PAPER NAME
Ansh_individual_majorproject.pdf

AUTHOR
Harshit Bhardwaj

WORD COUNT
7341 Words

CHARACTER COUNT
40119 Characters

PAGE COUNT
30 Pages

FILE SIZE
416.2KB

SUBMISSION DATE
May 19, 2024 5:15 PM GMT+5:30

REPORT DATE
May 19, 2024 5:16 PM GMT+5:30

10% Overall Similarity

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

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

Excluded from Similarity Report

- Bibliographic material

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

I. INTRODUCTION

I 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. I 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 I 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.

I 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. I want to get a better understanding of these algorithms efficacy in identifying epilepsy and forecasting seizures by analysing their advantages and disadvantages.

I want to see the potential of machine learning and deep learning, including algorithms like logistic regression and CNN, in classification as I 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

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[11]. 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.

Research was carried out by Laura Abraira et al. [12] 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.

The importance of EEG signals for researching brain-related disorders is emphasized in this review [13] 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. [14] 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. [15]

The primary aim of the paper authored by Supriya, S., Siuly, S., Wang, H. et al. [16] 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 [17] 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 [18] 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. [19] 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.

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

I have applied two domains of artificial intelligence that are machine learning and deep learning in this report. I 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, I 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 I have also use deep learning algorithms so that I 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 I can also see the conclusion table 5.1 I 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. I will go through the "BONN University-EEG Dataset" dataset frequently utilized in epilepsy in the part after that.

The BONN EEG Time Series Epilepsy Dataset constitutes an important tool for epilepsy research and neurology. The dataset [20] 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 [20]

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

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

B. Experimental Setup

The hardware used here is of an apple MacBook Air with M1 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. I 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.

I have reported details on the experimental results of the applied Machine Learning Algorithms. Table 4.1 shows training over 70 % of data and testing for 30%, Table 4.2 shows training over 60 % of data and testing for 40%.

C. 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

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

12 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, I 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.

D. 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, I 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.

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

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.

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

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.

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

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.

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%.

A. Performance Measures

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

1) Precision

Measures the accuracy of positive predictions.

$$Precision = \frac{TP}{(TP+FP)} \tag{5}$$

2) Recall

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

$$Recall = \frac{TP}{(TP+FN)} \tag{6}$$

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.

$$F-1\ Score = 2 \times \frac{Recall\ X\ Precision}{(Recall+Precision)} \tag{7}$$

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.

$$Accuracy\ Score = \frac{TP+TN}{TP+FN+TN+FP} \quad (8)$$

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.

$$Macro\ Avg = \frac{1}{N}\sum_{i=1}^{N} Metric_i \quad (9)$$

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.

$$Weighted\ Avg = \frac{1}{N}\sum_{1}^{N}\frac{Metric_i X Support_i}{Total\ Support} \quad (10)$$

"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*

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 I 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, I 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 I 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.

10% Overall Similarity

Top sources found in the following databases:

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

TOP SOURCES

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

1	global.oup.com Internet	2%
2	link.springer.com Internet	<1%
3	Palak Handa, Monika Mathur, Nidhi Goel. "EEG Datasets in Machine Le... Crossref	<1%
4	Swetha Lenkala, Revathi Marry, Susmitha Reddy Gopovaram, Tahir Ceti... Crossref	<1%
5	New Mexico Highlands University on 2024-02-25 Submitted works	<1%
6	dokumen.pub Internet	<1%
7	Taylor's Education Group on 2023-11-19 Submitted works	<1%
8	CSU, San Jose State University on 2023-05-15 Submitted works	<1%

9	Siuly, , and Yan Li. "A novel statistical algorithm for multiclass EEG sig... Crossref	<1%
10	Swiss School of Business and Management - SSBM on 2024-05-13 Submitted works	<1%
11	arxiv.org Internet	<1%
12	Tinu Theckel Joy, Santu Rana, Sunil Gupta, Svetha Venkatesh. "Fast hy... Crossref	<1%
13	braininformatics.springeropen.com Internet	<1%
14	University of Leeds on 2024-05-05 Submitted works	<1%
15	iaeme.com Internet	<1%
16	kkwagh on 2024-05-02 Submitted works	<1%
17	export.arxiv.org Internet	<1%
18	quarxiv.authorea.com Internet	<1%
19	sciencebuddies.org Internet	<1%
20	Indian Institute of Technology, Ropar on 2021-09-13 Submitted works	<1%

21	smartech.gatech.edu Internet	<1%
22	kluniversity.in Internet	<1%
23	A. S. Muthanantha Murugavel, S. Ramakrishnan. "Hierarchical multi-cla... Crossref	<1%
24	Fortis College - Centerville on 2024-01-17 Submitted works	<1%
25	University of Ulster on 2023-05-05 Submitted works	<1%
26	docplayer.com.br Internet	<1%
27	ijrpr.com Internet	<1%
28	vdoc.pub Internet	<1%
29	Colorado State University, Global Campus on 2023-05-15 Submitted works	<1%
30	Khalifa University of Science Technology and Research on 2023-10-31 Submitted works	<1%
31	jitm.ut.ac.ir Internet	<1%
32	mdpi-res.com Internet	<1%

33	frontiersin.org Internet	<1%
34	"Machine Intelligence and Big Data Analytics for Cybersecurity Applica... Crossref	<1%
35	Acharya, U. Rajendra, S. Vinitha Sree, G. Swapna, Roshan Joy Martis, a... Crossref	<1%
36	Victoria University on 2017-04-05 Submitted works	<1%
37	ar5iv.labs.arxiv.org Internet	<1%
38	infoscience.epfl.ch Internet	<1%
39	hindawi.com Internet	<1%
40	nature.com Internet	<1%
41	ncbi.nlm.nih.gov Internet	<1%
42	Roozbeh Zarei, Jing He, Siuly Siuly, Guangyan Huang, Yanchun Zhang. ... Crossref	<1%
43	Taghi M. Khoshgoftaar. "Indirect classification approaches: a compara... Crossref	<1%