MAJOR PROJECT [ETMJ100]

on

Unveiling Epilepsy: Machine Learning and Deep Learning Approaches for EEG Signal-Based Patient Detection

Submitted to

Amity University Uttar Pradesh
In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

By:
Mantra Jain (A2305220412)

Ansh Srivastav (A2305220390)

(2020-24)

Under the guidance of

Dr. Harshit Bhardwaj

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY
AMITY UNIVERSITY, NOIDA, UTTAR PRADESH (201301)
April-May 2024

DECLARATION

We, **Mantra Jain** and **Ansh Srivastav**, students of Bachelor of Technology in Computer Science and Engineering hereby declare that the project titled **"Unveiling Epilepsy: Machine Learning and Deep Learning Approaches for EEG Signal-Based Patient Detection"** which is submitted by us to Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, in partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

Noida

Date:

Mantra Jain [A2305220412] Ansh Srivastav [A2305220390]

Batch (2020-24)

CERTIFICATE

On the basis of declaration submitted by **Mantra Jain** and **Ansh Srivastav**, students of B. Tech (Computer Science and Engineering), I hereby certify that the project titled **"Unveiling Epilepsy: Machine Learning and Deep Learning Approaches for EEG Signal-Based Patient Detection"** which is submitted to Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering, is an original contribution with existing knowledge and faithful record of work carried out by them under my guidance and supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Noida

Date:

Dr. Harshit Bhardwaj

Assistant Professor

Department of Computer Science and Engineering

Amity School of Engineering and Technology

Amity University Uttar Pradesh

ACKNOWLEDEGEMENT

I would like to express my sincerest gratitude and indebtedness to **Prof. (Dr.) Sanjeev Thakur (Head of Computer Science and Engineering)**, It is high privilege for me to express my deep sense of gratitude to those entire faculty Members who helped me in the completion of the project, specially my internal guide **Dr. Harshit Bhardwaj** who was always there at hour of need and all the professors of our Computer Science Department.

My special thanks to all other faculty members, batchmates & seniors of Amity School of Engineering and Technology, Amity University Uttar Pradesh for helping me in the completion of project work and its report submission.

Sincerely,

Mantra Jain

A2305220412

Ansh Srivastav

A2305220390

Table of Contents

ABSTRACT

The intricate neurological condition known as epilepsy, which is common across the world, presents considerable difficulties in accurately identifying and differentiating between non-epileptic and epileptic activity using electroencephalograms (EEGs). To customize successful therapies, it is essential to accurately identify the kinds of epileptic activity. Since epilepsy affects about 50 million people worldwide according to latest update of WHO and is typified by spontaneous seizures, early identification and prediction are vital in enabling people to minimize possible harm.

This report provides a brief overview of the report on epilepsy diagnosis and classification analysis, which includes various machine learning algorithms such as K-Nearest Neighbour (KNN), Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine (SVM) and Decision Trees. This report explores the evolving field of epilepsy diagnosis and reviews the various machine learning algorithms, datasets, and computational techniques currently in use.

To identify small patterns in EEG data, this study combines cutting edge technologies, like Long Short-Term Memory (LSTM) and 1D-CNN (Convolutional Neural Network) leveraging data from five hundred patients acquired from the UCI Machine Learning Repository. To optimize the 1D-CNN LSTM architecture and hyper-parameters, Bayesian optimization is employed, allowing for efficient exploration of the parameter space. Its effectiveness is not only limited to enhancing the performance metrics of a particular model but also minimizing the computing power required for fine tuning. The research evaluates the effectiveness of the 1D-CNN LSTM-based model, showcasing its potential as a reliable tool for automated epilepsy detection with accuracy of 99.47% (≈ 100%), average sensitivity of 99.45%, and average specificity of 99.57%. This approach, emphasizes the significance of anticipating seizures in advance, attempts to provide epileptics the tools they need to control and avoid seizures in advance, so ultimately enhancing their quality of life for patients.

Keywords: Epilepsy, Seizures, 1D-CNN, LSTM, Bayesian Optimization, electroencephalogram

List of Tables

List of Figures

I. INTRODUCTION

We discovered at the outset of this research that it would be quite helpful to clarify a few things. To provide them a brief overview of what they will read about in the next chapters, as well as the nature of the examination's subject and the solution's structure. We will concentrate on identifying epileptic seizures in electroencephalogram (EEG) data. All users are welcome to utilize this information, which was gathered at the German university of Bonn. Several well-known machine learning techniques that have been suggested in the literature for comparable tasks will be used in the identification procedure. To evaluate them, seven different measures will be applied. Python 3.7 was used to implement the whole procedure. The objective is to contrast some of the approaches put out in the literature and expand them from patient-specific to datasets with numerous cases.

The current seizure prediction methods lack, with particular emphasis on their limited performance on small training datasets and their disregard for time-series data. It is a mental disorder characterized by seizures and uncertainty, remains a significant medical problem. Timely and accurate detection of epilepsy is very important for diagnosis, treatment, and patient management. Considering that seizures can occur suddenly and without warning, it is important to have a system that can detect seizures. A comprehensive review of the electroencephalogram (EEG) recording is required to accurately identify these seizures. In recent years, the intersection of machine learning and medicine has shown promise in improving the diagnosis and classification of epilepsy.

Epilepsy is a mental disorder characterized by sudden and unpredictable events that affects millions of people worldwide. These seizures are caused by electrical malfunctions in the brain and often present with symptoms that vary in intensity and duration.[1] Epilepsy is a chronic brain disorder affecting nerves cell activity for an individual of all ages. It has an impact over 50 millions of worldwide population, positioning it as one of the most prevalent neurological conditions. Almost 80% of epileptics belongs to blue collar class, if given the right diagnosis and care in early stages, there are chances of making up to 70% of epileptics enjoy a seizure free life. Individuals with epilepsy have a threefold increased risk of dying young compared to a healthy human being. In developing or underdeveloped nations, 75% of epileptic patients do not undergo proper treatment and many may die undiagnosed. People suffering from epilepsy along with their families and relatives must face stigma and prejudice in many parts of the world.

As per WHO [2] in India, the average incidence of epilepsy is 5.59–10 per 1,000 individuals. In India, there are more than ten million epileptic sufferers, or more than 1% of the total population. The incidence is higher in rural areas 1.9% in contrast to urban areas 0.6%. Since 2015, February's second Monday has been marked as International Epilepsy Day (IED), an internationally recognized healthcare event aimed at uniting epileptic patients and fostering a community where knowledge of the condition's epidemiological profile, diagnosis, and treatment options is exchanged. [3] Electroencephalogram (EEG) is one of the most common diagnostic measures used in medical industry to diagnose epilepsy which is a highly intricate disease. This condition is so complex that it makes understanding EEG signals or results very difficult. Integration of such techniques with machine learning is important in differentiating epileptic seizures from other types and identifying particular forms of epileptic activity.

The optimal treatment and management of epilepsy requires its diagnosis to be accurate and within the golden time period i.e. before we can see the external symptoms of epilepsy like staring, jerking movements of the arms and legs or stiffening of the body. Medical applications for machine learning have been growing rapidly, providing a wide range of opportunities for the analysis, diagnosis and classification of epilepsy. With this integration comes a new way of enhancing diagnostic accuracy, predicting epilepsy occurrences as well as coming up with personalized and customized treatment plans. This comprehensive report encompasses various topics on machine learning for classification of epileptic and non-epileptic signals, stressing on the significance of artificial intelligence (AI) in unravelling complex medical conditions. However, traditional epilepsy diagnosis relied entirely on neurologist clinical observations, physical examinations, and electroencephalography (EEG) data analyses that are all human dependent and are prone to be mistaken thus chances of detection of epilepsy within golden hour is very much reduced, but incorporating Machine Learning techniques such as K-Nearest Neighbour (KNN) or logistic regression is expected to offer a much more precise and efficient way to diagnose the disease.

Both the search and classification scenario can be modified by various machine learning and deep learning algorithms. These algorithms help us distinguish seizures from other conditions, predict the frequency of seizures, and use data from big, complex datasets from pattern recognition and data analytics to develop customized treatment plans for everyone.

We will examine the state of the art models in epilepsy detection and classification in this report, focusing on the many types of system mastery algorithms employed as well as the statistical and computational techniques. We want to get a better understanding of these algorithms efficacy in identifying epilepsy and forecasting seizures by analysing their advantages and disadvantages.

We want to see the potential of machine learning and deep learning, including algorithms like logistic regression and CNN, in classification as we further explore the merger of technology and health. Readers will have a better grasp of the field's present status, upcoming difficulties, and usefulness for machine and deep learning to enhance patient care by having an epilepsy management method from the information in this report.

In this light, artificial intelligence intersects [4] with medical research as promising pathways towards advancing the comprehension and handling of epilepsy. With regards to machine learning, recurrent neural networks particularly those involving Long- Short Term Memory (LSTM) [5] networks have shown potential in decoding complex patterns within time series data. This improves accuracy and efficiency in detection and prediction of epileptic seizure A significant obstacle in the earlier research on seizure prediction is the insufficient analysis of time-series data. One kind of neural network that retains information from earlier instances is the Recurrent Neural Network (RNN), which uses past outputs as inputs [6]. RNNs have been more popular recently in studies on speech recognition and natural language processing. Normally RNN faces the gradient vanishing problem which is not an issue with LSTM, one of the RNN designs, which makes it easier to learn long-term relationships in time series data [7].

T. Sainath et al. [8] improved the performance metrics of the DNNs by making an ensemble model of RNN and CNN into a convolutional neural network. In some large problems, this led to a 4 to 6% relative improvement over independent implementation of LSTMs. Numerous studies that have looked at the combination of CNN and LSTM to extract temporal and spatial properties have shown how successful this approach could be by giving prominent outputs in classification. [9]

This report highlights the need of using 1D-CNN LSTM ensembled (our suggested final model based on deep learning algorithms) networks in order to understand the temporal dynamics found in EEG data. Because these networks are designed to detect long term correlations in sequential data, they are perfect for exposing minute patterns that are suggestive of impending epileptic activity. Also Bayesian optimization is used to optimize the performance of the suggested ensemble model. This is a useful technique for adjusting hyper-parameters. The model is ensured to attain optimal configurations that optimize projected accuracy while utilizing the least amount of processing resources that is achieved by employing Bayesian optimization. Interestingly, 500 patients EEG recordings were made available by the UCI Machine Learning Repository, each file contained 4097 data points over a 23.5-seconds period. [10]

II. Related Work

The extension of machine learning in epilepsy-focused sectors, including seizure detection and monitoring, has been the subject of numerous studies. By utilizing methods such as multilayer artificial neural networks, support vector machine (SVM), and deep learning, machine learning shows potential in enhancing the ability to handle and evaluate EEG and imaging data that was once considered too complex for experts. Furthermore, in this paper Abbasi, Bardia and Goldenholz, Daniel M [11] supports applying machine learning techniques to optimize medication selection, improve the precision of clinical outcome predictions, and streamline surgical planning. Predictive models produced by machine learning are a source of concern for the authors due to the limited number of validation studies published. It's worth considering the applicability and generalizability of these models in light of this deficiency. Broader datasets that take into account greater diversity are recommended by the authors in order to fill this void. Furthermore, the expected increase in investment in external validation studies to make the application of machine learning in medicine, particularly in epilepsy, more reliable was highlighted.

Amin, Ushtar, and Benbadis, Selim R [12], highlight the complexity involved in reversing an epilepsy diagnosis, emphasizing the necessity of examining "unusual" EEG patterns, which can pose challenges. A major factor contributing to misinterpretation of regular EEGs as abnormal is the lack of practical experience in neurology residency programs. They argue against prioritizing tests like EEG over medical expertise, as certain seizure types may evade detection, complicating epilepsy identification. For instance, hypermotor seizures in the frontal lobe might be mistaken for psychogenic episodes, while focal unaware cognitive seizures in older adults could be misdiagnosed as dementia. Additionally, epilepsies affecting the frontal and temporal lobes may manifest as psychotic symptoms, leading to misdiagnosis as primary mental disorders. Diagnostic errors are common across medical specialties, carrying significant consequences for both patients and physicians. In neurology, errors often stem from an overemphasis on assessments rather than considering the clinical context. Epilepsy diagnosis typically relies on clinical evaluation and medical history, with overdiagnosis being more prevalent than underdiagnosis. Lack of adequate medical background and atypical EEG findings can contribute to erroneous epilepsy diagnoses. Patients previously diagnosed with epilepsy may fail to improve with antiepileptic medications if they do not truly have the condition. In reality, many individuals receiving incorrect epilepsy diagnoses ultimately experience syncope or psychogenic nonepileptic events.

Chen, Hai and Koubeissi, Mohamad Z reviewed how electroencephalogram (EEG) is linked to Epileptic seizures and provided physiologic basis of EEG and intracranial EEG studies. They talked about pointed contoured waveforms or complexes that are different from background waves and mimic those observed in a part of human people with epileptic diseases are referred to as interictal epileptiform discharges. The most extensively studied interictal epileptiform discharges consist of spikes and sharp waves [13]. They elaborated on rhythmic discharges, which usually need to persist for a minimum of 10 seconds to be classified as an electrographic seizure. BIRDs (Brief Potentially Ictal Rhythmic Discharges) are described as “Concisely, this refers to short bursts of rhythmic brain activity exceeding 4 Hz, which may appear abruptly and do not match any recognized normal or harmless patterns" Their research frequently identifies interictal or ictal abnormalities, and how EEG is still an essential tool for diagnosis of epilepsy. However, the absence of interictal epileptiform discharges or ictal symptoms does not necessarily exclude epilepsy. Seizures can manifest in two forms: focal or generalized. Electrographic patterns may vary, and ictal activity typically evolves over the course of a seizure. For accurate diagnosis and treatment of nonconvulsive status epilepticus (NCSE)—a condition characterized by continuous seizure activity lasting at least 30 minutes, accompanied by cognitive or behavioural alterations—continuous EEG monitoring plays a vital role. When scalp EEG findings are inconclusive, intracranial EEG monitoring proves invaluable, especially in surgical planning, as it often enables earlier detection of seizures and offers superior spatial resolution compared to scalp recordings.

Lahmiri, Salim indicated how epilepsy is becoming more common, and its prevalence is rising. Designing precise computerized procedures for the identification and categorization of electroencephalogram (EEG) data from epileptic patients is therefore very helpful in the diagnostic process. There work aims to propose a machine-learning diagnosis method that can quickly and accurately identify between normal and abnormal EEG data with seizure-free periods using the extended Hurst exponent approaches, fractal features of EEG signals are computed at various scales to better describe their dynamics[14]. Generic Hurst exponent estimations between healthy and epileptic EEG signals with seizures uninterrupted durations are statistically different, according to parametric and nonparametric statistical tests. Support vector machine classifiers that were trained using extended Hurst exponent estimates. There suggested system has potential and can be expanded for other biomedical applications such as differentiating between normal brain waves and those with intervals of seizure or between epileptic EEG signals with seizure free intervals because this problem is challenging and has not been addressed in the literature.

Mesraoua et al. indicated how EEG in comparison to the conventional method of eye assessment alone, scalp electroencephalography has the potential to provide additional spatial and temporal information. Fortunately, this information is easier to acquire because to contemporary digital EEG technology and computer-assisted analysis. A potential method to enhance non-invasive EEG localisation in focal epilepsies is to look at the spike voltage topography of interictal spikes [15]. Another additional method for locating the epileptogenic zone in individuals who are candidates for epilepsy surgery is electrical source imaging. Quantitative EEG offers a simplified and a static visualization of the extensive amount of data contained in continuous EEG. In recent times scalp EEG analysis has improved significantly with the use of computer assisted techniques and technological advancements. Scalp EEG recordings have been enhanced by including spike voltage topography, electrical source imaging and quantitative EEG to offer more consistent spatial and temporal information especially in epilepsy. Modern digital EEG equipment and sophisticated computer algorithms have provided neurologists with additional information to aid in the accurate diagnosis and therapy of epilepsy. This study to maximize the use of scalp EEG in epilepsy demonstrates the necessity of encouraging technological advancements identification and treatment.

Research was carried out by Laura Abraira et al. [16] as a part of group divided into three parts first, being the patient affected by the loss of consciousness, secondly, there were 41 patients who had experienced transient ischemic attack and at last, there were a bunch of 26 healthy people. The gender distribution for the LOE group was such that 57.6% of the subjects consist of men having an average age of 70.9 years. The most prevalent and the vascular risk factor was of 72.7% for the hypertension. Patients had a higher prevalence of mild cognitive impairment than those of the previous groups.

However, there was no difficulty in the daily activities of the patients. The most often reported form of seizures (54.4%) were focal impaired awareness seizures, which are characterized by an epigastric aura followed by unresponsiveness. In 57.5% of LOE patients, the EEG showed no epileptiform activity. Remarkably, in 93.3% of instances, seizures were successfully managed with a single epileptic medication. This proves the medical need for an automated system for detecting epilepsy, such that its early recognition could lead to its early medication.

All components of the seizure prediction scheme are pre-processing, feature extraction and classification of EEG data. A number of academicians have proposed a variety of deep learning and machine learning techniques to exploit EEG scalp signals which are recorded by installing electrodes on patients' heads in order to detect epilepsy. Several scholars have recently presented strategies for predicting seizures involving epilepsy using scalp EEG data. Preictal and interictal state categorization, feature identification, and EEG data processing comprise the three fundamental stages of all these methods.

The importance of EEG signals for researching brain-related disorders is emphasized in this review [17] by U. Rajendra Acharya, S. Vinitha Sree, G. Swapna, Roshan Joy Martis and Jasjit S. Suri, they also discussed about the difficulties posed by the signals non-linearity and the subjective interpretations that follow. The authors provide a thorough introduction of signal analysis approaches, including linear, time-frequency, nonlinear and frequency domain methods, to enable better analysis. The research is primary focused on the field of epilepsy detection, a neurological condition distinguished by its abrupt and erratic symptoms. The authors support automated systems that can classify states as normal, interictal, and ictal and identify seizures in their early stages. They strongly believe that by taking preventative steps, these systems may improve patients quality of life. Their overview presents the results of many automated methods for classifying epileptic activities using EEG as the basis signal. Interestingly, a combination of features from the evaluated methods—in particular, the non-linear features from the EEG segments shows impressive classification accuracies. Even with these developments, the review highlights the unresolved problems and ongoing difficulties that need to be resolved before a fully automated Computer-Aided Detection (CAD) system for seizure monitoring and epilepsy detection can be clinically implemented. They highlighted the importance of ongoing research and development in this crucial field.

A deep CNN model for EEG seizure detection was presented by Hossain, M. Shamim et al. [18] Their approach was able to automatically identify strong and significant EEG characteristics. That encouraged to include their deep learning model in the suggested methodology for end-to-end learning of EEG data. Additionally, their research demonstrates that CNN are a useful tool for brain imaging. It’s common for people with epilepsy to record every single occurrence of there epilepsy in a paper or electronic diary so that they can later on receive an appropriate therapy. Using a publicly available EEG epilepsy dataset from Boston Children’s Hospital, the study evaluates how well a deep CNN trained model can identify seizures. With little sensitivity to patient changes, this model can identify seizure patterns since it has been trained to extract spectral and temporal information from raw EEG data. In order to highlight the distinct qualities of band power attributes that are recognized by the CNN model, all new visualization approaches have been presented. Medical practitioners are able to get brain mapping pictures for additional study quickly by using correlation maps, which establish a connection between spectral amplitude characteristics and output images.

These visualization techniques improve the deep learning model’s findings and interpretability, and are useful tools in therapeutic contexts. When deep CNN model are used to identify seizures in EEG data, accurate and patient-generalizable results are obtained. This study demonstrates how deep learning models can identify strong characteristics from unprocessed EEG data, outperforming traditional techniques in seizure detection. Furthermore, the visualization approaches have been developed to enhance the interpretative ability of the model’s predictions that offers a significant assistance to medical practitioners in the diagnosis and management of epilepsy. Overall, this paper emphasizes how deep learning has significantly advanced EEG based seizure detection and shows how it might improve patient outcomes. [19]

The process known as Bayesian optimization is an effective way to optimize objective functions that take hours or even minutes to examine [20].In function evaluations, it can tolerate stochastic noise; it is best appropriate for optimization over constant domains with fewer than 20 dimensions. Using an acquisition function derived from the surrogate, it establishes an objective surrogate and uses a Bayesian machine learning technique known as Gaussian process regression to assess the unpredictability in the surrogate. Bayesian optimization stands out as a potent method for optimizing objective functions that are both computationally intensive and subject to stochastic noise. This instructional guide delves deeply into Bayesian optimization, offering an extensive examination of its principles, methodologies, and diverse applications across various fields. Covering fundamental concepts as well as advanced techniques, the tutorial furnishes researchers and practitioners with a comprehensive toolkit for effectively harnessing Bayesian optimization in diverse optimization endeavours. Additionally, it underscores the tutorial's contributions to refining and formalizing acquisition functions, highlighting its pivotal role in advancing optimization methodologies. Concluding with insights into available Bayesian optimization software and future research directions, the tutorial stresses the imperative of advancing Bayesian optimization methodologies to tackle evolving challenges and opportunities within optimization and machine learning spheres. It elucidates the foundational principles of Bayesian optimization, particularly its applicability in optimizing objective functions within high-dimensional continuous domains with limited evaluations. The tutorial introduces Gaussian process regression as a surrogate model for objective functions, facilitating uncertainty quantification and informed decision-making processes.

Tinu Theckel Joy, Santu Rana, Sunil Gupta, Svetha Venkatesh [21] gave a detailed and mathematical explanation of the Optimization algorithm, they indicated the efficiency of this algorithm through several bench marked datasets and explained its application on various sate of art technologies. They concluded a test error of less than 0.2% on a CNN algorithm. The authors of the paper introduce an innovative Bayesian optimization framework tailored for hyperparameter tuning, drawing inspiration from principles rooted in statistical learning theory. By employing insights from PAC learning theory, the framework initially optimizes hyperparameters on small subsets of data and then progresses to explore more intricate models using the entire dataset, resulting in enhanced classifier performance. Furthermore, the framework's effectiveness is further reinforced by the deliberate addition of directional derivative signals to the hyperparameter search field. This study incorporates learning theory notions into optimization, which makes a substantial contribution to the progress of hyperparameter tuning approaches. The authors demonstrate the usefulness of their suggested approach in hyperparameter tweaking, which eventually results in the enhanced classifier performance that they explained through experimental validation across a range of machine learning methods. The innovation that they showed by using the directional derivative signs is a remarkable feature of their suggested framework as when we place them in the hyper parameter search, it enables the exploration of more complex models that are consistent with learning theory, insights, which further guide the hyper parameter tuning.

The primary aim of the paper authored by Supriya, S., Siuly, S., Wang, H. et al. [22] is to disseminate knowledge to researchers regarding the current methodologies utilized for detecting epilepsy from EEG data. Their paper provides a concise overview of the existing techniques within the realm of automated epilepsy detection, focusing on various domains of EEG signal analysis including time domain, frequency domain, time–frequency domain, and non-linear approaches. Moreover, the paper delves into the limitations of these current methods, highlighting the need for automated seizure detection techniques. Such techniques would aid clinicians in diagnosing epilepsy through computer-based EEG analysis, ultimately reducing costs, inaccuracies, and the lengthy duration of examinations.

In 1993, H Qu, J Gotman [23] introduced an innovative approach utilizing the K Nearest Neighbour classifier for automated seizure detection. This method was personalized for individual patients, aiming to enhance detection accuracy by leveraging the consistency of EEG recordings unique to each patient. While this strategy proved effective in distinguishing between seizure and non-seizure activities for individual patients, it encountered challenges in latency detection. Qu et al. continuously refined this method over time through multiple revisions. Nonetheless, a notable limitation of patient-specific approaches arose when applied to heterogeneous epileptic patient cohorts, leading to less favorable outcomes. Moreover, in instances of multiple seizures within a single individual, improving sensitivity required the integration of diverse classifiers. Subsequent researchers have since proposed various techniques for epileptic seizure detection, which will be briefly summarized below.

Wavelet transform was used by P. Jahankhani, V. Kodogiannis and K. Revett [24] to extract parameters from EEG data, and a neural network-based classifier was used to classify the signals. They combined an expert model with a wavelet transform-based feature extraction technique to detect epilepsy in EEG recordings. Their results showed that when the expert model was included, accuracy was higher than when the neural network-based model was used alone. To diagnose epilepsy from EEG signals, a method utilizing discrete wavelet transform is employed, which calculates approximation and detail coefficients as features. With a 96% classification accuracy, this technique effectively identified seizure activity. The nonlinear features of EEG signals during ictal activity—which contrast with the Gaussian linear stochastic patterns seen in regular EEG data—were another focus of their study. They also noticed that during epileptic convulsions, approximate entropy decreased. They discovered that when there was an epileptic discharge, entropy measurements dropped.

Polat K, Güneş S. [25] employed a decision tree classifier in combination with the Welch technique based on Fast Fourier Transformation (FFT) to identify epileptic EEG data. Afterwards, they introduced a novel hybrid method that extracts parameters from epileptic EEG data using the Welch FFT methodology and reduces dimensionality using Principal Component Analysis (PCA). An AI recognition system that using this method achieved 95% classification accuracy. They developed a decision tree based logistic model technique for seizure detection. Also, a principal component analysis based optimal allocation technique was offered to differentiate between normal and epileptic EEG data. Their study's objective was to minimize the dimensionality of the dataset and generate independent components using Principal Component Analysis. In addition, they presented a novel technique based on time-frequency (T-F) pictures for the diagnosis of epilepsy from EEG signals. This advanced approach consistently produces high quality results by using the Fisher Vector as an encoder and the Grey Level Co-occurrence Matrix as a descriptor.

Belhadj S, Attia A, Adnane AB, Ahmed-Foitih Z, and Taleb AA [26] introduced an unsupervised clustering approach for epilepsy identification, employing potential-based hierarchical agglomerative clustering alongside empirical mode decomposition. Together with the Kolmogorov distance using the Bhattacharya distance, the Euclidean distance between the intrinsic mode functions (IMFs) was computed and supplied as input for the clustering method. They reported an accuracy of 98.84% in categorization with this strategy utilizing the CHB-MIT epileptic database. They created a seizure detection processor with wavelet energy as a parameter by utilizing an SVM classifier. Under the direction of knowledgeable neurologists examined two forms of epileptic seizures: partial epilepsy and primary generalized epileptic disease. Using a multi-layer perceptron neural network classifier and a radial basis function neural network classifier, they obtained 95.2% and 89.2% accuracy, respectively. They classified EEG signals into normal, interictal, and ictal forms of epilepsy using the Largest Lyapunov Exponent parameter for both feed-forward and recurrent neural networks. More encouraging outcomes were obtained by the recurrent neural network, which achieved a 96% classification accuracy overall, 97.38% specificity, and 96% classification sensitivity.

A Douglas–Peucker algorithm (DP)-based methodology for epilepsy identification from raw EEG data was suggested by Zarei R, He J, Siuly S, Huang G, Zhang Y [27]. In order to minimize dimensionality and find uncorrelated variables, principal component analysis, or PCA, was utilized. The University of Bonn's epileptic EEG patient database was used for the experiments, and four machine learning classifiers decision tree, k-NN, random forest (RF) and SVM classifiers were used to assess performance. Larger EEG signal data volumes cause this framework's computing complexity to grow, which is a disadvantage.

They developed a novel method for detecting epilepsy by breaking down epileptic EEG signals into Q, R, and J levels using the tunable Q-factor wavelet transform (TQWT) and five sub-bands and. From each epoch, ten statistical signals were taken out and evaluated with SVM, k-NN, and bagging tree (BT) classifiers. With 3750 samples of Bonn University focal and non-focal epileptic data, there approach achieved good accuracy with the epileptic EEG data. Although there are implementation issues in real-time systems, its main benefit is reduced computing costs and data.

A technique for detecting epileptic seizures was presented by them which used the Information Gain (InfoGain) algorithm on fast Fourier transform (FFT) and discrete wavelet transform (DWT) separately. Using the LS-SVM classifier, the excellent accuracy indicates that seizure activity may be detected with efficacy when FFT and InfoGain are combined.

III. METHODOLOGY

We have applied two domains of artificial intelligence that are machine learning and deep learning in this report. We have used several machine learning and algorithms on the freely available dataset our goal is to find the best machine learning algorithm to detect epileptic signals in the real time and at the end of the report, we conclude that decision tree is the best algorithm in the domain of machine learning for detecting the epileptic brain signals, a table with multiple factors of evaluation is shown in the results part that provides us with the accuracy of different machine learning algorithms. Further we have also use deep learning algorithms so that we can read the epileptic signals more deeply, though it will require a significant use of extensive hardware, but the results provided by the deep learning algorithms would be also significantly much better that we can also see the conclusion table 5.1 we have used an ensemble technique that combines one dimensional convolutional neural network along with a long short term memory, deep learning algorithm.

The goal of the suggested architecture is to create deep learning model that is accurate and reliable in identifying epileptic episodes. This is made possible by the separation of two types of brain states into interictal and ictal. The model proposed in this study is an ensemble model, which is combination of 1D-CNN followed by LSTM. Prior to the introduction of the 1D-CNN and LSTM, initially a pre-processing of the raw EEG is necessary. Next, the 1D-CNN LSTM model is created and used to identify epileptic seizures. The initial data set was pre-processed and reorganized by a UCI official, as explained more in section below “Freely Accessible Dataset” Therefore, a normalization of the EEG signal data is done in the pre-processing step which is acquired from the UCI dataset set before feeding it to the suggested model.

A. Freely Acessible Datasets

The utilization of dataset is crucial for data scientists and academics to evaluate the success of the models they have presented. The detection of a tumour should similarly pick up on our brain signals. The most popular way to track brain activity is through EEG recordings. These recordings are crucial for machine learning classifications that investigate novel techniques for detecting tumours in a variety of ways, including early tumours detection, quick tumour detection, patient tumour detection, and tumour localization. Data sets that are accessible to the general public are crucial for analysis, comparison, and inference. We will go through the well-known dataset frequently utilized in epilepsy in the part after that.

1) *BONN University-EEG Dataset:*

The BONN EEG Time Series Epilepsy Dataset constitutes an important tool for epilepsy research and neurology. The dataset [28] was developed at the University of Bonn in Germany towards enhancing computational analysis of epileptic seizure and improve its detection. Here are some more detailed aspects of the dataset. Data Source: Two major sources of the dataset are; EEG recordings.

a) Epileptic Patients:

Epileptic EEG data from people. These recordings are very valuable for understanding epileptic seizures because they document the activity at the level of the brain during such events.

b) Vigorous Individuals:

Control: Data from EEG recordings of humans who do not have epileptic seizures.

Annotations: Annotation has been applied in this dataset, indicating epileptic seizures and other events worthy of note. Such annotations are important for training and testing of automated seizure detection software in EEG data.

Contributions: With the introduction of the BONN EEG Time Series Epilepsy Dataset, it is possible to develop computer-aided tools for epilepsy diagnosis and management. This allowed refining the algorithms that had the effect of giving better results when working with other patients having this condition.

This dataset consists of 100 single-channel EEG recordings, each lasting 23.6 seconds and sampled at a rate of 173.61 Hz. The spectral bandwidth of the data ranges from 0.5 Hz to 85 Hz, and it was originally obtained using a 128-channel acquisition system. These EEG recordings were extracted from larger multi-channel EEG recordings of five patients and designated as Sets A to E.

- Sets A and B represent surface EEG recordings during periods of closed and open eyes, respectively, in healthy patients.

- Sets C and D comprise intracranial EEG recordings, with C obtained from a seizure-free zone within an epileptic patient's brain and D from a non-seizure-generating area of the same patient.

- Set E contains intracranial EEG data from an epileptic patient captured during epileptic seizures.

Each set contains 100 text files, each with 4097 samples representing a single EEG time series in ASCII code format. The data has undergone bandpass filtering with cut-off frequencies at 0.53 Hz and 40 Hz. It is noteworthy that this dataset is devoid of artifacts, and thus, no prior pre-processing steps are necessary for classifying healthy (non-epileptic) and unhealthy (epileptic) signals. Strong eye movement artifacts have been removed. This dataset was made publicly available in 2001 and has been extended as part of the EPILEPSIA project.

Figure 3.1 BONN University [28]

Indeed, the dataset is very important because it provides an opportunity to conduct further investigations into epilepsy which translates into development of effective computational tools meant.

TABLE 3.1 SUMMARY OF BONN DATASET

<table>
<tr><th>Set</th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th></tr>
<tr><td>Subject</td><td>Vigorous</td><td>Vigorous</td><td>Epilepsy</td><td>Epilepsy</td><td>Epilepsy</td></tr>
<tr><td>Subject Condition during Readings</td><td>Not asleep with eyes opened</td><td>Not asleep with eyes closed</td><td colspan="2">Seizure-free (interictal)</td><td>Seizure-free (ictal)</td></tr>
<tr><td>Electrode Type</td><td colspan="2">Surface</td><td colspan="3">Intracranial</td></tr>
<tr><td>Electrode Placement</td><td colspan="5">International 10-20 System</td></tr>
<tr><td>Channels</td><td colspan="5">100</td></tr>
<tr><td>Duration</td><td colspan="5">23.6 Seconds</td></tr>
</table>

2) *UCI Machine Learning Dataset*

Figure 3.2 UCI Machine Learning Repository[10]

Each of the five folders in the original dataset [10] has one hundred files, each of which represents a particular topic or individual. Every file contains a 23.6-second observation of neural activity. Data points totalling 4097 are collected from the related time-series. The value of the EEG recording at a particular moment in time is represented by each data point. There are five hundred distinct individuals in all, and every one of them having 4097 data points for 23.5 seconds. The 4097 data points were split up into 23 segments, with each segment holding 178 data points in a single second. Each segment had an EEG record value recorded at a distinct time period. Which gives 23 x 500 = 11500 data points in 1 second (column) for each item, and a label with y = 1, 2, 3, 4, and 5 in the last row. Thus, subject from class 2 to 5 are categorized as non-epileptic EEG signals and category 1 belongs to epileptic EEG signals.

TABLE 3.2 SUMMARY OF UCI DATASET

Subject Category	**Subject State**	**Epileptic/Not Epileptic**
1	*Epileptic Patient*	*Epileptic*
2	*Brain With Tumour*	*Not Epileptic*
3	*Healthy Brain*	*Not Epileptic*
4	*Eyes Closed*	*Not Epileptic*
5	*Eyes Open*	*Not Epileptic*

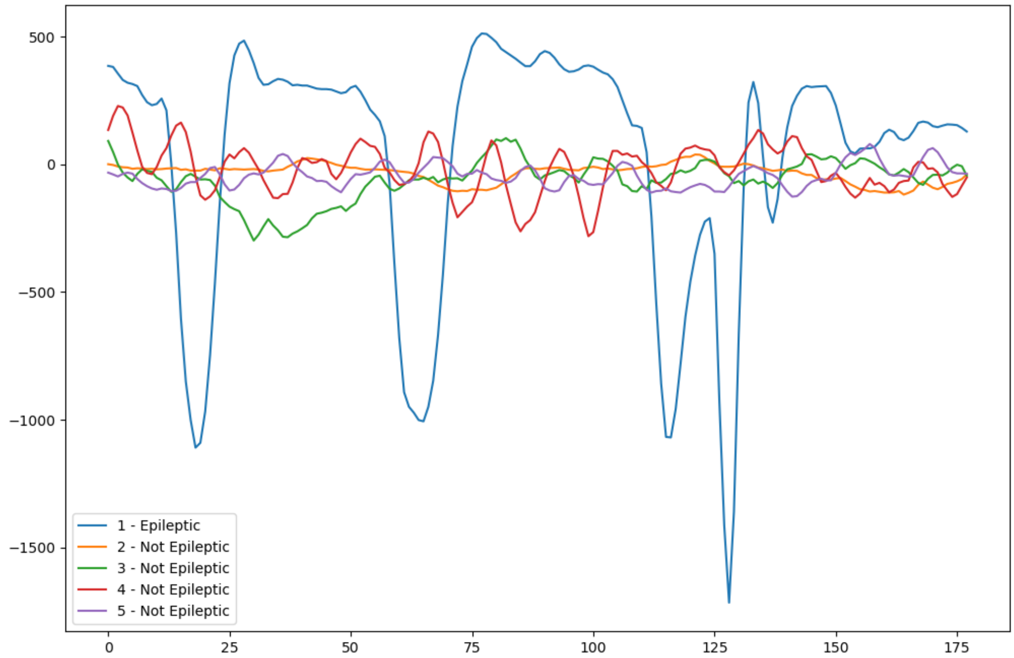


Figure 3.3 The raw EEG signal waveform of four healthy subjects and one epileptic subject.

B. 10-20 Electrode System

A 10-20 electrode arrangement serves as a standardized technique for the strategic placement of electrodes on the human scalp, primarily for electroencephalography (EEG) measurements as shown in Fig 3.4 It divides the scalp into defined zones and positions electrodes at precise coordinates relative to anatomical landmarks. The nomenclature "10-20" signifies that the separation between these landmarks are uniformly either 10% or 20% of the total measurements from right to left or front to back on the skull. This approach offers a consistent methodology for capturing brainwave activity and finds extensive application in clinical and scientific investigations, supporting the examination and exploration of neurological conditions and brain functioning.

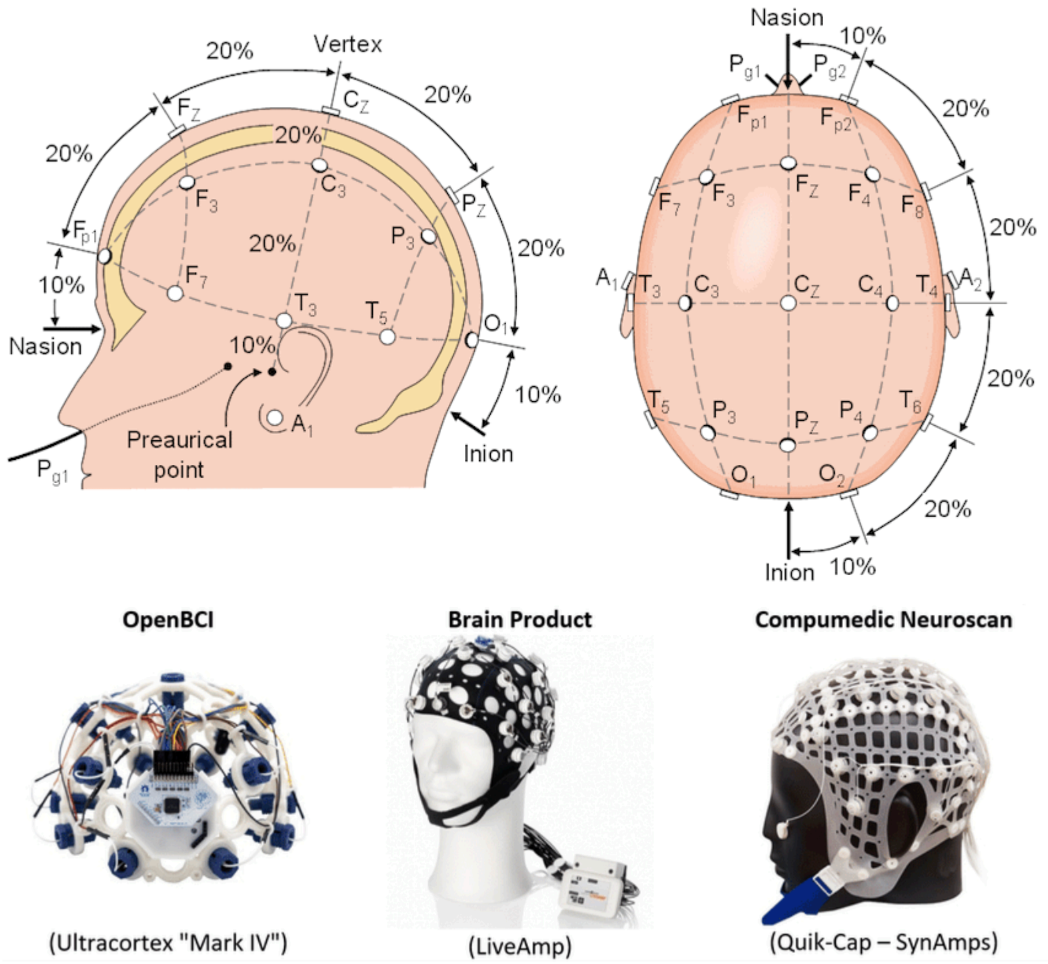


Figure 3.4 The 10-20 system with front-back (nasion to inion) 10% and 20% electrode separation [29]

The 10-20 electrode system is a standardized method used for the placement of electrodes on the scalp for electroencephalography (EEG) recordings. Figure 3.4 shows the equipment which is used to measure the brain's electrical activity in a non-invasive manner, it is used very commonly by the neurologist in their clinics to understand the functioning of the brain in real time.

The 10-20 electrode ensures that there is similar placement of the equipment for all the patients, this method is medically proven. Its name is derived from the percentage percentages of certain cranial lens that are used to quantify the relative distance is between the reference points as shown on the scalp of the statue in figure 3.4.

The 10-20 electrode system defines the nasion and the inion as the primary point of the reference for all the distance or the percentages. The outward bump at the base of this skull is called inion and the nose, where the frontal and the nasal bones meet is called nasion. These two points are the primary reference point for this system that could be understood by any researcher of this field.

10% and 20% respectively refers to the distance between the nasion, inion and distance along the sides of the skull (mastoid). The setting of the electrode placement is done in respect to these percentages, such as if the distance between the nasion, inion is of 10 equal segments. Then the electrodes are positioned at a specific percentage along the line.

The complete working of the 10–20 electrode system is shown below:

1) *Measurement Point*

The nation, inion are the two main reference points of the system, placement of electrodes is determined with the help of the separation or the distance between these sides and certain points on the scalp.

2) *Measures of Percentage:*

Measurement of percentage is made along the length of the side of the head, which are referred as mastoid and separation between the nasion and inion. The exact location of the electrode is calculated by dividing the distance into two distinct ratios that is of 10 and 20.

3) *Naming Convention*

The names to the electrode placement is in respect to the hemisphere and mid line, where hemispheres represents the even numbers and the midline represents the odd numbers such as C_Z electrode will be situated at the upper center of the skull and F_{p1} would be situated in the front polar area of the left hemisphere.

4) Standardization

10-20 system offers a standardization among the researchers of whole world to use a single system for understanding the brain signals, it ensures that the electro placement is consistent throughout the world, which acts as a mathematical language for understanding the brain signals. That's why offering a unique framework for positioning of EEG electrodes it establishes a standardization in the medical field.

5) Even Distribution

The 10-20 system also ensures that the brain signals are collected evenly from the brain as the distribution of the electrodes is done evenly around the scalp because each electrode is placed on a distinct area of the brain we get the least overlapping signals. Moreover, the recordings are also made at different times that ensures any disturbance in the signal.

The evenly spaced electrodes around the scalp ensures to capture the complete electrical activity from all the parts of the brain. Because if different researchers use different variety of electrode placement, we cannot compare the signals of even a single patient that it is necessary to have a standardization for the accurate EG data interpretation.

The greatest advantage of the 10-20 electrode system is that it is globally accepted by all the researchers and physicians or neurologist, which facilitates the communication about the brain signals even cross-border. It's acceptance by neurologist also gives it a medical support and helps the machine learning researchers to collaborate with the worldwide neurologist. Thus, acts as a common language for anyone who is trying to understand the brain signals.

It is not only restricted to a specific combination. It provides a researcher freedom to alter the electrode combinations according to his own research or the theoretical requirements, do we know 10–20 electrode system is accepted worldwide, but one can specify his own system being used on the basis of the percentage or the ratio of the distances.

a) Advantages:

- Global Acceptance: Researchers or medical professionals, irrespective of their country or language spoken follows a consistent system for recording the brain electrical signals.

- Customization: It allows different position of the electrodes, if one want to specifically target a portion of a brain according to his area of interest.

- Persistent Recording: provided a constant electrical supply ensures that electrical signals from the brain arch captured without any disturbance throughout the scalp.

- Reusability: Multiple EG experiments can be done with the help of minor changes in the clinical set up.

b) Limitations:

- Limited Coverage: The signal is coming from the front, and the back of the head are not as good as the signal captured by other systems.

- Intricate: It requires very deep knowledge of understanding the human scalp, which every machine learning researcher might not have.

This shows that the 10–20 electrode system is a standard approach which is used widely across the world for recording the EEG signals of a brain that also have medical support and proves. The guaranteed and the uniform placement makes it easier to read the data of any patient from various sources. Studies and subjects. Thus, proving it is still a relevant mechanism that is being used for recording the brain electrical activity by both the researchers and the neurologist despite its drawback. It's the best method developed clinically yet.

C. Experimental Setup

The hardware used here is of an apple MacBook Air with M1GB chipset having an integrated graphic card that consist of 8 cores, it is an integrated part of the recently developed chip named as M1 SoC by Apple. We have used Keras version 2.12.0 and Python version 3.7 for all the algorithms applied are hardware or version of the library used hasn't changed.

A 90-10 train test split on the data is performed throughout the experiment. The number of training epochs for the CNN, 1D convolutional LSTM, and deep neural network (DNN) models is 100. The suggested method also employs the dropout strategy to enhance the generalization performance and prevent the issue of overfitting. Distribution of the data in a random manner is done before training, and then subsequently forwarded to the network. Furthermore, checkpoints are incorporated into the training process. During training, the model's accuracy for each epoch's training data set and test data set shall be calculated, that at the end of each epoch to allow us to assess whether that model is overfitted or in order to verify its generalization potential. Should the model's capacity to generalize not increase after ten training cycles, the learning rate will be reconciled.

D. Hyperparameter-tuning

A tabular model is employed for epilepsy detection using Fastai's tabular learner. The model architecture is determined by hyperparameters such as the learning rate (lr), weight decay (wd), dropout rate (dp), and the number and sizes of layers. Through rigorous trials of various combinations the Bayesian Optimization instance is configured to maximize the accuracy by iteratively exploring the hyperparameter space. It conducts 100 iterations to discover the hyperparameter metrics that yield the highest accuracy on the validation set. A proxy function is used to find the minimum objective value based on previous metrics of the objective function

$$X_{s,b}^{*} = \arg max_{x \in \mathbb{X}} X f_{s,b}(x) \tag{1}$$

The performance of the model trained on subset b on the validation dataset for a hyperparameter configuration x is shown by equation 1 by the notation $f_{s,b}(x)$. To represent everything pertaining to a subset of data, 's' is utilized. If these ideal hyperparameters are developed with a limited amount of data, they may be noisy. Thus, Bayesian optimization is performed on several smaller subsets to select a robust estimate of the hyperparameter, and then select the best hyperparameters.

A key component of developing machine learning models is hyperparameter optimization, which focuses on optimizing parameters that have a big impact on the model's performance but aren't explicitly learnt during training. The computing efficiency accuracy and the generalizability of the model are highly dependent on learning rate, regularization strengths, and the tree depth specified while defining the model.

In hyper parameter optimization, the idea is to determine an ideal collection of hyper parameters which produces the best model performance. For this a predetermined search space must be explored first. Other evolutionary algorithms, such as random search and great search are also used for this.

The reason for not using the great searches that it is computationally very expensive, and even after that, it remains efficient for only a limited search space because it examine every possible combination of the hyper parameters within the set limits, which is itself a theoretically and practically a very lengthy process.

Thus, we need to use the probabilistic models for searching where comes the bay Bayesian Optimization as it is the best probabilistic model. It dynamically chooses the hyper parameter based on the previous experiences to effectively explore the search space. In order to develop the hyper parameter settings for an optimal solution evolutionary algorithm, imitate the principle of natural selection.

Hyper parameter optimization aims to maximise the accuracy of the model while preventing the overfitting. Thus, it finds a balance between the models complexity and generalized performance.

Target variables, also known as hyperparameter measurements, are optimised by hyperparameter tuning. Model correctness is a commonly used statistic that is determined by an assessment pass. Numeric metrics are required.

Establishing the purpose and label for every metric is essential when setting up a hyperparameter tuning task. Whether you wish to optimise your model to maximise or minimise the value of this statistic is specified in the aim.

E. Classifiers Theory

1) Decision Tree

A decision tree approach could be useful in detecting the availability as it is a good classifier by recursively dividing the data. According to the distant qualities, we can create a tree like structure which can further be used to identify whether the patient is epileptic or non-epileptic based on the distinct characteristics shown by the data in the training phase, this approach makes the use of internal structure tree as a tool for the decision making. Epilepsy may be diagnosed using EEG using a decision tree approach, which is frequently used for classification tasks. A tree like structure is produced by recursively splitting the data based on unique attributes, and this structure is utilized to determine the class labels of individual instances. This method functions by exploiting the internal structure as a decision making tool. In order to create a decision tree, the recursive process involves determining which characteristic to use to separate the data at each node.

In order to get homogeneous subsets, it is necessary to decrease the disorder and impurity in the data, which may be done with the criterion measure. Until every sample in a node has the same class label, the recursive process goes on for every subset. It then finally stops until a stopping condition such as the maximum depth or the minimum number of samples per leaf is met. The epileptic or non-epileptic status of fresh EEG data may be determined by moving up the decision tree from the root node to a leaf node. This is the mathematical justification for the classification procedure that follows decision tree construction as shown in equation 2.

$$Entropy(S) = \sum_{i=1}^{N} p_i \log_2 p_i \tag{2}$$

N = count of unique class values

Pi = event probability

2) K-Nearest Neighbours

KNN is k-nearest neighbours algorithm otherwise known as a supervised learner with nonparametric characteristics. This involves determining an approximate class or value of a data point by comparing it to other data points. It is applicable in both regression and classification purposes but the general use of clustering similar points makes this tool mainly a classifier. “K” in KNN stands for the number of nearest neighbours, which is taken into account in case of a certain record classification. Choice of ‘K’ depends upon various parameters of the input data. Most of such data generally benefit from a higher ‘K’ value. For a classification technique, it’s usually advisable to use one ‘K’ value for this purpose; besides, some cross validation methods can help choose the best ‘K’ for a dataset.

3) Logistic Regression

For binary classification issues, logistic regression is a popular statistical and machine learning technique.

$$Euclidean = \sqrt{\sum_{i=1}^{i=k}(x_i - y_i)^2} \qquad (3)$$

Based on a variety of input variables, it estimates the likelihood that an output will fall into one of two classifications. Logistic regression limits its output to a range between 0 and 1, reflecting probabilities. Logistic regression measures the effect of each input feature on the likelihood of class membership by calculating coefficients for each feature. It can be understood, is computationally effective, and acts as a base for more sophisticated methods. Numerous industries, including as healthcare, finance, and marketing, use logistic regression because decision-making and predictive modelling require an understanding of the likelihood of binary outcomes.

4) Naïve Bayes

Naïve Bayes is an easy-to-use probabilistic classification technique for simple applications. The latter relies on the Bayes theorem and provides the probability for a point to belong to a specific class. The naive assumption is that every attribute is independent and makes calculation easier, but this notion does not hold true in real cases. While it is somewhat oversimplified, still many techniques applied in the text classification of spam and opinion mining lose against naive Bayes. This particular algorithm is specifically ideal for high dimensional datasets that have just enough labelled data. Naive Bayes is of considerable importance because of the capability of handling multiple classifications as well as the ease with which it can be trained and implemented for machine learning and many natural language processing applications.

$$P\left(\frac{C}{X}\right) = \frac{P\left(\frac{X}{C}\right) \times P(C)}{P(X)} \tag{4}$$

P(C/X) = Posterior Probability

P(X/C) = Likelihood

P(C) = Class Prior Probability

P(X) = Predictor Prior Probability

5) Random Forest

It is an excellent machine learning’s ensemble learning technique. This method will result in good and accurate prediction by incorporating several decision making trees in it. The second point is that trees are randomly train-ed one at a time on a specifically chosen portion of the data, based on randomly selected features reducing thus overfitting and improving generalizedness. In this case, the final forecast is created using projections of different trees. Random Forest is able to provide accurate and highly reliable predictions on many applications which include both classification and regression. One such tool exists for numerous branches of study such as image classification, finances, and healthcare and needs minor changes done in respects with hyperparameters unlike a solitary decision tree.

6) Support Vector Machine

Creating an ideal hyperplane that divides the data into two groups of two. SVM has advantage in high dimensional domain; thus appropriate for tasks like text and image classification. The algorithm operates on such kernel functions as radial basis function (RBF) kernel for handling both linearly and non-linearly classified data. The biggest advantage of SVM is that it is very robust, particularly when working with small and unbalanced datasets.

7) 1D-CNN

In order to obtain representations and effective features from 1D time series convolution sequence data, a 1D-CNN may perform 1D operations with different filters. Fig 3.5 shows the technicalities of the 1D-CNN process. To confirm to the single dimensional nature of the raw EEG signal data, the feature maps and convolutional filters of the 1DCNNs used in this work are entirely one dimensional. By increasing the number of convolutional layers, CNN is capable of progressively generating higher level features for epileptic seizure detection tasks which are resistant and discriminable.

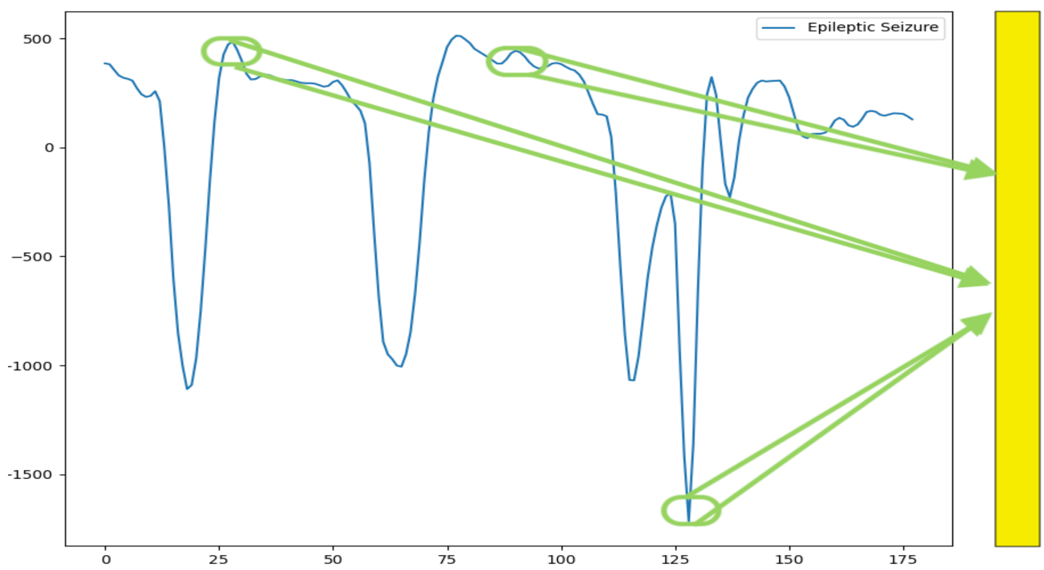


Figure 3.5 1D Convolution operation

8) LSTM

The standard LSTM block structure is shown in Fig 3.6 The LSTM block consists of four gates: input gate zi, here, a sigmoid function receives the input states the previous hidden state and the current input state and determines which values should be updated by converting them to a range of 0 to 1. One indicates importance, whereas zero indicates not much; forget gate zf, this gate determines what data should be retained or discarded. The sigmoid function processes data from the present input as well as data from the prior hidden state; gate z in the cell state that retains the data throughout time; and output gate zo, which determines the value of the subsequent hidden state keeping in mind that information about prior inputs is contained in the concealed state and predictions are also made using the concealed state.

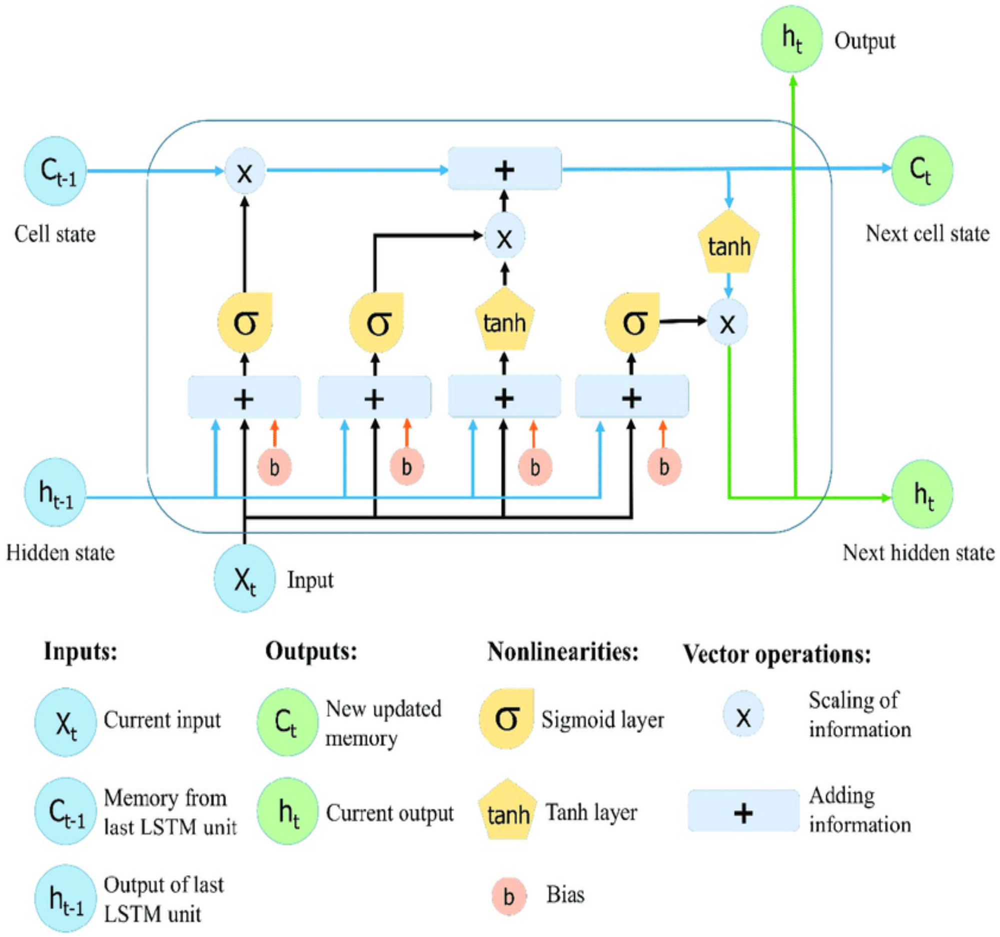


Figure 3.6 LSTM Block Structure [30]

9) 1D-CNN LSTM

4 convolutional layers, 2 LSTM layers, 1 input layer, 1 pooling layer, 4 fully connected (FC) layers, and a SoftMax output layer make up the suggested ensemble model. First, as the data source for the proposed model, the 45 X 1 form of the one-dimensional EEG signal data is used after that, to extract abstraction features from the raw signal data, input data is transmitted through an initial convolution layer composed of 64 one dimensional convolution kernels with a shape of 3 X 1 and a length of 1, respectively. A ReLU (Rectified Linear Unit) activation layer comes after this convolutional layer, which adds non-linearity to the suggested model. Here, the one-dimensional convolutional operation and the ReLU activation are defined mathematically as follows:

$$y_i^k = \sigma\left(\sum_{j=1}^{N_{k-1}} b_i^k + \mathrm{conv1D}\left(w_{j,i}^k , x_j^{k-1} \right) \right) \tag{5}$$

In the kth layer is the i^{th} feature map; the activation function ReLU, which can assist prevent over fitting, is represented by $\sigma()$; $w_{j,i}^k$ is the trainable convolutional kernel; In the $(k-1)^{th}$ layer x_j^{k-1} is the jth feature map; where N^{k-1} depicts how many feature maps are there in the $(k-1)^{th}$ layer; Since conv1D is a representation of the one-dimensional convolution process without zero-padding, the size of feature map in the kth layer is smaller than its corresponding dimension in the $(k-1)^{th}$ layer.

After the convolution and activation, (45 X 1) sized 64 feature maps are produced. Subsequently, a max-pooling layer receives the output of convolutional layer 1. The following is a description of the one-dimensional max-pooling operation's mathematical definition:

$$p_i^a = \max\,(p_i^{a'} : a < a' < (a + s) \tag{6}$$

Here "s" represents pooling window size; max pooling action leads to pai which represents the a^{th} neuron; before that $p_i^{a'}$ represents a'^{th} neuron in the i^{th} feature map.

Both the size and the stride of the pooling windows in the Pooling Layer one are 2. It can speed up the training process and drastically lower the overall training parameters in the suggested model. 21 × 64 is the size of 64 feature maps produced after pooling. Subsequently, 3 convolutional layers are employed to additionally extract advanced characteristics that may aid in categorization. ReLU is also used for convolution procedure and the non-linear activation.

After passing through each of the one-dimensional convolution layers, the resulting 1024 feature maps, each measuring 43 X 64, will be fed into a single 256 neuron FC layer, after which a dropout of 0.3 will be applied to the FC's output. With aim of fitting the results of LSTM layers, FC Layer 1 can integrate the results of the convolution layers, minimize the size of feature maps, and, to some degree, mitigate the overfitting problems through dropout.

In order to prevent the prolonged dependence of the conventional RNN, the output features are sent to the LSTM layers after going across the FC layer1. There are 4 gates in the LSTM cell: forget, input data, output logic and cell state gate. In order to protect the earlier data and to increase the capacity to obtain meaningful incites from the EEG time series data, they can cooperate. Both the LSTM layers 1 and 2 have 64 neurons each.

After their passage by the LSTM layers, result characteristics shall be supplied at three FC levels. Lastly, a layer of SoftMax output is applied in the ensemble model for better results. In line with the specific results produced by Bayesian hyperparameter optimization, the proposed model's more detailed structure has been modified. The architecture depicted in Fig. 3.7 [31] is employed, when the recommended model has been successfully built and trained, to recognize epileptic seizure activity.

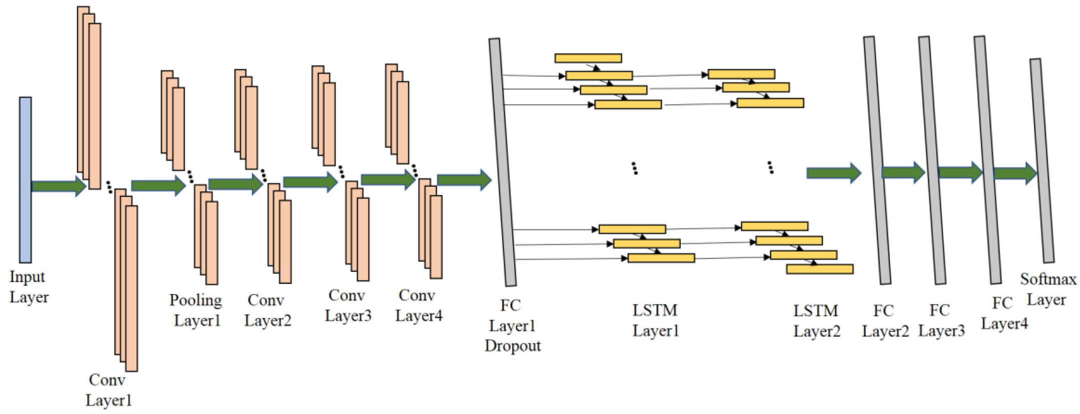


Figure 3.7 1D-CNN LSTM Model [31]

Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are two potent deep learning architectures that are combined in a 1D-CNN LSTM ensemble model to handle sequential and temporal data processing tasks. It uses a group of CNN layers to capture the special information and then uses the current layers to collect the temporal information and then utilize the advantages of both CNN and LSTM. A deep explanation is given below:

a) 1D Convolutional Neural Networks (CNNs):

One dimensional filter is used to extract the features from the input sequence when we are working with one dimensional data like in the case of epileptic signals, these filters uses the convolution operation to capture the local changes in the patterns when they move over the data. This is the reason it is mostly suited for the tasks like classification. It has an exceptional ability to recognize the special patterns in the data which is most important in capturing the special or the temporal differences between the signals. Pooling layers, such as Maxpooling is used to down sample the features to control the computational complexity.

b) Long Short Term Memory (LSTM) Networks:

LS team is an advanced version of an RNA that is a neural network, which is crafted, specially to capture the long range relationships and the temporal nuance contained in special sequential data. The idea to use an LSTM is because of its memory cells and getting mechanisms which are designed to control the information flow within the network, making it the best algorithm available for the classification of sequential data. Because of its unique architectural design, it is able to efficiently store the important data for a long period of time compare to other algorithms where we face the issue of vanishing gradient.

This special feature of maintaining the memory over such a long period of time is one of the most astonishing feature of the LSTM this makes time series forecasting very easy, which helps in understanding temporary relationships within the data. This is why LSTM are the most important part of a deep learning algorithm, providing it the unmatched efficacy in the task that involve sequential data interpretation or classification.

c) Ensemble Approach:

The objective of a 1D-CNN LSTM ensemble model is to enhance prediction performance by merging the complementing advantages of CNNs and LSTMs. This is accomplished by utilizing the same input data to train independent CNN and LSTM models and then finally combining their predictions using an ensemble approach (e.g. weighted combination or average).

10) Workflow of 1D-CNN-LSTM Ensemble Model:

a) Input Data Preparation

Pre-processing and model training are done on the input data, which might be temporal or sequential data represented as one-dimensional signals. Normalization, feature extraction, and segmentation could be required for this.

b) CNN Model Training

The input data is then processed into a CNN model, which consists of pooling layers after one or more convolutional layers as shown in figure 3.7 then through convolution and pooling processes the CNN gains the ability to extract spatial characteristics from the input sequence. The CNN model finally generates a series of feature maps collecting relevant patterns in the data.

c) LSTM Model Training

In parallel, an LSTM network is trained to recognize long-range patterns and temporal relationships in sequential data by running the same input data through it. By adjusting its internal state in response to both the current input and earlier states, the LSTM model iteratively processes the input sequence.

d) Ensemble Combination

Following training, an ensemble technique is used to integrate the predictions of the CNN and LSTM models. A weighted combination depending on each model's performance on validation data could be used, or the predictions from the two models might be averaged.

e) Evaluation and Prediction

To measure the performance of the ensemble model, an independent test dataset is used. After that it may be used to forecast fresh or unknown data utilizing the complementary abilities of LSTMs and CNNs to provide predictions that are more reliable and accurate.

In conclusion, a 1D-CNN LSTM ensemble model combines the temporal modelling or complementing skills of LSTMs with the spatial feature extraction capabilities of CNNs to produce a potent framework for the analysis of sequential and temporal data. The ensemble strategy takes advantage of the complimentary characteristics of both models to enhance overall predictive performance by integrating their forecasts.

IV. Experimental Results

Information on the experimental outcomes derived from the used machine learning methods is presented in this section. Tables 4.1 demonstrate training with 70% of the data and testing with 30% . Table 4.2 demonstrate training with 60% of the data and testing with 40%. Now further in Deep Learning this report shows Validation, Training Loss, and Validation and Training Accuracy of suggested method i.e. 1D-CNN LSTM ensemble model is shown in Fig 4.1 Additionally, 2 deep learning models—a conventional CNN and a DNN—for the identification of epilepsy have been created and may be compared with the proposed model. Finally, Table 4.3 compares and calculates the accuracy, precision, recall, and F1-score metrics to further assess the seizure categorization performance of these three models.

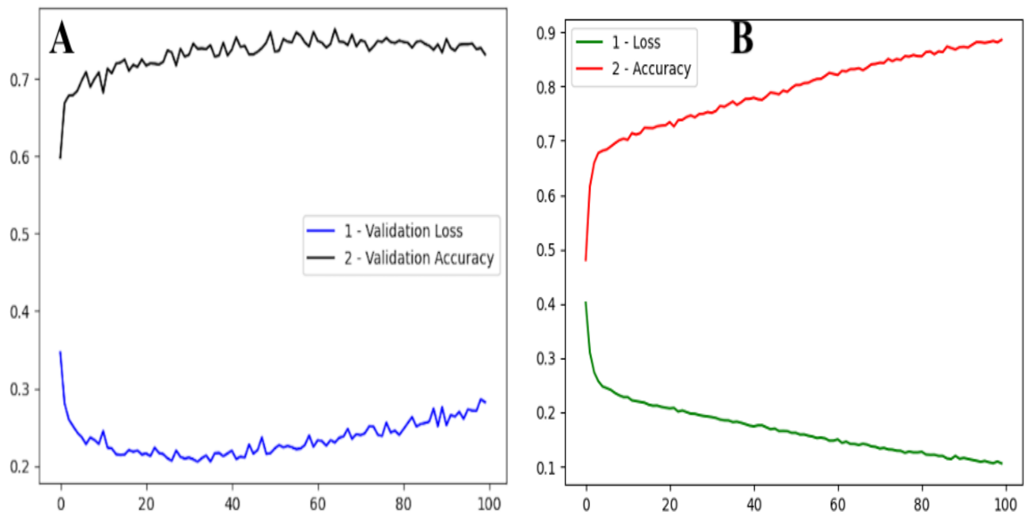


Figure 4.1 Training and Validation Accuracies (A) Validation Loss and Accuracy (B) Training Loss and Accuracy

A. Performance Measures

Here's a brief explanation of the terms in a classification report:

1) Precision

Measures the accuracy of positive predictions.

$$Precision = \frac{TP}{(TP+FP)} \tag{7}$$

2) Recall

Measures the ability of the model to correctly identify all positive instances.

$$Recall = \frac{TP}{(TP+FN)} \tag{8}$$

3) F1-score

It's the harmonic mean of precision and recall and is useful when you want to balance both FP and false negatives.

$$F - 1\ Score = 2 \times \frac{Recall\ X\ Precision}{(Recall+Precision)} \tag{9}$$

4) Support

The number of samples in each class, which can help you understand the dataset's class distribution.

5) Accuracy

It assesses the overall accuracy of a model's predictions, computed as the proportion of correctly predicted instances to the total number of instances.

$$Accuracy\ Score = \frac{TP+TN}{TP+FN+TN+FP} \tag{10}$$

TP= True Positives

FP= False Positives

FN= False Negatives

6) Macro Average

Macro average is a way to calculate an average of a metric (e.g., precision, recall, F1-score) across multiple classes in a multi-class classification problem.

$$Macro\ Avg = \frac{1}{N}\sum_{i=1}^{N} Metric_i \tag{11}$$

7) Weighted Average

Unlike macro average, weighted average takes into account the class distribution. Classes with more instances have a greater influence on the weighted average than classes with fewer instances. It calculates the metric for each class, but the contribution of each class to the weighted average is proportional to the number of instances in that class.

$$Weighted\ Avg = \frac{1}{N}\sum_{1}^{N} \frac{Metric_i X Support_i}{Total\ Support} \tag{12}$$

"N" is the total number of classes.

Metric is the recall, precision, F1-score for class i.

Support is the number of instances in class i.

Total Support is the total number of instances in the dataset.

TABLE 4.1 SEIZURE DETECTION ON BONN DATASET USING 70-30 SPLIT ML CLASSIFIERS

Metric	Classifier	Precision	F1-score	Support	Accuracy
Baseline	Decision Tree	*0.857*	*0.923*	*42.0*	*0.889*
	K-Nearest Neighbors	*0.737*	*0.848*	*42.0*	*0.762*
	Logistic Regression	*0.894*	*0.944*	*42.0*	*0.921*
	Naive Bayes	*0.933*	*0.966*	*42.0*	*0.952*
	Random Forest	*0.857*	*0.923*	*42.0*	*0.889*
	Support Vector Machine	*0.84*	*0.913*	*42.0*	*0.873*
Seizure	Decision Tree	*1.0*	*0.8*	*21.0*	*0.889*
	K-Nearest Neighbors	*1.0*	*0.444*	*21.0*	*0.762*
	Logistic Regression	*1.0*	*0.865*	*21.0*	*0.921*
	Naive Bayes	*1.0*	*0.923*	*21.0*	*0.952*
	Random Forest	*1.0*	*0.8*	*21.0*	*0.889*
	Support Vector Machine	*1.0*	*0.765*	*21.0*	*0.873*

Metric	Classifier	Recall	Macro Avg	Weighted Avg
Baseline	Decision Tree	*1.0*	*0.862*	*0.882*
	K-Nearest Neighbors	*1.0*	*0.646*	*0.714*
	Logistic Regression	*1.0*	*0.904*	*0.918*
	Naive Bayes	*1.0*	*0.944*	*0.951*
	Random Forest	*1.0*	*0.862*	*0.882*
	Support Vector Machine	*1.0*	*0.839*	*0.864*
Seizure	Decision Tree	*0.667*	*0.833*	*0.889*
	K-Nearest Neighbors	*0.286*	*0.643*	*0.762*
	Logistic Regression	*0.762*	*0.881*	*0.921*
	Naive Bayes	*0.857*	*0.929*	*0.952*
	Random Forest	*0.667*	*0.833*	*0.889*
	Support Vector Machine	*0.619*	*0.81*	*0.873*

TABLE 4.2 SEIZURE DETECTION ON BONN DATASET USING 60-40 SPLIT ML CLASSIFIERS

Metric	**Classifier**	**Precision**	**F1-score**	**Support**	**Accuracy**
Baseline	Decision Tree	*0.96*	*0.98*	*48.0*	*0.972*
	K-Nearest Neighbors	*0.889*	*0.941*	*48.0*	*0.917*
	Logistic Regression	*0.98*	*0.99*	*48.0*	*0.986*
	Naive Bayes	*1.0*	*0.989*	*48.0*	*0.986*
	Random Forest	*0.96*	*0.98*	*48.0*	*0.972*
	Support Vector Machine	*0.96*	*0.98*	*48.0*	*0.972*
Seizure	Decision Tree	*1.0*	*0.957*	*24.0*	*0.972*
	K-Nearest Neighbors	*1.0*	*0.857*	*24.0*	*0.917*
	Logistic Regression	*1.0*	*0.979*	*24.0*	*0.986*
	Naive Bayes	*0.96*	*0.98*	*24.0*	*0.986*
	Random Forest	*1.0*	*0.957*	*24.0*	*0.972*
	Support Vector Machine	*1.0*	*0.957*	*24.0*	*0.972*

Metric	**Classifier**	**Recall**	**Macro Avg**	**Weighted Avg**
Baseline	Decision Tree	*1.0*	*0.958*	*0.972*
	K-Nearest Neighbors	*1.0*	*0.875*	*0.917*
	Logistic Regression	*1.0*	*0.979*	*0.986*
	Naive Bayes	*0.979*	*0.99*	*0.986*
	Random Forest	*1.0*	*0.958*	*0.972*
	Support Vector Machine	*1.0*	*0.958*	*0.972*
Seizure	Decision Tree	*0.917*	*0.968*	*0.972*
	K-Nearest Neighbors	*0.75*	*0.899*	*0.913*
	Logistic Regression	*0.958*	*0.984*	*0.986*
	Naive Bayes	*1.0*	*0.985*	*0.986*
	Random Forest	*0.917*	*0.968*	*0.972*
	Support Vector Machine	*0.917*	*0.968*	*0.972*

With a 99.47% accuracy rate, the suggested model outperforms KNN by 7.7%, SVM by 5%, and DT by 2.27% as compared to the machine learning algorithms applied in paper [19]. This drives to the fact that suggested 1D-CNN LSTM ensemble model has great potential in the field of epileptic seizure detection research through EEG signals, as demonstrated by all these results.

TABLE 4.3 PERFORMANCE METRIC OF DEEP LEARNING MODELS

Model	Accuracy	Precesion	Recall	F1-Score
CNN	*97.13%*	*94.24%*	*92.34%*	*0.9328*
DNN	*96.35%*	*95.18%*	*87.50%*	*0.9118*
Suggested 1D-CNN LSTM	*99.47%*	*99%*	*99%*	*0.9959*

V. CONCLUSION

The increasing prevalence of epilepsy underscores the growing importance of accurate detection. A significant challenge lies in effectively identifying seizures from extensive datasets. Given the intricate nature of EEG signals within such datasets, ML classifiers prove to be a fitting solution for precise seizure detection. However, the critical aspects are the judicious choice of classifiers and features as shown in the results of table 4.1 and 4.2.

This report has conducted a comprehensive examination of machine learning methodologies for seizure detection. Consequently, it is concluded that "non-black-box" classifiers, specifically the decision forest, exhibit superior effectiveness. This choice is motivated by their ability to generate several logical and informative rules while maintaining a higher prediction accuracy. Moreover, decision forests facilitate the exploration of valuable insights, including seizure localization and the investigation of various seizure types.

On the other hand, despite their high predicted accuracy, "black-box" classifiers are unable to provide unambiguous rules. Regarding feature selection, it is recommended to opt for features that yield logical outcomes. Effective knowledge discovery may not be supported by reducing the dataset's dimensionality by using only one or two characteristics, such as line length and energy.

In essence, this report offers fresh insights for data scientists engaged in the domain of epileptic seizure detection through EEG signals. To sum up, this report centres on the assessment of machine learning classifiers and the selection of appropriate features as key factors in enhancing seizure detection methodologies.

A 1D-CNN LSTM ensemble epilepsy seizure detection model is proposed in this study using EEG signal as input. The proposed ensemble model will build an entire network i.e. by combining a LSTM with 1D-CNN, it will be able to distinguish precisely between the ordinary and epileptic seizures EEG data. The LSTM model is successful in identifying and interpreting the individual EEG signals, whereas the 1D-CNN picks out features from EEG data very well. Experiments on one of the popular dataset i.e.

UCI epileptic seizure data set validate the effectiveness of the suggested approach. Furthermore, when compared to other approaches such as DNN, CNN, KNN, SVM, and DT, the suggested model improves accuracy by 3.12%, 2.34%, 7.7%, 5.0%, and 2.27%, respectively. The suggested model has made significant strides toward recognising epileptic seizures but there are still some issues that need to be resolved in the future. The suggested model requires a significant quantity of labelled EEG signal data from a reliable source for its supervised training.

TABLE 5.1 DIFFERENCE BETWEEN DEEP LEARNING AND MACHINE LEARNING MODEL

Model	Accuracy	Precesion
1D-CNN LSTM	*99.47%*	*99%*
Decision Tree	*97.2%*	*96%*
Difference	*2.05%*	*3%*

Table 5.1 shows dominance of suggested deep learning algorithm over the best performing Machine learning algorithm though the suggested model used an extensive hardware and overloaded it, but it also provides a significant rise in the results. As we know that in the real world, it is difficult to get such filtered and clean epileptic signals. So, there is a significant chance of a dip in the accuracy of the model. Thus, we want to achieve as high as possible accuracy in theory so that any robber in the real-time data should cause the least deviation possible from the theoretical accuracy. This signifies the importance in difference of 2.05% accuracy and 3% precision. On theory, these minor differences may not justify the over-utilization of the hardware resources, but in practicality, these can prove as the game changers of our model. As we are dealing with the human health here, so even a 0.1% accuracy is a great step for saving the human lives.

VI. FUTURE PROSPECTS

However, gathering EEG data is a tedious work because it requires sensitive information of patients. The next study will be concentrating on two areas in light of these limitations: first, the transfer learning technique that could have been incorporated into the suggested model to lessen its reliance on labelled signal data; second, the suggested model can be improved more and adjusted further to perform better on increasingly difficult epileptic seizure recognition tasks, which will enhance its capacity to classify data from a variety of sources.

In contemporary research, the adoption of graph-theory methodologies has ushered in novel perspectives in the realm of epilepsy detection through EEG signals, leveraging distinct graph parameters. These graph-theory-based approaches offer valuable insights into the latent dynamics of brain activity and the mapping of brain behaviours. They facilitate a comprehensive understanding of EEG signal dynamics across various scales—microscopic, mesoscopic, and macroscopic—while also establishing meaningful correlations among them. Graph theory serves as a crucial tool in pinpointing anomalies within EEG patterns and extracting significant information regarding the underlying brain connectome through specific topological attributes of the EEG signal network. Statistical features derived from constructing networks from EEG signals furnish indispensable insights into dysfunctions associated with the structural and functional aspects of the brain in epilepsy research.

REFERENCES

[1] Abeg Kumar Jaiswal and Haider Banka. Epileptic seizure detection in eeg signal using machine learning techniques. Australasian physical & engineering sciences in medicine, 41:81–94, 2018.

[2] WHO, "World health organization 2024." https://www.who.int/ news-room/fact-sheets/detail/epilepsy. Accessed: Feb 12, 2024.

[3] P.Hospitals,"Internationalepilepsyday12feb,2024:Themeandimportance." https://www.pacehospital.com/international- epilepsy- day. Ac- cessed: Feb 12, 2024.

[4] H. Bhardwaj, P. Tomar, A. Sakalle, and U. Sharma, "Chapter 20 - principles and foundations of artificial intelligence and internet of things technology," in Artificial Intelligence to Solve Pervasive Internet of Things Issues (G. Kaur, P. Tomar, and M. Tanque, eds.), pp. 377–392, Academic Press, 2021.

[5] A. Sakalle, P. Tomar, H. Bhardwaj, M. A. Alim, et al., "A modified lstm framework for analyzing covid-19 effect on emotion and mental health during pandemic using the eeg signals," Journal of Healthcare Engineering, vol. 2022, 2022.

[6] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," Journal of the American Medical Informatics Association, vol. 24, no. 2, pp. 361–370, 2017.

[7] J. Hanson, Y. Yang, K. Paliwal, and Y. Zhou, "Improving protein disor- der prediction by deep bidirectional long short-term memory recurrent neural networks," Bioinformatics, vol. 33, no. 5, pp. 685–692, 2017.

[8] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 4580–4584, Ieee, 2015.

[9] Y.Sun,F.P.-W.Lo,andB.Lo,"Eeg-baseduseridentificationsystemus- ing 1d-convolutional long short-term memory neural networks," Expert Systems with Applications, vol. 125, pp. 259–267, 2019.

[10] Q. Wu and E. Fokoue, "Epileptic Seizure Recognition." UCI Machine Learning Repository, 2017. DOI: https://doi.org/10.24432/C5G308.

[11] Bardia Abbasi and Daniel M Goldenholz. Machine learning applications in epilepsy. Epilepsia, 60(10):2037–2047, 2019.

[12] Ushtar Amin and Selim R Benbadis. The role of eeg in the erroneous diagnosis of epilepsy. Journal of Clinical Neurophysiology, 36(4):294–297, 2019.

[13] Hai Chen and Mohamad Z Koubeissi. Electroencephalography in epilepsy evaluation. Continuum: lifelong learning in neurology, 25(2):431–453, 2019.

[14] Salim Lahmiri. An accurate system to distinguish between normal and abnormal electroencephalogram records with epileptic seizure free intervals. Biomedical Signal Processing and Control, 40:312–317, 2018.

[15] Boulenouar Mesraoua, Dirk Deleu, Hassan Al Hail, Gayane Melikyan, Paul Boon, Hiba A Haider, and Ali A Asadi-Pooya. Electroencephalography in epilepsy: look for what could be beyond the visual inspection. Neurological Sciences, 40:2287–2291, 2019.

[16] L. Abraira, L. L. Gramegna, M. Quintana, E. Santamarina, J. Salas-Puig, S. Sarria, A. Rovira, and M. Toledo, "Cerebrovascular disease burden in late-onset non-lesional focal epilepsy," Seizure, vol. 66, pp. 31–35, 2019.

[17] U. R. Acharya, S. Vinitha Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated eeg analysis of epilepsy: A review," Knowledge-Based Systems, vol. 45, pp. 147–165, 2013.

[18] M. S. Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, "Applying deep learning for epilepsy seizure detection and brain mapping visualization," ACM Trans. Multimedia Comput. Commun. Appl., vol. 15, feb 2019.

[19] A. Aghaei-Lasboo and R. S. Fisher, "Methods for measuring seizure frequency and severity," Neurologic Clinics, vol. 34, no. 2, pp. 383– 394, 2016. Optimal Care for Patients with Epilepsy: Practical Aspects.

[20] P. I. Frazier, "A tutorial on bayesian optimization," 2018.

[21] T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, "Fast hyperparameter tuning using bayesian optimization with directional derivatives,"Knowledge-Based Systems, vol. 205, p. 106247, 2020.

[22] Supriya, S., Siuly, S., Wang, H. et al. Automated epilepsy detection techniques from electroencephalogram signals: a review study.Health Inf Sci Syst 8, 33 (2020). https://doi.org/10.1007/s13755-020-00129-1

[23] Qu, H, and J Gotman. “Improvement in seizure detection performance by automatic adaptation to the EEG of each patient.” Electroencephalography and clinical neurophysiology vol. 86,2 (1993): 79-87. doi:10.1016/0013-4694(93)90079-b

[24] P. Jahankhani, V. Kodogiannis and K. Revett, "EEG Signal Classification Using Wavelet Feature Extraction and Neural Networks," IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA’06), Sofia, Bulgaria, 2006, pp. 120-124, doi: 10.1109/JVA.2006.17.

[25] Polat K, Güneş S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Appl Math Comput. 2007;187:1017–26.

[26] Belhadj S, Attia A, Adnane AB, Ahmed-Foitih Z, Taleb AA. Whole-brain epileptic seizure detection using unsupervised classification. In: Model- ling, Identification and Control (ICMIC), 2016 8th International Confer- ence on (pp. 977–982). IEEE. 2016.

[27] Zarei R, He J, Siuly S, Huang G, Zhang Y. Exploring Douglas-Peucker algorithm in the detection of epileptic seizure from multicategory EEG signals. Biomed Res Int. 2019; 2019:1–19.

[28] Ralph Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian Elger. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Physical review. E, Statistical, nonlinear, and soft matter physics, 64:061907, 01 2002.

[29] TMSi. “The 10-20 System for EEG.” TMSi, 21 Feb. 2023, info.tmsi.com/blog/the-10-20-system-for-eeg. Accessed 01 Feb. 2024.

[30] S. Yan, "Understanding LSTM and its diagrams." https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714, March 2016. Accessed: Feb 16, 2024.

[31] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, “Collaborative fuzzy clustering from multiple weighted views,” IEEE Transactions on Cybernetics, vol. 45, no. 4, pp. 688–701, 2015.

APPENDIX A

Minor Project Report

On

An Analysis of Epilepsy Detection and Classification using Machine Learning Techniques

Submitted to Amity University, Uttar Pradesh

in partial fulfilment of the requirement for the award of the degree of

Bachelor of Technology in

Computer Science and Engineering (2020-24) By

Mantra Jain

(A2305220412)

Ansh Srivastav

(A2305220390)

Under the guidance of

Dr Harshit Bhardwaj

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY
AMITY UNIVERSITY NOIDA UTTAR PRADESH
July-October 2023

DECLARATION

We, Mantra Jain and Ansh Srivastav, students of B.Tech (7-CSE-8Y) hereby declare that the project titled "An Analysis of Epilepsy Detection and Classification using Machine Learning Techniques" which is submitted by me to Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

The Author attests that permission has been obtained for the use of any copyrighted material appearing in the Dissertation / Project report other than brief excerpts requiring only proper acknowledgement in scholarly writing and all such use is acknowledged.

Date:
Mantra Jain
A2305220412
7-CSE-8Y (2020-24)

CERTIFICATE

On the basis of declaration submitted by **Mantra Jain (A2305220412) and Ansh Srivastv (A2305220390)**, student of B.Tech. CSE, I hereby certify that the in-house project titled "**An Analysis of Epilepsy Detection and Classification using Machine Learning Techniques**", submitted to Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology in Computer Science & Engineering, is an original contribution with existing knowledge and faithful record of work carried out by him under my guidance and supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: NOIDA
Date:

Dr. Harshit Bhardwaj
Assistant Professor-II
(Guide)
Department of Computer Science and Engineering
Amity School of Engineering and Technology
Amity University Uttar Pradesh, Noida

An analysis of Epilepsy detection and classification using machine learning techniques

Abstract

Epilepsy, a mental disorder characterized by seizures and uncertainty, remains a significant medical problem. Timely and accurate detection of epilepsy is very important for diagnosis, treatment and patient management. Considering that seizures can occur suddenly and without warning, it is important to have a system that can detect seizures. A comprehensive review of the electroencephalogram (EEG) recording is required to accurately identify these seizures. In recent years, the intersection of machine learning and medicine has shown promise in improving the diagnosis and classification of epilepsy. This summary provides a brief overview of the report on epilepsy diagnosis and classification analysis, which includes various machine learning algorithms such as K-Nearest Neighbor (KNN), Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine (SVM) and Decision Trees. This study provides a brief summary of a report on epilepsy detection and classification through machine learning. This report explores the evolving field of epilepsy diagnosis and reviews the various machine learning algorithms, datasets, and computational techniques currently in use. The overall aim of this report is to demonstrate the potential of machine learning to improve our understanding and management of epilepsy.

Keywords: electroencephalogram, Epilepsy, Seizures

1. Introduction

We discovered at the outset of this research that it would be quite helpful to clarify a few things. To provide them a brief overview of what they will read about in the next chapters, as well as the nature of the examination's subject and the solution's structure. We will concentrate on identifying epileptic seizures in electroencephalogram (EEG) data. All users are welcome to utilize this information, which was gathered at the German university of Bonn. Several well-known machine learning techniques that have been suggested in the literature for comparable tasks will be used in the identification procedure. To evaluate them, seven di↵erent measures will be applied. Python 3.7 was used to implement the whole procedure. The objective is to contrast some of the approaches put out in the literature and expand them from patient-specific to datasets with numerous cases. Epilepsy is a mental disorder characterized by sudden and unpredictable events that a↵ects millions of people worldwide. These seizures are caused by electrical malfunctions in the brain and often present with symptoms that vary in intensity and duration. Accurate and timely diagnosis of epilepsy is important for e↵ective diagnosis, treatment and management. In recent years, the integration of machine learning techniques into medicine has opened new avenues in the analysis, diagnosis and classification of epilepsy. This integration provides a revolutionary way to increase diagnostic accuracy, predict epilepsy events, and tailor personalized treatment

Email addresses: mj.mantrajain@gmail.com (Mantra Jain), srivastavansh253@gmail.com (Ansh Srivastav)

strategies. This comprehensive document covers various areas of machine learning for search and classification and highlights the important role of artificial intelligence (AI) in improving our understanding of complex medical conditions. Although epilepsy diagnosis has historically relied on neurologists' clinical observations, physical examination, and electroencephalography (EEG) data analysis, machine learning algorithms such as nearest neighbor (KNN) and logistic regression promise to provide better and more e↵ective diagnosis. path. - E cient and scalable solution. Machine learning and various algorithms such as KNN and Logistic Regression have the potential to change the entire search and classification landscape. These algorithms use information from large, complex datasets from pattern recognition and data analysis to help us di↵erentiate epilepsy, predict seizure incidence, and develop personalized treatment plans.

In this report we will examine the current state of the art in detecting and classifying epilepsy, paying attention to the di↵erent types of machine learning algorithms used and the data and computational methods used. By examining the strengths and limitations of these algorithms, we aim to better understand their e↵ectiveness in diagnosing epilepsy and predicting seizures.

As we delve deeper into the integration of medicine and technology, we aim to see the potential of machine learning, including algorithms such as KNN and logistic regression, in search and classification. From the content of this report, readers will gain a deeper understanding of the current state of the field, future challenges, and deadlines for machine learning to improve epilepsy management and improve patient care.

2. Related Work

The extension of machine learning in epilepsy-focused sectors, including seizure detection and monitoring, has been the subject of numerous studies, including this paper. By utilizing methods such as multilayer artificial neural networks, support vector machine (SVM), and deep learning, machine learning shows potential in enhancing the ability to handle and evaluate EEG and imaging data that was once considered too complex for experts. Furthermore, this paper [1] supports applying machine learning techniques to optimize medication selection, improve the precision of clinical outcome predictions, and streamline surgical planning. Predictive models produced by machine learning are a source of concern for the authors due to the limited number of validation studies published. It's worth considering the applicability and generalizability of these models in light of this deficiency. Broader datasets that take into account greater diversity are recommended by the authors in order to fill this void. Furthermore, the expected increase in investment in external validation studies to make the application of machine learning in medicine, particularly in epilepsy, more reliable was highlighted.

U. Rajendra Acharya et al. explains that seizures that occur frequently are the hallmark of epilepsy, an electrophysiological brain condition. Epileptic seizures are often detected and studied using an electroencephalogram (EEG) [2] , a technique that tracks and analyzes electrical responses in the brain. Nevertheless, it is frequently challenging to spot minute but significant shifts in the EEG pattern by visual examination, opening up a wide study area for biomedical researchers to create and apply a number of clever algorithms enabling the detection of such modest alterations. The EEG signals are also irregular and unpredictable in character, which adds to the di culty of manually identifying both normal and aberrant (interictal and ictal) activity. Therefore, a Computer Aided Diagnostic (CAD) method must be created in order to e↵ortlessly di↵erentiate between healthy and sick behaviors utilizing a minimal amount of highly distinguishing classifiers. It has been discovered that nonlinear characteristics can capture complicated physiological processes in the EEG signals, such as sudden shifts and unpredictable activity.

Akut, Rohan is talking about the avoidance of overfitting in modeling with a multiple validation of X folds to train each model. They did 150 epoch training loops and a batch size of 5 and used Batch Normalization to try and mitigate the e↵ect of any covariance variations happening in the model [3]. This modification includes another normalization e↵ect that allows for faster model training with less overfitting probability risk. Particular emphasis is on the Max Pooling Layer using a 1x2 filter which results in output size half of a input size. Its last layer of computation is passed through a SoftMax function in order to classify the data. A dropout layer value of 0.25 was used for more fine-tuning and overfitting prevention. Binary classified results were then compared with pre-existing models based on accuracy, sensitivity, and specificity, which suggests an overarching aim of assessing how well the model performs in small datasets. This procedure can help to get the detailed information about the capability of the model. Particularly, we assessed the validity of the model with respect to a two-level classification utilizing the Bonn dataset. The Bonn dataset is considered a standard benchmark for EEG data, as well as one of the benchmark datasets to evaluate classifier's performance on, which enables us to test our CNN based on a highly reliable source. The CNN model is trained using cross-validation of 10-folds and the optimal batch sizes/epochs that can generate best performance.

Amin et al. points out that reversing the epilepsy diagnosis is complex and requires examining the ”unusual” EEG, which can be tricky. One of the primary causes for regular EEGs being mistakenly interpreted as disturbed is the absence of required practical experience in neurology residency programs. Tests, such as the EEG, shouldn't take precedence over medical expertise [4]. Some kinds of seizures may go undiagnosed, and identification of epilepsy may prove di cult. Hypermotor seizures in the frontal lobe may be mistaken for psychogenic episodes. Focal oblivious cognitive seizures in the elderly may be misdiagnosed as dementia, and frontal and temporal lobe epilepsies may present with ictal or interictal psychosis that is misdiagnosed as a main mental disease. Diagnostic mistakes happen often in medicine and a↵ect all disciplines. Both patients and doctors may have severe repercussions. In neurology, mistakes frequently result from a focus on ”assessments over the actual clinical situation. Epilepsy is often diagnosed clinically and based on a patient's medical history. Overdiagnosis of epilepsy occurs more frequently than underdiagnosis. Poor medical background and an unusual EEG can result in an incorrect epilepsy diagnosis. Patients who had been diagnosed previously with epilepsy, failed to get better after receiving their first antiepileptic medication as they don't actually have the condition. Most people who receive an incorrect epilepsy diagnosis end up having syncope or psychogenic nonepileptic events.

Amin, Ushtar and Benbadis, Selim R came to the conclusion that Regular EEG has a sensitivity of minimum 80 percent for epilepsy. Except for the neglected issue of over-reading, the specificity for interictal epileptiform emition is strong i.e. grater than 90 percent [5]. When they occur, typically aid in identifying the epilepsy disorder type. One of the independent recurrent indicators is the existence of epileptiform EEG discharges, although this information must be evaluated in a clinical setting. This is a reason to record an EEG since it helps with risk assessment before deciding whether to discontinue ASM (Anti Seizure

Medications) from patients who are seizure-free. For various applications, there are many EEG recording methods. The video-EEG recording of the alleged occurrences is the most conclusive examination to identify the seizure type. When the cause of changed mental state and atypical behaviors is unclear, an EEG in the intensive care unit is a helpful tool for investigating the root cause. They explained how EEG continues to be a crucial test for the diagnosis of epilepsy despite developments in imaging. It can clarify the kind of epilepsy as well as confirm the diagnosis. Depending on the length, the availability of video, and the inpatient or outpatient surgery scenario, there are many distinct forms of EEG recordings, each with advantages and disadvantages. Although interictal epileptiform aberrations are particularly unique to epilepsy, unskilled readers may overinterpret them. EEG has a role in the diagnosis of epilepsy, the decision to stop therapy in seizure-free individuals, and the evaluation of critically sick patients for potential encephalopathies and status epilepticus. EEG results must be somewhat uniform and understandable to the doctor who ordered the EEG.

Bajaj et al. have introduced to an innovative personalisable approach to detecting epileptic seizures with the aid of an individual. It is based on the analysis of EEG signal's analytical intrinsic mode functions (IMF). It [7] is directed towards classifying EEG recordings corresponding to seizure free as well as seizure events, which can be informative in terms of diagnostics and therapy. The method we propose uses the extraction of an instantaneous area using the curve of the analytical IMFs of EEG calculated with a movable window of the minimal width. According to the study, this method demonstrated promising results in detecting focal temporal lobe epilepsy from intracranial EEG signals using rule-based technique. In addition, the authors contrasted their approach against existing detection techniques examined using the same EEG dataset. As demonstrated the proposed approach achieved better detection performance which pave the way for future research in epilepsy diagnosis. Statistical metrics like sensitivity (SEN), specificity (SPE), Positive predictive value (PPV), Negative Predictive Value (NPV) and Error rate detection (ERD) were used to evaluate the performance of the proposed system. The indices were calculated according to correctly and incorrectly detect positive and negative events in total. The performance was good for seizure detection and suggests the feasibility of this method to be deployed in clinics for practical use. Therefore, the proposed work is an additional contribution in the personalized epilepsy detection literature. This highlights the possibility of utilizing instantaneous area of IMFs obtained from EEG signals and opens up an avenue towards better understanding and manipulation of epileptic seizures.

Chakraborti et al. analyses artificial neural networks (ANNs), a popular machine learning technique used to discover knowledge and patterns from exponentially growing data sets. This discussion [8] centers on two specific types of neural networks: back-propagation (perceptron) and back-propagation networks. The learning rule of the former is used for updating the weights and the biases of the network. This network has an output that's controlled by the interaction between a transfer function, the weight, the bias and the input. Then there's the multilayer feed-forward back-propagation network which generalizes the least mean square algorithm. Backward propagation is done for the sensitive parts of the model. They use the back-propagation network to detect the epileptic and non-epileptic signals (Figure 2). Functionality and application aspects of both networks are discussed in detail followed by a comparative study highlighting the broader aspects and benefits of deploying artificial neural networks in the context of machine learning applications.

Chen et al. reviewed how electroencephalogram (EEG) is linked to Epileptic seizures and provided physiologic basis of EEG and intracranial EEG studies. They talked about pointed contoured waveforms or complexes that are di↵erent from background waves and mimic those observed in a part of human people with epileptic diseases are referred to as interictal epileptiform discharges. The most widely characterized interictal epileptiform discharges are spikes and abrupt waves [9]. They explained about rhythmic discharge that typically has to last at least 10 seconds to qualify as an electrographic seizure. BIRDs are described as ”Concisely, this refers to short bursts of rhythmic brain activity exceeding 4 Hz, which may appear abruptly and do not match any recognized normal or harmless patterns” Their research frequently identifies interictal or ictal abnormalities, and how EEG is still an essential tool for diagnosis of epilepsy. Yet, epilepsy cannot be ruled out in the lack of interictal epileptiform discharges or ictal symptoms. There are two types of seizures: namely, focal or generalized. Electrographic patterns may di↵er, and ictal activity often develops throughout a seizure. In order to properly diagnose and treat nonconvulsive status epilepticus i.e is a state of continuous seizure activity for at least 30 minutes, with cognitive or behavioral changes, continuous EEG monitoring is crucial. When scalp EEG results are ambiguous, intracranial EEG monitoring is extremely helpful for planning surgery since it frequently provides for earlier seizure identification and higher spatial resolution than scalp recordings.

Guo et al. indicated that Epilepsy a↵ects around 1 percent of the worldwide population. Recurring seizures are epilepsy’s primary symptom. The processes behind epileptic diseases can be better understood by carefully analyzing electroencephalogram (EEG) samples [10]. Automatic seizure identification in EEG recordings is crucial because epileptic seizures happen erratically and without warning. An e cient analytic method for signals that are not stationary, such as EEGs, is the wavelet transform (WT). The line length characteristic is a measurement that is responsive to variations in frequency as well as amplitude and depicts shifts in waveform dimensionality. The automated epileptic seizure detection approach described in this study employs line length characteristics based on wavelet transform multiresolution decomposition in conjunction with an artificial neural network (ANN) to categorize the EEG signals about seizure presence or absence. The authors are aware of no other works in the literature that are comparable to this one. The suggested approach was assessed using an established public dataset. The excellent accuracy achieved for three separate categorization issues attested to the method’s outstanding performance.

Jaiswal et al. presents detailed comparison of SpPCA (Sparse Principle Component Analysis) against SubXPCA (Sub space Extend Princpale Component Analysis) with SVM (Support vector machine) for classification purpose. It turns out that adding more projection vectors is positively correlated with how well they can predict the given task [11]. In terms of classification accuracy, the paper shows that the SubXPCA method yielded very slightly better results in most cases with respect to SpPCA. A great finding was: As the number of PVs grew larger in SpPCA, its accuracy didn’t raise dramatically. On the other hand, SubXPCA had an increasing number of accurate features. In the paper we also included figures showing an out-of-sample performance comparison among PCA + SVM, SpPCA + SVM, and SubXPCA + SVM; these confirmed the best outcomes of subspace XPCA + SVM. The authors end with putting the results into perspective and compare them to what had been found in prior literature. Thusly this document is an incredible expansion to the computational documentation in class calculations issues, especially for Principle Components Analysis, Spare Principle Corresponding Matrix Evaluation and Subspace Stretched out Principle Corresponding

Matrix Assessment. It demonstrates how these algorithms perform di↵erently with varying conditions which is quite an excellent base for future research on these algorithms.

Lahmiri, Salim indicated how epilepsy is becoming more common, and its prevalence is rising. Designing precise computerized procedures for the identification and categorization of electroencephalogram (EEG) data from epileptic patients is therefore very helpful in the diagnostic process. There work aims to propose a machine-learning diagnosis method that can quickly and accurately identify between normal and abnormal EEG data with seizure-free periods using the extended Hurst exponent approaches, fractal features of EEG signals are computed at various scales to better describe their dynamics [12]. Generic Hurst exponent estimations between healthy and epileptic EEG signals with seizures uninterrupted durations are statistically di↵erent, according to parametric and nonparametric statistical tests. Support vector machine classifiers that were trained using extended Hurst exponent estimates. There suggested system has potential and can be expanded for other biomedical applications such as di↵erentiating between normal brain waves and those with intervals of seizure or between epileptic EEG signals with seizure free intervals because this problem is challenging and has not been addressed in the literature.

Lahmiri et al. told that epilepsy is becoming more common, and its prevalence is rising. They designed precise computerized procedures [13] for the identification and categorization of electroencephalogram (EEG) data from epileptic patients, therefore they showed its importance in the diagnostic process. There work proposed an automated diagnosis method that can quickly and accurately identify between normal and abnormal EEG data with seizure-free periods. The system is built on extended pointed exponent estimations at various scales that are used to describe EEG recordings, and it is then based on a support vector machine classifier that will be used for categorization and have various kernels. At last, cross-validating studies using the ten-fold classification was suggested.

Mesraoua et al. indicated how EEG in comparison to the conventional method of eye assessment alone, scalp electroencephalography has the potential to provide additional spatial and temporal information. Fortunately, this information is easier to acquire because to contemporary digital EEG technology and computer-assisted analysis. A potential method to enhance non-invasive EEG localisation in focal epilepsies is to look at the spike voltage topography of interictal spikes [14]. Another additional method for locating the epileptogenic zone in individuals who are candidates for epilepsy surgery is electrical source imaging. The vast amount of data in continuous EEG is streamlined by quantitative EEG by providing a static graphical depiction. These are just a few instances of the wealth of information that scalp EEG recordings may give that goes beyond simple eye assessment.

In dissecting epileptic seizures, this paper "A review on the pattern detection methods for epilepsy seizure detection from EEG signals" o↵ers an outline on the use of Discrete Wavelet Transform (DWT). The early portion of the article regards the struggles involved with raw EEG data due to poor spatial resolution and a low signal-to-noise ratio. It is therefore crucial to initiate pre-processing as a means of enhancing the resolution and ratio of the data. The wavelet transform method is presented as a useful technique based on multi-resolution analysis that addresses both of these problematic elements adeptly [15]. Signals from epileptic seizures can be analyzed more e↵ectively using a technique that captures both low-frequency and high-frequency time information. This technique involves decomposing the signal into various levels using a filter bank, as the authors explain in their discussion on the functionality of the method. The paper delves into the importance of statistical time-domain features extracted from DWT sub-bands, specifically focusing on the mean absolute value (MAV) and standard deviation. To provide further insight on the effectiveness of DWT in EEG pre-processing, the paper references an existing study from the University of Bonn, Germany, which utilizes DWT to identify epileptic seizures. Overall, the discussion in the paper is well-reasoned and emphasizes the potential benefits of DWT in enhancing signal-to-noise ratios and spatial resolution for analyzing epileptic seizures in EEG data.

Smith SJM says Hans Berger, a German psychiatrist, made the discovery of a person electroencephalogram (EEG) in 1929 [17]. When Gibbs and colleagues in Boston showed 3 per second spike wave discharge in what was then called petit mal epilepsy, its potential uses in epilepsy quickly became evident. Because it is an easy and reasonably priced compared to later treatments, way to show the physiological manifestations of abnormal cortical excitability that underlie epilepsy, EEG persists on playing a crucial role in the detection and treatment of patients with epileptic seizures. EEG assists patients with epilepsy in identifying the kind of seizure and epilepsy syndrome, which helps with antiepileptic drug selection and prognostic prediction. In order to determine is the seizure condition focal or generalized, idiopathic or symptomatic, or a component of a particular epilepsy syndrome, EEG results are important to the multi-axial diagnosis of epilepsy. The logical distinction between partial and generalized seizures/epilepsy types is relevant and helpful in the therapeutic setting. Based on the patient and witness accounts, the doctor will often be able to determine the kind of seizure. However, EEG can assist distinguish between a complex partial seizure with focal IED (Interictal Epileptiform Discharges) and an absence type seizure with generalized IED when the history is uncertain (unobserved "blackouts" or transient impairment of awareness).

The research article "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis" [18] describes a method of data analysis that is used to categorize segments, this tool is therefore highly effective. It primarily utilizes the time domain, frequency domain, time-frequency domain, and nonlinear analysis. The outcomes are considered via the perspective of different classifiers, including K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machines (SVM). Each of these models is evaluated via two different techniques: 5-fold CV and 10-fold CV. In the time domain, the majority of classifiers have a high degree of accuracy, with KNN and SVM having the greatest success. In the domain of frequency, the accuracy is slightly lower, with the LR and SVM models providing the greatest performance. Within the domain of time-frequency, KNN and SVM have a high degree of success, this demonstrates the effectiveness of these classifiers. Overall, the effort suggests that these methods of classification, combined with a wavelet denoising method, are capable of producing accurate and legitimate results in various disciplines.

3. Methodology

3.1. Freely acessible datasets

The utilization of datasets is crucial for data scientists and academics to evaluate the success of the models they have presented. The detection of a tumor should similarly pick up on our brain signals. The most popular way to track brain activity is through EEG recordings. These recordings are crucial for machine learning classifications that investigate novel techniques for detecting tumors in a variety of ways, including early tumor detection, quick tumor detection, patient tumor detection, and tumor localization. Data sets that are accessible to the general public are crucial for analysis, comparison, and inference. We'll go through the well-known datasets frequently utilized in epilepsy in the part after that.

1. **Bonn University—EEG dataset** The BONN EEG Time Series Epilepsy Dataset constitutes an important tool for epilepsy research and neurology. The dataset [6] was developed at the University of Bonn in Germany towards enhancing computational analysis of epileptic seizure and improve its detection. Here are some more detailed aspects of the dataset: Data Source: Two major sources of the dataset are; EEG recordings.
    (a) Epileptic Patients: Epileptic EEG data from people. These recordings are very valuable forunderstanding epileptic seizures because they document the activity at the level of the brain during such events.
    (b) Healthy Individuals: Control: Data from EEG recordings of humans who do not have epilepticseizures.

    Annotations: Annotation has been applied in this dataset, indicating epileptic seizures and other events worthy of note. Such annotations are important for training and testing of automated seizure detection software in EEG data. Contributions: With the introduction of the BONN EEG Time Series Epilepsy Dataset, it is possible to develop computer-aided tools for epilepsy diagnosis and management. This allowed refining the algorithms that had the e↲ect of giving better results when working with other patients having this condition. This dataset consists of 100 single-channel EEG recordings, each lasting 23.6 seconds and sampled at a rate of 173.61 Hz. The spectral bandwidth of the data ranges from 0.5 Hz to 85 Hz, and it was originally obtained using a 128-channel acquisition system. These EEG recordings were extracted from larger multi-channel EEG recordings of five patients and designated as Sets A, B, C, D, and E.
    (a) Sets A and B represent surface EEG recordings during periods of closed and open eyes, respectively, in healthy patients.
    (b) Sets C and D comprise intracranial EEG recordings, with C obtained from a seizure-freezone within an epileptic patient's brain and D from a non-seizure-generating area of the same patient.
    (c) Set E contains intracranial EEG data from an epileptic patient captured during epilepticseizures.

    Each set contains 100 text files, each with 4097 samples representing a single EEG time series in ASCII code format. The data has undergone bandpass filtering with cuto↲ frequencies at 0.53 Hz and 40 Hz. It is noteworthy that this dataset is devoid of artifacts, and thus, no prior preprocessing steps are necessary for classifying healthy (non-epileptic) and unhealthy (epileptic) signals. Strong eye movement artifacts have been removed. This dataset was made publicly available in 2001 and has been extended as part of the EPILEPSIA project. Indeed, the dataset is very important because it provides an opportunity to conduct further investigations into epilepsy which translates into development of e↲ective computational tools meant

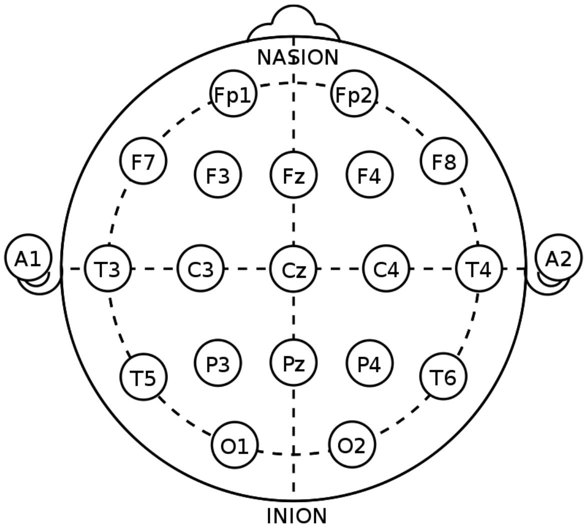


Figure 1: 10 - 20 Electrode System

	Set A	Set B	Set C	Set D	Set E
Subject	Vigorous	Vigorous	Epilepsy	Epilepsy	Epilepsy
Subject Conditions during Readings	Not asleep with eyes opened	Not asleep with eyes closed	Seizure-free (interictal)	Seizure-free (interictal)	Seizure-free (ictal)
Type of Electrode	Surface	Surface	Intracranial	Intracranial	Intracranial
Placement of Electrode	International 10 - 20 System	International 10 - 20 System	International 10 - 20 System	International 10 - 20 System	International 10 - 20 System
Channels	100	100	100	100	100
Duration (Secsond)	23.6	23.6	23.6	23.6	23.6

Table 1: Summary of BONN Dataset

2. **CHB-MIT Scalp EEG Database** The CHB-MIT Scalp EEG Database encompasses data extracted from EEG recordings of 22 young epileptic patients with unpreventable seizures. These [16] were followed by subjects undergoing withdrawal of antiseizure medicines with time spans ranging from two days to determine instance of seizures and operation suitable. 182 seizure startings and closures were recorded. The young epileptic patients were recorded on EEG at the Children's hospital in Boston. Following withdrawal of antiepileptic drugs, the patient had to be watched for several hours or even several days, in order for the type and frequency of epilepsy attacks could be identified and determine eligibility for surgical intervention. These 129 files which contain at least one episode of a seizure recorded in the book Records with Seizures form the subset of the recordings marked by (#) in the table on page two There are 198 seizures in these records as compared to the 182 listed among the first 23 instances in the seizure annotation files in RECORDS-WITH-SEIZURES. Files with names "chbnn-summary.txt" describe the montage used for each recording, as well as the duration in seconds between the commencement and the beginning of each seizure contained in di↲erent "edf(European Data Format)" files.

3.2. Classifiers Theory

1. Decision Tree

Using a decision tree technique, which is commonly used for classification jobs, epilepsy may be classified with EEG. Recursively partitioning the data according to distinct characteristics yields a tree-like structure that is used to identify the class labels of instances. By using the built-in structure as a decision-making tool, this algorithm operates. The decision tree for EEG epilepsy categorization and its mathematical justification are given below. The recursive process entails selecting the best characteristic to split the data at each node in order to construct a decision tree. The criteria measure may be used to reduce the disorder and impurity in the data, which is essential to achieving homogenous subsets. The recursive procedure continues for each subset until all samples in a node have the same class label, at which point it ends when a stopping requirement, such as the maximum depth or the minimum samples per leaf, is satisfied. The epileptic or non-epileptic status of fresh EEG data (X _new) may be determined by moving up the decision tree from the root node to a leaf node. This is the mathematical justification for the classification procedure that follows decision tree construction.

$$Entropy(S) = \sum_{i=1}^{N} p_i log_2 p_i$$

S- set of all instances in the dataset N -
number of distinct class values

$$p_i = eventprobability$$

2. K-Nearest Neighbors

KNN is k-nearest nearest algorithm otherwise known as a supervised learner with nonparametric characteristics. This involves determining an approximate class or value of a data point by comparing it to other data points. It is applicable in both regression and classification purposes but the general use of clustering similar points makes this tool mainly a classifier. "K" in KNN stands for the number of nearest neighbors, which is taken into account in case of a certain record classification. Choice of 'K' depends upon various parameters of the input data. Most of such data generally benefit from a higher 'k' value. For a classification technique, it's usually advisable to use one 'k' value for this purpose; besides, some cross-validation methods can help choose the best 'k' for a dataset.

$$Euclidean = \sqrt{\sum_{i=1}^{i=k} (x_i \quad y_i)^2}$$

3. Logistic Regression

For binary classification issues, logistic regression is a popular statistical and machine learning technique. Based on a variety of input variables, it estimates the likelihood that an output will fall into one of two classifications. Logistic regression uses the logistic function to limit its output to a range between 0 and 1, reflecting probabilities, in contrast to linear regression, which predicts continuous values. Logistic regression measures the e↲ect of each input feature on the likelihood of class membership by calculating coe cients for each feature. It can be understood, is computationally e↲ective, and acts as a base for more sophisticated methods. Numerous industries, including as healthcare, finance, and marketing, use logistic regression because decision-making and predictive modeling require an understanding of the likelihood of binary outcomes.

4. Naive Bayes

Na¨ıve Bayes is an easy-to-use probabilistic classification technique for simple applications. The latter relies on the Bayes theorem and provides the probability for a point to belong to a specific class. The naive assumption is that every attribute is independent and makes calculation easier, but this notion does not hold true in real cases. While it is somewhat oversimplified, still many techniques applied in the text classification of spam and opinion mining lose against naive Bayes. This particular algorithm is specifically ideal for high-dimensional datasets that have just enough labeled data. Naive Bayes is of considerable importance because of the capability of handling multiple classifications as well as the ease with which it can be trained and implemented for machine learning and many natural language processing applications.

$$P(C/X) = \frac{(P(X/C) \times P(C))}{P(X)}$$

P(C/X) = Posterior Probability

P(X/C) = Likelihood

P(C) = Class Prior Probability

P(X) = Predictor Prior Probability

5. Random Forest

It is an excellent machine learning's ensemble learning technique. This method will result in good and accurate prediction by incorporating several decision making trees in it. The second point is that trees are randomly train-ed one at a time on a specifically chosen portion of the data, based on randomly selected features reducing thus overfitting and improving generalizedness. In this case, the final forecast is created using projections of di↵erent trees. Random Forest is able to provide accurate and highly reliable predictions on many applications which include both classification and regression. One such tool exists for numerous branches of study such as image classification, finances, and healthcare and needs minor changes done in respects with hyperparameters unlike a solitary decision tree.

6. Support Vector Machine

SVM is a powerful supervised learning technique that can be used in classification and regression problems. Creating an ideal hyperplane that divides the data into two groups of two. SVM has advantage in high-dimensional domain; thus appropriate for tasks like text and image classification. The algorithm operates on such kernel functions as radial basis function (RBF) kernel for handling both linearly and non-linearly classified data. The biggest advantage of SVM is that it is very robust, particularly when working with small and unbalanced datasets; it can also cope with several non-linear relationships in data.

4. Experimental Results

This section details on the experimental results of the applied Machine Learning Algorithms. Table 2 shows training over 70 % of data and testing for 30%, Table 3shows training over 60 % of data and testing for 40%.

4.1. Performance Measures

Here's a brief explanation of the terms in a classification report:

1. Precision: Measures the accuracy of positive predictions.

$$Precision = \frac{TruePositives}{(TruePositives + FalsePositives)}$$

2. Recall: Measures the ability of the model to correctly identify all positive instances.

$$Recall = \frac{TruePositives}{(TruePositives + FalseNegatives)}$$

3. F1-score: A balanced metric that combines both precision and recall. It's the harmonic mean ofprecision and recall and is useful when you want to balance both false positives and false negatives.

$$F1 \quad score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4. Support: The number of samples in each class, which can help you understand the dataset's classdistribution.
5. Accuracy: It measures the overall correctness of predictions made by a model. It is calculated asthe ratio of correctly predicted instances to the total number of instances.

$$AccuracyScore = \frac{TruePositives + TrueNegatives}{TruePositives + FalseNegatives + TrueNegatives + FalsePositives}$$

6. Macro Average (Macro Avg): Macro average is a way to calculate an average of a metric (e.g., precision, recall, F1-score) across multiple classes in a multi-class classification problem. It calculates the metric independently for each class and then takes the unweighted average (arithmetic mean) of these class-wise metric values.

$$MacroAvg = \frac{1}{N}\sum_{i=1}^{N} Metric_i$$

- N is the total number of classes.
- Metric is the metric (e.g., precision, recall, F1-score) for class i

7. Weighted Average (Weighted Avg): Unlike macro average, weighted average takes into account theclass distribution. Classes with more instances have a greater influence on the weighted average than classes with fewer instances. It calculates the metric for each class, but the contribution of each class to the weighted average is proportional to the number of instances in that class.

$$WeightedAvg = \frac{1}{N}\sum_{1}^{N}(\frac{Metric_i \times Support_i}{TotalSupport})$$

- N is the total number of classes.
- Metric is the metric (e.g., precision, recall, F1-score) for class i.
- Support is the number of instances in class i.
- Total Support is the total number of instances in the dataset.

Metric	Classifier	precision	recall	f1-score	support
Baseline	Decision Tree	0.857	1.0	0.923	42.0
	K-Nearest Neighbors	0.737	1.0	0.848	42.0
	Logistic Regression	0.894	1.0	0.944	42.0
	Naive Bayes	0.933	1.0	0.966	42.0
	Random Forest	0.857	1.0	0.923	42.0
	Support Vector Machine	0.84	1.0	0.913	42.0
Seizure	Decision Tree	1.0	0.667	0.8	21.0
	K-Nearest Neighbors	1.0	0.286	0.444	21.0
	Logistic Regression	1.0	0.762	0.865	21.0
	Naive Bayes	1.0	0.857	0.923	21.0
	Random Forest	1.0	0.667	0.8	21.0
	Support Vector Machine	1.0	0.619	0.765	21.0
accuracy	Decision Tree			0.889	0.889
	K-Nearest Neighbors			0.762	0.762
	Logistic Regression			0.921	0.921
	Naive Bayes			0.952	0.952
	Random Forest			0.889	0.889
	Support Vector Machine			0.873	0.873
macro avg	Decision Tree	0.929	0.833	0.862	63.0
	K-Nearest Neighbors	0.868	0.643	0.646	63.0
	Logistic Regression	0.947	0.881	0.904	63.0
	Naive Bayes	0.967	0.929	0.944	63.0
	Random Forest	0.929	0.833	0.862	63.0
	Support Vector Machine	0.92	0.81	0.839	63.0
weighted avg	Decision Tree	0.905	0.889	0.882	63.0
	K-Nearest Neighbors	0.825	0.762	0.714	63.0
	Logistic Regression	0.929	0.921	0.918	63.0
	Naive Bayes	0.956	0.952	0.951	63.0
	Random Forest	0.905	0.889	0.882	63.0
	Support Vector Machine	0.893	0.873	0.864	63.0

Table 2: Seizure Detection on BONN dataset using 70-30 split ML Classifiers

Metric	Classifier	precision	recall	f1-score	support
Baseline	Decision Tree	0.96	1.0	0.98	48.0
	K-Nearest Neighbors	0.889	1.0	0.941	48.0
	Logistic Regression	0.98	1.0	0.99	48.0
	Naive Bayes	1.0	0.979	0.989	48.0
	Random Forest	0.96	1.0	0.98	48.0
	Support Vector Machine	0.96	1.0	0.98	48.0
Seizure	Decision Tree	1.0	0.917	0.957	24.0
	K-Nearest Neighbors	1.0	0.75	0.857	24.0
	Logistic Regression	1.0	0.958	0.979	24.0
	Naive Bayes	0.96	1.0	0.98	24.0
	Random Forest	1.0	0.917	0.957	24.0
	Support Vector Machine	1.0	0.917	0.957	24.0
accuracy	Decision Tree			0.972	0.972
	K-Nearest Neighbors			0.917	0.917
	Logistic Regression			0.986	0.986
	Naive Bayes			0.986	0.986
	Random Forest			0.972	0.972
	Support Vector Machine			0.972	0.972
macro avg	Decision Tree	0.98	0.958	0.968	72.0
	K-Nearest Neighbors	0.944	0.875	0.899	72.0
	Logistic Regression	0.99	0.979	0.984	72.0
	Naive Bayes	0.98	0.99	0.985	72.0
	Random Forest	0.98	0.958	0.968	72.0
	Support Vector Machine	0.98	0.958	0.968	72.0
weighted avg	Decision Tree	0.973	0.972	0.972	72.0
	K-Nearest Neighbors	0.926	0.917	0.913	72.0
	Logistic Regression	0.986	0.986	0.986	72.0
	Naive Bayes	0.987	0.986	0.986	72.0
	Random Forest	0.973	0.972	0.972	72.0
	Support Vector Machine	0.973	0.972	0.972	72.0

Table 3: Seizure Detection on BONN dataset using 60-40 split ML Classifiers

5. Conclusion

The increasing prevalence of epilepsy underscores the growing importance of accurate detection. A significant challenge lies in effectively identifying seizures from extensive datasets. Given the intricate nature of EEG signals within such datasets, machine learning classifiers prove to be a fitting solution for precise seizure detection. However, the critical aspects are the judicious choice of classifiers and features.

This research paper has conducted a comprehensive examination of machine learning methodologies for seizure detection. Consequently, it is concluded that “non-black-box” classifiers, specifically the decision forest (an ensemble of decision trees), exhibit superior effectiveness. This choice is motivated by their ability to generate multiple sensible and explanatory logic rules while maintaining a high prediction accuracy. Moreover, decision forests facilitate the exploration of valuable insights, including seizure localization and the investigation of various seizure types.

In contrast, “black-box” classifiers lack the capacity to generate explicit logic rules, although they can achieve notable predictive accuracy. Regarding feature selection, it is recommended to opt for features that yield logical outcomes. A review of existing literature reveals that features like entropy, line length, energy, skewness and standard deviation can attain a classification accuracy of 100%. It is advisable to avoid irrelevant features, especially as the dimensionality of the data increases, as this can lead to heightened computational costs and the emergence of unintelligible patterns. While employing only one or two features such as line length and energy may reduce the dataset’s dimensionality, it may not be conducive to effective knowledge discovery.

In essence, this paper offers fresh insights for data scientists engaged in the domain of epileptic seizure detection through EEG signals. To sum up, the paper centers on the assessment of machine learning classifiers and the selection of appropriate features as key factors in enhancing seizure detection methodologies.

References

[1] Abbasi, B., & Goldenholz, D. M. (2019). Machine learning applications in epilepsy. *Epilepsia*, *60*, 2037–2047.

[2] Acharya, U. R., Vinitha Sree, S., Swapna, G., Martis, R. J., & Suri, J. S. (2013). Automated eeg analysis of epilepsy: A review. *Knowledge-Based Systems*, *45*, 147–165. URL: https://www.sciencedirect.com/science/article/pii/S0950705113000798. doi:https://doi.org/10. 1016/j.knosys.2013.02.014.

[3] Akut, R. (2019). Wavelet based deep learning approach for epilepsy detection. *Health information science and systems*, *7*, 8.

[4] Amin, U., & Benbadis, S. R. (2019). The role of eeg in the erroneous diagnosis of epilepsy. *Journal of Clinical Neurophysiology*, *36*, 294–297.

[5] Amin, U., & Benbadis, S. R. (2019). The role of eeg in the erroneous diagnosis of epilepsy. *Journal of Clinical Neurophysiology*, *36*, 294–297.

[6] Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. (2002). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical review. E, Statistical, nonlinear, and soft matter physics*, *64*, 061907. doi:10.1103/PhysRevE.64.061907.

[7] Bajaj, V., & Pachori, R. B. (2013). Epileptic seizure detection based on the instantaneous area of analytic intrinsic mode functions of eeg signals. *Biomedical Engineering Letters*, *3*, 17–21.

[8] Chakraborti, S., Choudhary, A., Singh, A., Kumar, R., & Swetapadma, A. (2018). A machine learning based method to detect epilepsy. *International Journal of Information Technology*, *10*, 257–263.

[9] Chen, H., & Koubeissi, M. Z. (2019). Electroencephalography in epilepsy evaluation. *Continuum: lifelong learning in neurology*, *25*, 431–453.

[10] Guo, L., Rivero, D., Dorado, J., Rabunal, J. R., & Pazos, A. (2010). Automatic epileptic seizure detection in eegs based on line length feature and artificial neural networks. *Journal of neuroscience methods*, *191*, 101–109.

[11] Jaiswal, A. K., & Banka, H. (2018). Epileptic seizure detection in eeg signal using machine learning techniques. *Australasian physical & engineering sciences in medicine*, *41*, 81–94.

[12] Lahmiri, S. (2018). An accurate system to distinguish between normal and abnormal electroencephalogram records with epileptic seizure free intervals. *Biomedical Signal Processing and Control*, *40*, 312–317.

[13] Lahmiri, S., & Shmuel, A. (2018). Accurate classification of seizure and seizure-free intervals of intracranial eeg signals from epileptic patients. *IEEE Transactions on Instrumentation and Measurement*, *68*, 791–796.

[14] Mesraoua, B., Deleu, D., Al Hail, H., Melikyan, G., Boon, P., Haider, H. A., & Asadi-Pooya, A. A. (2019). Electroencephalography in epilepsy: look for what could be beyond the visual inspection. *Neurological Sciences*, *40*, 2287–2291.

[15] Sharmila, A., & Geethanjali, P. (2019). A review on the pattern detection methods for epilepsy seizure detection from eeg signals. Biomedical Engineering / Biomedizinische Technik, 64, 507–517. URL: https://doi.org/10.1515/bmt-2017-0233. doi:doi:10.1515/bmt-2017-0233.

[16] Shoeb, A. (2000). Chb-mit scalp eeg database. URL: https://doi.org/10.13026/C2K01R.

[17] Smith, S. J. (2005). Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*, ii2–ii7.

[18] 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*, 222.

Appendix B: Plag Report

Similarity Report

PAPER NAME
MAJOR PROJECT 123.pdf

AUTHOR
a a

WORD COUNT
12459 Words

CHARACTER COUNT
68259 Characters

PAGE COUNT
49 Pages

FILE SIZE
1.9MB

SUBMISSION DATE
Apr 16, 2024 10:34 AM GMT+5:30

REPORT DATE
Apr 16, 2024 10:35 AM GMT+5:30

11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

Excluded from Similarity Report

- Bibliographic material

11% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	accc.gov.au Internet	1%
2	Gaowei Xu, Tianhe Ren, Yu Chen, Wenliang Che. "A One-Dimensional C... Crossref	<1%
3	link.springer.com Internet	<1%
4	mdpi.com Internet	<1%
5	Palak Handa, Monika Mathur, Nidhi Goel. "EEG Datasets in Machine Le... Crossref	<1%
6	dokumen.pub Internet	<1%
7	"Applied Intelligence and Informatics", Springer Science and Business ... Crossref	<1%
8	Fiji National University on 2022-11-20 Submitted works	<1%

9	Swetha Lenkala, Revathi Marry, Susmitha Reddy Gopovaram, Tahir Ceti... Crossref	<1%
10	New Mexico Highlands University on 2024-02-25 Submitted works	<1%
11	University of Rome Tor Vergata on 2023-06-28 Submitted works	<1%
12	undp.org Internet	<1%
13	Victoria University on 2017-04-05 Submitted works	<1%
14	nature.com Internet	<1%
15	Bournemouth University on 2023-05-19 Submitted works	<1%
16	Taylor’s Education Group on 2023-11-19 Submitted works	<1%
17	Yol, Şeyma. "Detection of Epileptic Seizures from EEG Signals by Using... Publication	<1%
18	CSU, San Jose State University on 2023-05-15 Submitted works	<1%
19	Siuly, , and Yan Li. "A novel statistical algorithm for multiclass EEG sig... Crossref	<1%
20	arxiv.org Internet	<1%

21	frontiersin.org Internet	<1%
22	Al Akhawayn University in Ifrane on 2021-12-18 Submitted works	<1%
23	Tinu Theckel Joy, Santu Rana, Sunil Gupta, Svetha Venkatesh. "Fast hy... Crossref	<1%
24	braininformatics.springeropen.com Internet	<1%
25	Indian Institute of Technology, Madras on 2024-02-15 Submitted works	<1%
26	University of Strathclyde on 2023-08-04 Submitted works	<1%
27	iaeme.com Internet	<1%
28	science.gov Internet	<1%
29	P. Srihari, V. Santosh, Sannasi Ganapathy. "An epileptic seizures diagn... Crossref	<1%
30	export.arxiv.org Internet	<1%
31	Supriya Supriya, Siuly Siuly, Hua Wang, Yanchun Zhang. "Automated ep... Crossref	<1%
32	Indian Institute of Technology, Ropar on 2021-09-13 Submitted works	<1%

33	University of Dundee on 2020-03-03 Submitted works	<1%
34	igi-global.com Internet	<1%
35	smartech.gatech.edu Internet	<1%
36	kluniversity.in Internet	<1%
37	A. S. Muthanantha Murugavel, S. Ramakrishnan. "Hierarchical multi-cla... Crossref	<1%
38	Federal University of Technology on 2021-09-12 Submitted works	<1%
39	Fortis College - Centerville on 2024-01-17 Submitted works	<1%
40	University of Ulster on 2023-05-05 Submitted works	<1%
41	academic-accelerator.com Internet	<1%
42	docplayer.com.br Internet	<1%
43	ijrpr.com Internet	<1%
44	vdoc.pub Internet	<1%

45	web.archive.org Internet	<1%
46	researchgate.net Internet	<1%
47	Heriot-Watt University on 2023-04-22 Submitted works	<1%
48	Khalifa University of Science Technology and Research on 2023-10-31 Submitted works	<1%
49	Sim University on 2015-03-01 Submitted works	<1%
50	Technological University Dublin on 2023-07-26 Submitted works	<1%
51	fastercapital.com Internet	<1%
52	jitm.ut.ac.ir Internet	<1%
53	mdpi-res.com Internet	<1%
54	testmagzine.biz Internet	<1%
55	Aayesha, Muhammad Bilal Qureshi, Muhammad Afzaal, Muhammad S... Crossref	<1%
56	Acharya, U. Rajendra, S. Vinitha Sree, G. Swapna, Roshan Joy Martis, a... Crossref	<1%

57	Chandigarh University on 2024-03-12 Submitted works	<1%
58	Devi Sri Prasad Puvvala, Gopichand Madala, Maheendra Kada, U Harih... Crossref	<1%
59	Liverpool John Moores University on 2020-03-03 Submitted works	<1%
60	Taghi M. Khoshgoftaar. "Indirect classification approaches: a compara... Crossref	<1%
61	Universidade do Porto on 2019-02-09 Submitted works	<1%
62	hindawi.com Internet	<1%
63	ijisae.org Internet	<1%
64	medrxiv.org Internet	<1%
65	ncbi.nlm.nih.gov Internet	<1%
66	Central Queensland University on 2013-05-22 Submitted works	<1%
67	University of Moratuwa on 2021-12-07 Submitted works	<1%
68	Brunel University on 2024-01-18 Submitted works	<1%

69	**Lecture Notes in Computer Science, 2016.** Crossref	<1%
70	**Liverpool John Moores University on 2021-09-05** Submitted works	<1%
71	**University of Southern Queensland on 2019-09-01** Submitted works	<1%

APPENDIX C: RESEARCH PAPER PROOF

IEEE11thICRITO/____________

AMITY UNIVERSITY
UTTAR PRADESH

IEEE
UP SECTION (INDIA)

AMITY INSTITUTE OF INFORMATION TECHNOLOGY

IEEE 11th INTERNATIONAL CONFERENCE
ON
RELIABILITY, INFOCOM TECHNOLOGIES AND OPTIMIZATION (ICRITO 2024)
(TRENDS AND FUTURE DIRECTIONS)

CERTIFICATE OF PARTICIPATION

This is to certify that ~~Prof./Dr./M~~s./Mr. **Mantra Jain** of **Department of Computer Science, Amity School of Engineering and Technology, Amity University, Noida** has participated and presented paper titled **Bayesian-Enhanced EEG Signal Analysis for Epileptic Seizure Recognition: A 1D - CNN LSTM Approach** during the **IEEE 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO 2024)** organised by Amity Institute of Information Technology from March 14-15, 2024 at Amity University Uttar Pradesh, India.

Prof. (Dr.) Rekha Agarwal
General Chair, ICRITO 2024

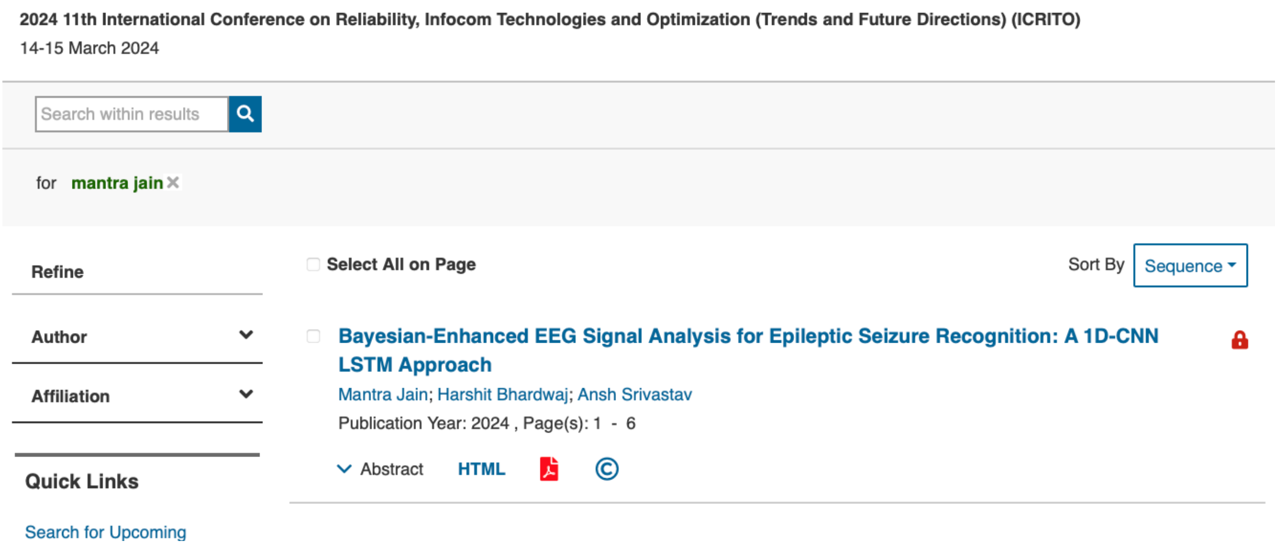
Prof. (Dr.) Sunil Kumar Khatri
General Chair, ICRITO 2024

Prof. (Dr.) K.M. Soni
General Chair, ICRITO 2024

Prof. (Dr.) Balvinder Shukla
Co-Patron, ICRITO 2024
Vice Chancellor, AUUP

15th March 2024

Published Paper Link: https://ieeexplore.ieee.org/document/10522346/authors#authors



COMMENT SHEET

Name and Signature of External Examiner: