilepsy

Epilepsy, an illness that is rather common, has a substantial influence on the lives of millions of people all over the globe. The purpose of this introductory part is to offer a comprehensive viewpoint on the difficulties that epilepsy presents, so laying the groundwork for a more in-depth investigation into the various seizure prediction approaches.

A. *Introduction to Epilepsy*

Epilepsy is characterised by recurrent seizures, affecting a considerable portion of the global population. Despite advancements in medical interventions, a substantial number of epilepsy patients face challenges in symptom management with traditional medications. Patients facing resistance to conventional treatments often endure a severely diminished quality of life. Some explore extreme measures, such as brain resection, in pursuit of relief, yet this drastic approach remains ineffective for a notable proportion of individuals. The Fig. 2 visually represents the intricate activity within the brain during an epileptic seizure.

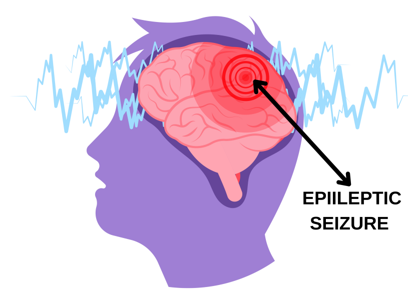

Fig. 2. Illustration of epileptic seizure activity in the brain.

B. *Seizure Types and Symptoms*

Epileptic seizures manifest in various types, including focal and generalised seizures, each presenting unique characteristics. A nuanced understanding of these manifestations is essential for accurate diagnosis and prediction. The diverse manifestations of epileptic seizures encompass various types, each characterised by distinct symptoms. Simple Partial Seizures, or Focal Onset Aware Seizures, affect a specific region of the brain, resulting in altered emotions, sensory perceptions, or movements without loss of consciousness. As depicted in Fig. 3, these seizures may progress to Complex Partial Seizures, or Focal Onset Impaired Awareness Seizures, where consciousness becomes altered, accompanied by involuntary repetitive movements. Notably, these partial seizures can evolve into Generalized Seizures, involving the entire brain.

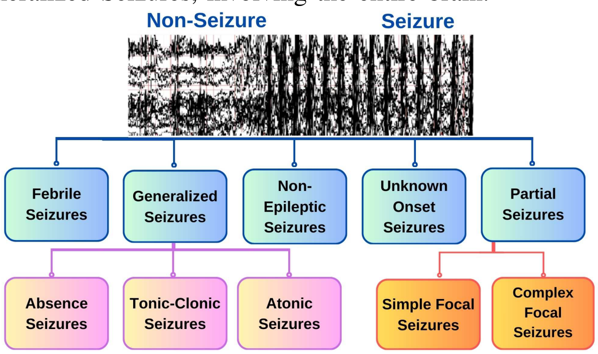

Fig. 3. Types of seizures.

The visualisation in Fig. 3 offers a thorough and inclusive depiction of the development and unique attributes associated with these many kinds of seizures. This visualisation serves to enhance comprehension and facilitate the categorization of epileptic occurrences. Recognising the diverse symptoms accompanying seizures is critical, emphasizing the need for tailored diagnostic and predictive approaches that consider the individualised nature of epilepsy.

C. Motivation for Advanced Seizure Prediction Approaches

The primary motivation behind this research stems from the pressing challenges in epilepsy management, particularly the limitations of current diagnostic tools and treatment strategies. One of the key difficulties lies in the unpredictable nature of seizures, which significantly affects patients' quality of life and complicates treatment planning. Current methods lack the precision and foresight needed for effective seizure management. This gap highlights the necessity for more accurate and timely seizure prognosis, where machine learning approaches hold significant promise. Machine learning's ability to analyze complex EEG data and identify patterns indicative of impending seizures presents a transformative opportunity in epilepsy care with minimum costs. Our research with the EpiNet model is directly motivated by these challenges. We aim to leverage the advanced capabilities of machine learning to enhance seizure prediction accuracy, ultimately leading to more personalized and effective epilepsy management strategies. This approach not only addresses a critical need in epilepsy care but also opens up new avenues for research and treatment methodologies in the field.

III. Literature Review

Kapoor et al. [3] introduce an innovative seizure prediction method using ensemble classifiers and hybrid search optimization, achieving a high accuracy of 96.61% on the CHB-MIT database. Their approach addresses existing limitations, sets a standard for researchers, and explores COVID-19-related data applications for enhanced seizure prediction. Savadkoohi et al.[4] used K-nearest neighbours (KNN) and support vector machine (SVM) prediction models to predict seizures, demonstrating efficiency, reliability, and flexibility across different frequency ranges. The technique has phase information and directionality issues, despite its merits. [5] introduced a spike rate-based seizure detection approach with great accuracy, improving the quality of life for epileptic patients. However, without a complete deep learning method for prediction, performance may decline.

Usman et al. [6] developed generative adversarial networks using LSTM units to improve sensitivity and reduce false positives. However, anticipation time improvement was inefficient. Author in [7] developed a seizure prediction component using deep learning approaches, improving sensitivity and specificity. The approach has limitations, including a low Signal-to-Noise Ratio (SNR) and dependency on several factors. Emara et al. [8] developed a technique to identify abnormalities in multi-channel EEG signals with high prediction accuracy. This method relied on samples, which was a drawback. Wang et al. [9] pioneered a CNN and DTF-based seizure prediction system, offering potential benefits for epilepsy patients in closed-loop therapy. However, the method was noted for its effectiveness despite time constraints. In a recent study [10], a DWT-transformed EEG data approach, coupled with a DenseNet-LSTM hybrid model, demonstrated improved seizure prediction accuracy, outperforming prior methods on the CHB-MIT scalp EEG dataset. Notably, this integration of Discrete Wavelet Transform and hybrid models signifies progress in the field. Viana et al. [11] explored remote subcutaneous EEG monitoring for individualized seizure forecasting.

Behnoush et al. [12] meticulously assessed machine learning algorithms for predicting seizures, emphasizing critical patient identification in emergency department data. A study enhanced signal-to-noise ratio by combining 23 EEG channels into one [13]. Ouichka et al. [14] delved into the challenges of predicting epileptic episodes using iEEG data. Deep learning models, including 3-CNN and 4-CNN, achieved 95% accuracy in autonomous seizure prediction, surpassing previous approaches. In [15], a successful seizure prediction method employs deep learning on preprocessed scalp EEG signals, achieving 92.7% sensitivity and 90.8% specificity in 24 patients. Meanwhile, [16] introduces a novel seizure detection approach using sparse representation with the Stein kernel, leveraging old data to simplify and understand new EEG samples.

The study by [17] employed a two-step strategy, utilizing multi-lead EEG samples to train SE-Net for short-term features and LSTM for long-term characteristics. Adversarial learning enhanced the LSTM feature mapping, resulting in a 5% improvement in classification accuracy on the TUH EEG Seizure Corpus and CHB-MIT databases. This study [18] introduces an innovative approach utilizing SLT and VGG-19 neural networks for precise seizure detection, achieving remarkable 100% accuracy in distinguishing seizure and non-seizure events across seven instances. Notably, its effectiveness extends to improved classification in three- and five-class scenarios, outperforming conventional methods. Tested on the CHB-MIT scalp EEG database, the proposed technique attains a notable 94.3% accuracy in distinguishing seizures from non-seizure episodes.

Studies by [19] highlight LSTM and Random Forest's efficacy in epileptic episode prediction with 97% and 98% accuracy. [20] conducts a comprehensive literature review emphasizing the critical role of ML in automating epilepsy diagnosis, discussing feature extraction methods, and advocating for relevant characteristics and classifiers. Researchers discovered a novel method for simultaneous epilepsy prediction in adults and children, leveraging a linear mixed model and recording over 1.2 million seizures [21]. Emphasizing the distinct seizure patterns in different age groups, they underscored the need for early diagnosis and treatment to prevent potential brain damage. In a separate investigation, a researcher achieved 100% accuracy in epileptic episode detection using innovative SVM-PCA methodologies, outperforming traditional algorithms [22]. The Local Binary Pattern (LBP) method, assessing key points in EEG data, proves effective for real-time seizure diagnosis [23]. Toraman et al. [24] successfully distinguish preictal and interictal occurrences using SVM, forecasting seizures up to 33 minutes in advance.

Weighted Majority Voting Ensemble (WMVE) stands out by dynamically evaluating and adjusting each classifier, prioritizing accurate categorization, particularly for challenging data. In comparison to Simple Majority Voting Ensemble (SMVE), WMVE demonstrates superior classification accuracy across diverse datasets [25]. Additionally, support vector machines (SVM), particularly when paired with a radial basis function kernel, emerge as highly effective in EEG signal categorization for epilepsy detection [26]. The proposed approach in [27] accurately predicted seizures with an average lead time of 23.6 minutes, utilizing "optimum allocated techniques" for sample selection. In [28], a three-part procedure combining LSTM, regression-based SNR enhancement, and statistical properties achieved superior results (94% accuracy) on the CHB-MIT dataset. Another study [29] employed three machine learning methods for effective seizure detection.

In one approach [30], a phase-space adjacency graph achieved a 97% success rate, while another method using hypergraph analysis reached 93% accuracy. A third method employing deep learning and CNN in phase-space analysis achieved perfect 100% accuracy. Additionally, an ensemble classifier in epilepsy research demonstrated superior performance with a 90% success rate, outperforming others in the range of 85% to 89.5%. Sharma et al. [31] introduce a novel hybrid method, combining higher-order statistics, sensitivity analysis, and the residual wavelet transform, to assess brain signal frequencies affected by transient events. This approach effectively detects and characterizes non-stationary time series alterations in neural activity across different brain areas.

Researchers in [32] developed a machine learning approach for epilepsy prediction using correlation dimension, which converges quickly due to its simplicity. The model predicts seizures by analyzing EEG spike rates, employing a mean filter to smooth spikes and activating an alert when preictal spike numbers exceed a threshold. The suggested technique in [33] predicts seizure activity with 92% accuracy using the CHB-MIT dataset for all patients. Researchers in [34] advocate using reconstructed phase space (RPS) over raw EEG data for seizure detection, citing improved accuracy rates of 95% for tertiary and 98.5% for binary classification. In [35], a two-layer LSTM model achieves an exceptional 98.14% average accuracy, enhancing the quality of life for epilepsy patients. Additionally, [36] introduces an approach to extract time-frequency features using STFT, addressing CNN and Transformer limitations with an innovative alternating structure for enhanced predictions.

To address the challenge of limited EEG data, the study in [37] employs a deep learning approach, specifically a DCGAN, to generate synthetic EEG data. The proposed method combines transfer learning with popular DL models and utilizes synthetic data for training, showcasing improved epileptic seizure prediction. The study [38] evaluated an EEG-based seizure detection model on CHB-MIT scalp data, demonstrating robustness with sensitivity and specificity values reaching approximately 100%. The updated XG Boost classifier surpassed previous methods, enhancing sensitivity by 0.05% and specificity by 1%. Syed Muhammad Usman et al. [39] created a more sensitive ensemble learning technique for epileptic prediction. Importantly, our technique doesn't use heart rate variability and EEG measurements.

IV. Data Processing and Feature Extraction

In this section, we delve into the details of the dataset employed in our study. We'll walk you through the different categories within the dataset, the distinctive features that shape its structure, and the meticulous steps we took to make sure the data is not only accurate but also well-suited for machine learning purposes.

A. Dataset Description

The Bonn EEG dataset [40], selected for our study, is renowned in epilepsy research for its high-quality and diverse data. This dataset is a benchmark in the field, offering a broad spectrum of EEG signal patterns from various types of epileptic seizures, which is crucial for developing robust and comprehensive seizure prediction models. Its widespread use and proven success in previous studies underscore its reliability and effectiveness. The practicality and accessibility of the Bonn dataset, combined with its ethical soundness, make it an ideal choice for our research, facilitating the advancement of machine learning approaches in epilepsy prognosis. It was collected from the brain activity recordings (EEG) of 500 different individuals. To make the data more manageable for analysis, we divided each 23.6-second EEG recording into 4097 data points. The dataset is neatly organized into 23 segments, each containing a total of 4097 data points. We then mixed things up a bit by shuffling the segments randomly, and within each segment, we have 178 data points. Each of these data points corresponds to a one-second interval in the EEG recording, helping us capture a more detailed snapshot of the brain activity. Consequently, 11,500 rows of data are generated, where each row stands for one second of an individual's EEG recording. A bandpass filter with a frequency range of 0.53–40 Hz and a sampling rate of 173.61 Hz was used in order to do the preprocessing on the dataset first. The preprocessed dataset is made up of electroencephalogram recordings that were taken under a variety of situations, with each condition being assigned a unique class label.

The EEG signals from the Bonn University dataset that represent the five classes are shown in Fig. 4 This picture displays exemplars of the EEG signals that reflect the five

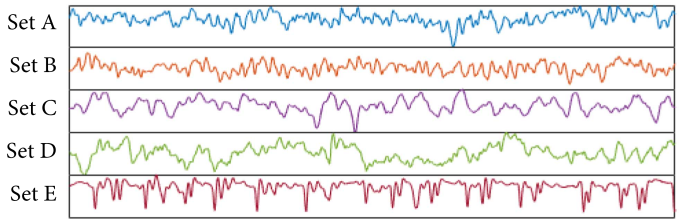

Fig. 4. Data set representing five classes EEG signals.

distinct categories. (a) Set A represents the state of having open eyes, (b) Set B represents the condition of having closed eyes, (c) Set C represents the state of the hippocampal formation during the period between seizures, (d) Set D represents the epileptogenic zone during the period between seizures, and (e) Set E represents the state of having a seizure, also known as the ictal state.

B. Dataset Structure

The Bonn University Epilepsy dataset comprises five sets, labeled A, B, C, D, and E. Each set contains 100 files, representing recordings from different individuals. Each file corresponds to a single subject and has a duration of 23.6 seconds. The EEG signals in each file are discretized into 4097 data points.

TABLE I. DATASET OVERVIEW

Class	Patient Status	Setup	Phase
A	Non-Epilepsy	Surface EEG	Open Eyes
B	Non-Epilepsy	Surface EEG	Close Eyes
C	Epilepsy	Intracranial EEG	Interictal Hippocampal Position
D	Epilepsy	Intracranial EEG	Interictal Epileptogenic Zone
E	Epilepsy	Intracranial EEG	Ictal

Table I displays the dataset with category labels. The dataset is categorised into five distinct classes, each representing various situations. As follows:

- Class A: EEG recorded with the patient's eyes open.
- Class B: EEG recorded with the patient's eyes closed.
- Class C: EEG recorded from the healthy brain area where the tumor was not present.
- Class D: EEG recorded from the area where the tumor was located.
- Class E: Seizure activity.

C. Data Cleaning and Missing Value Handling

For the purpose of ensuring that the data contains no errors, a data cleaning operation was carried out. For the purpose of removing rows from the dataset that were missing values, the dropna() function was used. It was essential to get rid of any data that was erroneous or missing. It made certain that we have trustworthy data.

D. Outlier Detection and Removal

Using the Local Outlier Factor (LOF) technique, we found and removed any strange data points that may have skewed our estimates. Basically a digital investigator. The "LocalOutlierFactor" algorithm from the scikit-learn module proved to be instrumental in resolving the problem. It helped us find the data points that didn't seem to fit and get rid of them so they wouldn't mess up the calculations later on.

E. Feature Scaling and Standardisation

Standardisation of the features was accomplished by using the StandardScaler function that is available in the scikit-learn package. As a component of this, the attributes were normalised by bringing their range to a value of one and with zero serving as the centre of their average. Through the use of a single scale, this step ensured that the features could be compared in a manner that was both accurate and relevant across a variety of characteristic categories.

F. Feature Selection

Selecting high-quality features is like assembling a winning dish in our machine learning adventure. For the purpose of epilepsy prediction, we are interested in the most relevant ones. Therefore, we used a method known as Recursive Feature Elimination (RFE) to determine the most important characteristics. It's as if we had a handy reference (the RFE class from the scikit-learn package) that highlighted the essential components for precise epilepsy prediction.

1) Recursive Feature Elimination (RFE): In the face of information overload, Recursive Feature Elimination (RFE) goes about its business. The process begins with an expansive viewpoint, seeing all characteristics as possible indicators. In order to determine the relative value of each feature, RFE uses a Random Forest classifier as its investigation tool. While RFE iteratively reduces the least significant features, the investigative process starts. Imagine it as a detective selecting the best number of characteristics by sorting through clues and eliminating irrelevant ones. To guarantee that the final collection of features reflects the most important bits of information for the work at hand, RFE's unique technique replicates a seasoned investigator's tactics.

2) Selected Feature Subset: When we applied the Recursive Feature Elimination (RFE) method, it helped us identify a subset of the most crucial features. We selected these variables because they demonstrated a remarkable ability to predict epileptic episodes. The idea behind choosing these particular features was to enhance the predictive performance of our models. By focusing on these key variables and narrowing down the feature space, we significantly improved the models' ability to make accurate predictions of epileptic episodes.

V. PROPOSED METHODOLOGY

In our study, we selected a diverse array of machine learning models, each chosen for its specific strengths in handling the complexities of EEG data. Support Vector Machines (SVM) and Gaussian Naive Bayes were chosen for their effectiveness in high-dimensional spaces and probabilistic approach, respectively. Gradient Boosting, XGBoost, and LightGBM were included for their robustness to overfitting and efficiency in processing large datasets. These models collectively address the challenges of EEG data analysis, such as noise, high dimensionality, and class imbalance. Training and evaluation were challenging due to the intricate nature of EEG signals, with specific strategies employed to mitigate issues like overfitting and ensure model accuracy and reliability. Figuring out when seizures may take place was the objective of this study. Therefore, the whole concept hinged on the utilisation of these ingenious models in order to properly forecast epileptic episodes.

A. Model Network

Support Vector Machines (SVM), Gaussian Naive Bayes (GNB), Gradient Boosting (GB), XGBoost, and LightGBM are some of the machine learning models that are part of the application. These models are carefully chosen for their ability to handle different aspects of the prediction.

B. Algorithmic Steps

The proposed algorithm comprises the following key steps:

1) Data Loading and Preprocessing:
   - Load the EEG dataset and Preprocess target variable for consistency.
   - Drop rows with missing values to maintain data integrity.
   - Apply the “Local Outlier Factor (LOF)” algorithm for outlier identification and exclusion.
2) Feature Scaling and Selection:
   - Scale features using “StandardScaler” for normalisation.
   - Use Recursive Feature Elimination (RFE) with “RandomForestClassifier” to select informative features.
3) Dataset Splitting:
   - Extract the dataset into training and test sets using “train_test_split” function.
4) Model Training and Evaluation:
   - Train various machine learning models on the training set, including SVM, Gaussian Naive Bayes, Gradient Boosting, XGBoost, and LightGBM.
   - Evaluate trained models using performance metrics (accuracy, sensitivity, specificity, precision, recall, and F1-score).
5) Hybrid Model Creation:
   - Average predictions from individual models to create a hybrid model.
6) Performance Analysis and Visualization:
   - Compare the performance of each model.
   - Plot ROC curves for model performance visualization.
   - Generate accuracy curves to illustrate accuracy improvement with increasing sample size.

Fig. 5 shows our proposed methodologies flowchart. In the end, by demonstrating the efficacy of ML models and FS methods on the Bonn dataset, this study advances the area of epilepsy prediction. The findings highlight the potential of these approaches in developing reliable and accurate prediction systems for epilepsy management.

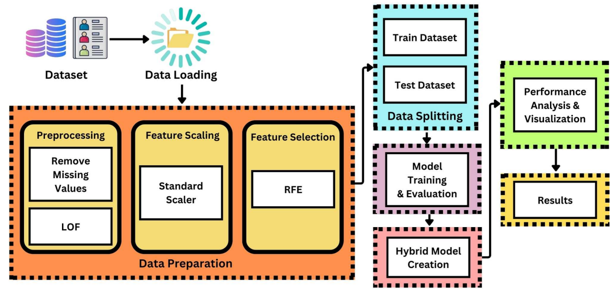

Fig. 5. Overall architecture of proposed methodology.

C. Setup

The suggested system requires importing numpy, pandas, scikit-learn, matplotlib, xgboost, and lightgbm. The target variable is preprocessed after loading the CSV dataset. Detect and eliminate outliers, then scale features using “StandardScaler”. Features are chosen using RFE. After separating the dataset into training and test sets using the ‘train_test_split’ function, we trained our models using the set. Next, we assessed accuracy, sensitivity, specificity, precision, recall, and F1-score. Then performance comparison has done to show the practicality of the proposed hybrid model.

D. EpiNet

Within the suggested approach, a simple averaging technique is used to build the hybrid model, named EpiNet. After training multiple machine learning models, including SVM, Gaussian Naive Bayes, Gradient Boosting, XGBoost, and LightGBM, the average predictions are pooled. This cooperative strategy uses each model’s advantages to ensure fair and efficient decision-making. For each instance, the hybrid model’s output is the average forecast from the group’s insights, improving epileptic seizure prediction accuracy and reliability. The hybrid model is built using simple averaging in the provided way. After training multiple machine learning models, including SVM, Gaussian Naive Bayes, Gradient Boosting, XGBoost, and LightGBM, the average predictions are pooled. This cooperative strategy uses each model’s advantages to ensure fair and efficient decision-making. The Hybrid model’s final output for a given instance is the average forecast from the group’s insights, improving epileptic seizure prediction accuracy and reliability.

VI. Results and Discussion

Our extensive epileptic episode prediction study comprised machine learning models and innovative feature selection methods. This method validated each model and explained seizure prediction. This discussion will explain the findings and any surprises. Linking our findings to earlier research aids comprehension. Comprehensive analysis is needed to improve epileptic seizure prediction.

A. Model Performance Dissection

To evaluate models’ prediction ability, the suggested technique uses major performance metrics. Model accuracy, sensitivity, specificity, precision, recall, and F1-score show performance. The accuracy metric the ratio of accurate predictions to total forecasts indicates correctness. Recall, or sensitivity, measures how effectively models detect positive cases. Specificity, which assesses negative projections, boosts sensitivity. Recall retrieves all positive events, whereas precision assesses positive prediction accuracy. The balanced F1-score combines recollection and accuracy for a detailed evaluation. The mathematical formulations of these metrics' equations give a solid foundation for assessing model performance in varied circumstances. The formulaś used to calculate the evaluation metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{5}$$

TABLE II. PERFORMANCE ANALYSIS OF THE PROPOSED EPINET MODEL

Model	Sensitivity	Specificity	Precision	Recall	F1-Score	Support	ROC AUC
SVM	0.909	1.000	0.998	0.998	0.998	22	0.995
GaussianNB	0.954	0.983	0.989	0.998	0.984	22	0.993
Gradient Boosting Classifier	0.909	0.999	0.997	0.997	0.997	22	0.974
XGB Classifier	0.909	1.000	0.998	0.998	0.998	22	0.996
LGBM Classifier	0.909	1.000	0.998	0.998	0.998	22	0.985
Hybrid	**0.909**	**1.000**	**0.998**	**0.998**	**0.998**	**22**	**0.971**

Our method is evaluated using SVM, Gaussian Naive Bayes, Gradient Boosting, XGBoost, LightGBM, and the innovative hybrid model in Table II. Each epileptic seizure prediction method has merits and downsides. SVM and XGBoost excel in accuracy and sensitivity. They recognise non-seizure circumstances well. For seizure prediction, Gaussian Naive Bayes and Gradient Boosting are more sensitive. Seizure treatment relies on their sensitivity to detect true positives. We now notice the hybrid model (EpiNet) for another reason. It balances several variables well. Its strength is combining the best features of various models without compromising others. It predicts epileptic seizures well due to its comprehensive approach. This hybrid model's success raises questions regarding its components' interactions. Exploring each model's strengths reveals their goal: accurate seizure prediction. Details on the hybrid model reveal its effectiveness and probable linkages, raising interest in its role in epilepsy forecasting.

B. *Insights into Hybrid Model Superiority*

Table III shows the mixed model ratings where the evaluation metrices such as accuracy, precision, recall, and F1-score of Class 0 (Non-Epilepsy) and 1 (Epilepsy) are measured. Also macro and weighted averages are provided. The hybrid model predicts epileptic episodes with 99.81% accuracy. The Hybrid Model adeptly amalgamates predictions from diverse models, showcasing superior performance in specificity and precision compared to individual models. Simultaneously, it achieves heightened sensitivity, signifying its efficacy in achieving an optimal balance. This amalgamation strategically leverages the strengths of individual models, providing a robust and superior framework for epileptic seizure prediction. Importantly, our findings demonstrate that the Hybrid Model outperforms individual models in key metrics, aligning with emerging research [3] advocating for hybridization to substantially enhance predictive accuracy. Our findings echo and extend the conclusions drawn by prior studies [21] that emphasized the significance of feature selection techniques in enhancing predictive accuracy. The nuanced performance variations observed in SVM models align with previous assertions regarding the impact of dataset characteristics on SVM effectiveness [22]. Additionally, the superior performance of the hybrid model resonates with recent research advocating for ensemble methods [10] in achieving superior predictive accuracy.

TABLE III. EVALUATION METRICS FOR THE EPINET MODEL

Metric	Precision	Recall	F1-Score	Accuracy
Class 0 (Non-Epilepsy)	1.00	1.00	1.00	**99.81%**
Class 1 (Epilepsy)	1.00	0.91	0.95	
Macro Avg	1.00	0.95	0.98	**99.81%**
Weighted Avg	1.00	1.00	1.00	

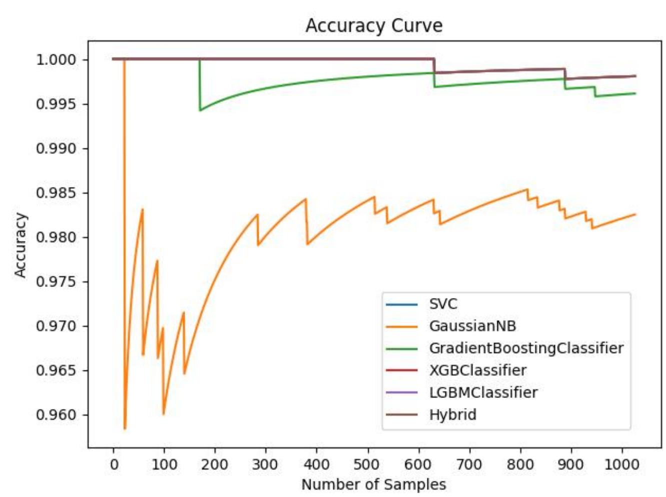

Fig. 6. Accuracy curve.

As shown in Fig. 6, the relationship among the amount of specimens and the precision of the predictions is represented by the accuracy curve. It illustrates how the precision of the models increases with the inclusion of additional samples. Sensitivity, denoted by the ROC AUC curve in (see Fig. 7) is the compromise between the FPR and the TPR. The graphical depiction illustrates the capacity of the models to distinguish positive from negative classifications. An increased AUC (Area Under the Curve) signifies superior data classifi-

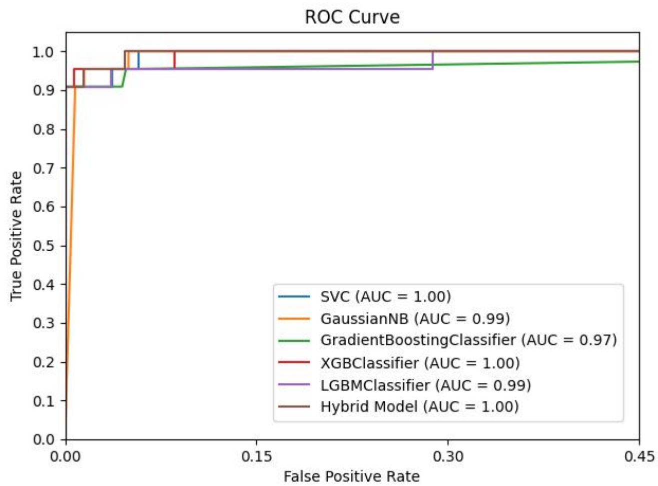

Fig. 7. ROC AUC curve.

cation performance. The ROC curve analysis shows not only the exceptional classification performance of individual models and EpiNet but also indicates their specificity and sensitivity balance. The perfect AUC score of 1.00 for EpiNet suggests no overlap between the true positive rate and false positive rate, indicating an ideal separation of classes. The hybrid model's confusion matrix (see Fig. 8) shows that the vertical elements show the true positive and true negative forecasts, which show cases that were correctly labelled as Non-Epilepsy and Epilepsy, respectively. Specifically, the model's perfect success rate of 100% for non-epilepsy instances demonstrated its exceptional ability to identify such situations. For the Epilepsy class, the model's sensitivity was 90.91%, indicating that it accurately detected 90.91% of actual instances. In contrast, the model's ability to accurately identify genuine negative instances is shown by the 100% accuracy for the Non-Epilepsy class. On the accuracy curve, EpiNet maintains a plateau of high accuracy across sample sizes, demonstrating robustness against overfitting—a challenge often encountered with smaller datasets. The nuanced fluctuations observed in other models' accuracy curves could be attributed to their individual handling of the dataset's complexity, which is mitigated in EpiNet's ensemble strategy. These insights into the model performance dynamics affirm the superiority of the hybrid approach, where collective intelligence effectively captures the intricate patterns in EEG data critical for seizure prognosis.

The hybrid model distinguishes epilepsy from non-epilepsy. This prepares for epilepsy prediction and classification. Our method diagnoses epilepsy patients with 99.81% accuracy. The findings demonstrate that our epilepsy prediction method is accurate and reliable. This confirms our model's applicability.

From Table IV, we can see that the EpiNet Model, a combination of different techniques for epileptic seizure prediction, outperforms individual Support Vector Machines (SVM) models, achieving an accuracy of 99.81%. Given the scarcity of studies on hybrid models, our approach, incorporating SVM alongside other methods, demonstrates enhanced prediction accuracy. Overall, our system, leveraging various machine learning models, presents promising results, with the Hybrid

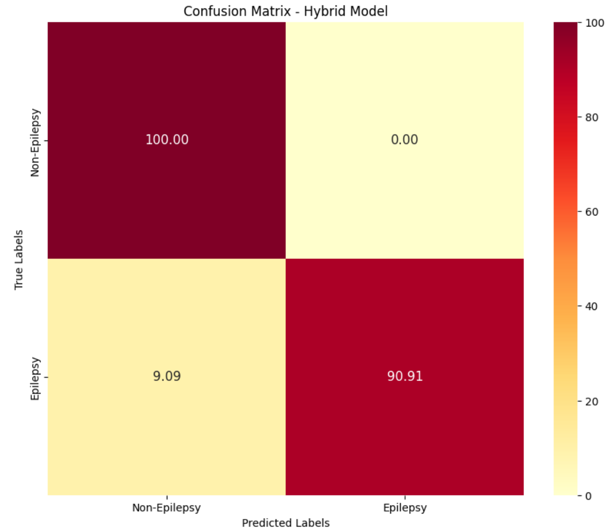

Fig. 8. Confusion matrix.

TABLE IV. PERFORMANCE ANALYSIS OF THE PROPOSED EPINET MODEL WITH STATE-OF-ARTS BASED ON SVM MODEL

Publication	Models	Accuracy
[26]	SVM	94%
[21]	"SVM" with "LBP"	97.80%
[10]	SVM	92.23%
[30]	SVM	97%
[23]	SVM	95.33%
[22]	SVM	94%
Proposed Method	**Hybrid Model**	**99.81%**

Model showing superior performance and accuracy compared to individual models.

C. *Identification of Model-Specific Strengths*

Delving deeper, the unexpected prominence of specific models in certain metrics prompts a nuanced understanding. SVM emerges as a stalwart in specificity, a characteristic well-documented in studies focusing on Support Vector Machines for epilepsy detection [26]. XGBoost, with its gradient boosting prowess, excels in overall accuracy, an attribute substantiated in recent research [38]. Identifying these model-specific strengths contributes valuable insights for tailored model selection based on the diagnostic emphasis.

VII. CONCLUSION AND FUTURE WORK

In the pursuit of advancing automatic epilepsy detection from EEG signal data, this study introduces a novel hybrid machine learning model, exhibiting a remarkable accuracy of 99.81% in the classification of five distinct classes (A-B-C-D-E) based on a dataset from the University of Bonn. In this study, we carefully prepared EEG data, addressed outliers, scaled features, and trained our model. What sets our approach apart is the use of a straightforward averaging method to combine model predictions, yielding outstanding classification results. This work establishes a strong foundation for advancing accurate epileptic seizure prediction techniques. By contributing to the landscape of medical diagnostics, this research holds the potential to significantly enhance healthcare strategies tailored to epilepsy patients.

While this study demonstrates promising results in automatic epilepsy detection, it is essential to acknowledge the limitations stemming from dataset specificity and potential challenges in generalization. Though our model excelled on the Bonn dataset, its generalizability to diverse datasets requires careful consideration. It signifies a need for further exploration and adaptation to diverse data sources. Adapting and fine-tuning the model for consistent performance across different data sources is necessary. Our findings in epileptic seizure prediction using EpiNet open new avenues in proactive healthcare management. By achieving a high accuracy rate, this model can be integrated into real-time monitoring systems, offering early warning signals for impending seizures. This advancement holds the potential to drastically improve patient outcomes, enabling timely interventions and personalized treatment plans. Future work will focus on enhancing the generalizability of EpiNet across diverse datasets, ensuring its applicability in various clinical settings. By doing so, we aim to contribute significantly to the evolution of healthcare strategies for epilepsy patients, making a tangible difference in their quality of life and care. Future research endeavors should focus on addressing these constraints by exploring diverse datasets, refining model robustness, and incorporating real-time monitoring systems. Overcoming these challenges will contribute to the continuous evolution of accurate and reliable epileptic seizure prediction techniques, further advancing the field of medical diagnostics and improving healthcare for individuals with epilepsy.

Acknowledgment

The researchers wish to extend their sincere appreciation to the University of Asia Pacific (UAP), Bangladesh and the Institute of Energy, Environment, Research, and Development (IEERD) for their invaluable support of this study.

References

[1] https://www.who.int/news-room/fact-sheets/detail/epilepsy (Accessed on: 17 January 2024).

[2] O. K. Esha, N. Haque, F. Ahmed and N. Begum, “A Comprehensive Review on Epilepsy Diagnosis and Prognosis using Machine Learning and Deep Learning Approaches,” 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Naya Raipur, India, 2022, pp. 159-164, doi: 10.1109/WIECON-ECE57977.2022.10151006.

[3] Kapoor, Bhaskar, et al. “Epileptic Seizure Prediction Based on Hybrid Seek Optimization Tuned Ensemble Classifier Using EEG Signals.” Sensors, vol. 23, no. 1, Dec. 2022, p. 423. Crossref, https://doi.org/10.3390/s23010423.

[4] Savadkoohi, Marzieh, Timothy Oladunni, and Lara Thompson. “A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) Signal.” Biocybernetics and Biomedical Engineering 40.3 (2020): 1328-1341.

[5] Slimen, Itaf Ben, Larbi Boubchir, and Hassene Seddik. “Epileptic seizure prediction based on EEG spikes detection of ictal-preictal states.” Journal of biomedical research 34.3 (2020): 162.

[6] Usman, Syed Muhammad, Shehzad Khalid, and Zafar Bashir. “Epileptic seizure prediction using scalp electroencephalogram signals.” Biocybernetics and Biomedical Engineering 41.1 (2021): 211-220.

[7] Usman, Syed Muhammad, Shehzad Khalid, and Muhammad Haseeb Aslam. “Epileptic seizures prediction using deep learning techniques.” Ieee Access 8 (2020): 39998-40007.

[8] Alshebeili, Saleh A., et al. “Inspection of EEG signals for efficient seizure prediction.” Applied Acoustics 166 (2020): 107327

[9] Wang, Gang, et al. “Seizure prediction using directed transfer function and convolution neural network on intracranial EEG.” IEEE Transactions on Neural Systems and Rehabilitation Engineering 28.12 (2020): 2711-2720.

[10] Ryu S, Joe I. A Hybrid DenseNet-LSTM Model for Epileptic Seizure Prediction. Applied Sciences. 2021; 11(16):7661. https://doi.org/10.3390/app11167661https://doi.org/10.3390/app11167661.

[11] Viana, Pedro F., et al. “Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized intrapatient models.” Epilepsia (2022).

[12] Behnoush, B., et al. “Machine learning algorithms to predict seizure due to acute tramadol poisoning.” Human & Experimental Toxicology 40.8 (2021): 1225-1233.

[13] Syed Muhammad Usman, Muhammad Usman, Simon Fong, “Epileptic Seizures Prediction Using Machine Learning Methods”, Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 9074759, 10 pages, 2017.https://doi.org/10.1155/2017/9074759.

[14] Echtioui A, Hamam H,et al. Deep Learning Models for Predicting Epileptic Seizures Using iEEG Signals. Electronics. 2022; 11(4):605. https://doi.org/10.3390/electronics11040605

[15] Harpale, Varsha, and Vinayak Bairagi. “An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states.” Journal of King Saud University-Computer and Information Sciences 33.6 (2021): 668-676.

[16] Tran, Ly V., et al. “Application of machine learning in epileptic seizure detection.” Diagnostics 12.11 (2022): 2879

[17] Cao, Xincheng, et al. “Automatic seizure classification based on domain-invariant deep representation of EEG.” Frontiers in Neuroscience 15 (2021): 760987.

[18] Tripathi, Prashant Mani, et al. “Automatic Seizure Detection and Classification Using Super-resolution Superlet Transform and Deep Neural Network-A Preprocessing-less Method.” Computer Methods and Programs in Biomedicine (2023): 107680.

[19] Ahmed, Mohammed Imran Basheer et al. “A Review on Machine Learning Approaches in Identification of Pediatric Epilepsy.” SN computer science vol. 3,6 (2022): 437. doi:10.1007/s42979-022-01358-9

[20] Farooq, Muhammad Shoaib, Aimen Zulfiqar, and Shamyla Riaz. “Epileptic Seizure Detection Using Machine Learning: Taxonomy, Opportunities, and Challenges.” Diagnostics 13.6 (2023): 1058.

[21] Tharayil JJ, Chiang S, Moss R, Stern JM, Theodore WH, Goldenholz DM. A big data approach to the development of mixed-effects models for seizure count data. Epilepsia. 2017;58(5):835-844. doi:10.1111/epi.13727

[22] Jaiswal, A.K., & Banka, H. (2017). Epileptic seizure detection in EEG signal with GModPCA and support vector machine. Bio-medical materials and engineering, 28 2, 141-157 .

[23] Tiwari, A.K., Pachori, R.B., Kanhangad, V., & Panigrahi, B.K. (2017). Automated Diagnosis of Epilepsy Using Key-Point-Based Local Binary Pattern of EEG Signals. IEEE Journal of Biomedical and Health Informatics, 21, 888-896.

[24] Toraman, Suat. “Preictal and Interictal Recognition for Epileptic Seizure Prediction Using Pre-trained 2DCNN Models.” Traitement du Signal 37.6 (2020).

[25] Satapathy, Sandeep & Jagadev, Alok & Dehuri, Satchidananda. (2017). Weighted Majority Voting Based Ensemble of Classifiers Using Different Machine Learning Techniques for Classification of EEG Signal to Detect Epileptic Seizure. Informatica. 41. 99-110.

[26] Lima, C.A., Coelho, A.L., Madeo, R.C., & Peres, S.M. (2015). Classification of electromyography signals using relevance vector machines and fractal dimension. Neural Computing and Applications, 27, 791 - 804.

[27] Kabir, E., Siuly, & Zhang, Y. (2016). Epileptic seizure detection from EEG signals using logistic model trees. Brain informatics, 3(2), 93–100. https://doi.org/10.1007/s40708-015-0030-2

[28] Aslam, M.H.; Usman, S.M.; Khalid, S.; Anwar, A.; Alroobaea, R.; Hussain, S.; Almotiri, J.; Ullah, S.S.; Yasin, A. Classification of EEG Signals for Prediction of Epileptic Seizures. Appl. Sci. 2022, 12, 7251. https://doi.org/10.3390/ app12147251

[29] Patrick H, Luckett BS. “Nonlinear methods for detection and prediction of epileptic seizures.” A Dissertation submitted in University of South Alabama, July, 2018.

[30] Abualsaud, Khalid et al. "Ensemble classifier for epileptic seizure detection for imperfect EEG data." TheScientificWorldJournal vol. 2015 (2015): 945689. doi:10.1155/2015/945689

[31] Sharma, Rahul. "Localization of epileptic surgical area using automated hybrid approach based on higher-order statistics with sensitivity analysis and residual wavelet transform." Biomedical Signal Processing and Control 86 (2023): 105192.

[32] Brari, Zayneb, and Safya Belghith. 'A Novel Machine Learning Approach for Epilepsy Diagnosis Using EEG Signals Based on Correlation Dimension'. IFAC-PapersOnLine, vol. 54, no. 17, 2021, pp. 7–11, https://doi.org10.1016/j.ifacol.2021.11.018.

[33] Slimen IB, Boubchir L, Seddik H. Epileptic seizure prediction based on EEG spikes detection of ictal-preictal states. J Biomed Res. 2020 Feb 17;34(3):162-169. doi: 10.7555/JBR.34.20190097. PMID: 32561696; PMCID: PMC7324272.

[34] Ilakiyaselvan, N et al. "Deep learning approach to detect seizure using reconstructed phase space images." Journal of biomedical research vol. 34,3 (2020): 240-250. doi:10.7555/JBR.34.20190043

[35] Singh K, Malhotra J. Two-layer LSTM network-based prediction of epileptic seizures using EEG spectral features. Complex Intell Syst. 2022;8(3):2405-2418. doi:10.1007/s40747-021-00627-z

[36] Li, Chang, et al. 'EEG-Based Seizure Prediction via Transformer Guided CNN'. Measurement, vol. 203, 2022, p. 111948, https://doi.org10.1016/j.measurement.2022.111948.

[37] Rasheed, Khansa et al. "A Generative Model to Synthesize EEG Data for Epileptic Seizure Prediction." IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society vol. 29 (2021): 2322-2332. doi:10.1109/TNSRE.2021.3125023

[38] Kumar, T., et al. 'A Modified XG Boost Classifier Model for Detection of Seizures and Non-Seizures'. WSEAS TRANSACTIONS ON BIOLOGY AND BIOMEDICINE, vol. 19, 01 2022, pp. 14–21, https://doi.org10.37394/23208.2022.19.3.

[39] Usman, Syed Muhammad, Shehzad Khalid, and Sadaf Bashir. "A deep learning based ensemble learning method for epileptic seizure prediction." Computers in Biology and Medicine 136 (2021): 104710.

[40] "Bonn Dataset" https://www.ukbonn.de/epileptologie/arbeitsgruppen/ag-lehnertz-neurophysik/downloads/