

Forecasting Epileptic Seizures Using XGBoost Methodology and EEG Signals

Sunkara Mounika¹ and Reeya S R^{2,*}

^{1,2}School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

Abstract

INTRODUCTION: Epilepsy denotes a disorder of neurological origin marked by repetitive and spontaneous seizures without any apparent trigger. Seizures occur due to abrupt and heightened electricity flowing through the brain, which can lead to physical and mental symptoms. There are several types of epileptic seizures, and epilepsy itself can be caused by various underlying conditions. EEG (Electroencephalogram) is one of the most important and widely used tools for epileptic seizure prediction and diagnosis. EEG uses skull sensors to record electrical signals from the brain, and it can provide valuable insights into brain activity patterns associated with seizures.

OBJECTIVES: Brain-computer interface technology pathway for analyzing the EEG signals for seizure prediction to eliminate the class imbalance issue from our dataset in this case, a SMOTE approach is applied. It is observable that there are more classes of one variable than there are of the others in the output variable. This will be problematic when employing different Artificial intelligence techniques since these algorithms are more likely to be biased towards a certain variable because of its high prevalence

METHODS: SMOTE approaches will be used to address this bias and balance the number of variables in the response variable. To develop an XGBoost (Extreme Gradient Boosting) model using SMOTE techniques to increase classification accuracy.

RESULTS: The results show that the XGBoost method achieves a 98.7% accuracy rate.

CONCLUSION: EEG-based model for seizure type using the XGBoost model for predicting the disease early. The Suggested method could significantly reduce the amount of time needed to accomplish seizure prediction.

Keywords: Machine Learning, Epileptic Seizures, XGBoost, Electroencephalogram, SMOTE

Received on 25 December 2023, accepted on 21 March 2024, published on 27 March 2024

Copyright © 2024 S. Mounika *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.5569

*Corresponding author. Email: reeya.sr@vitap.ac.in

1. Introduction

The electrical patterns in the cerebellum might suddenly and uncontrollably vary during epileptic attacks that can result in a broad spectrum of physical and mental symptoms. They are the primary clinical manifestation of epilepsy, a neurological disorder affecting millions of people worldwide. An updated database with the EEG signals from 14 epilepsy patients, gathered at the University of Siena, Italy, through scalp measurements is presented [1]. Logistic regression offers a simple and interpretable model that can develop a precise decision boundary by utilizing frequency data.

On the other hand, CNNs are powerful models for image-like data (such as spectral power images), and they can capture complex spatial patterns, making them well-suited for EEG signal analysis. Integrating both methods can offer a comprehensive and accurate approach to epileptic seizure prediction. It might be argued that a preferable strategy is to employ machine intelligence to find patterns in data that are invisible to the human eye and to detect seizure states and medication regimens objectively and automatically [2]. The efficient performance of popular deep learning models by employing transfer learning on Enhanced data, with a typical duration lag of 10 minutes between accurate predictions and the actual onset of seizures [3].

A real-time patient-specific estimation system is created on a System-on-Chip that incorporates a built-in

processor and configurable logic blocks utilizing surface electrode-recorded unipolar EEG data. Field-Programmable Gate Arrays (FPGAs), can be used to process unipolar EEG data obtained from surface electrodes. FPGAs offer flexibility and high-performance capabilities, making them suitable for real-time signal processing tasks in various applications, including EEG data analysis [4]. The author describes a novel technique for automatically identifying and classifying EEG signals recorded during epileptic seizures and EEG signals obtained from seizure-free individuals using a clever computer-aided method [5]. A description of features based on the cross-bi spectrum is applicable for identifying epileptic seizure activity [6]. The suggested system is a strong contender for real-time use since a novel algorithm for selecting channels is proposed to choose the most pertinent channels. To identify the subset of EEG channels that carry the most discriminative information for seizure detection. A channel selection algorithm is a crucial component of an EEG-based seizure detection system, especially when considering real-time usage. In EEG data not all channels may contribute equally to seizure detection. Choosing the most pertinent EEG channels can simplify processing, improve efficiency, enhance the accuracy of the seizure detection system [7].

A method for predicting electroencephalography (EEG) seizures utilizing various wavelet transform families. Wavelet transform is a powerful signal processing technique that can reveal the time-frequency information present in EEG signals, making it well-suited for analyzing brain activity during pre-seizure states [8]. The architecture of the ANN is designed, and it usually involves recurrent layers to capture temporal dependencies effectively. RNNs and LSTMs are widely favored in the context of modeling sequential data because of their proficiency in handling such data patterns. A tracking-based learning system specifics are activated by the ANN, a computational model that has evolved biologically. For developing the forecast outcomes, it is frequently used [9]. The seizure prediction system that has been presented integrates deep learning algorithms with the evaluation of EEG scores [10]. The GGN model is designed to operate on the graph representation of the EEG signals. GGN specifically tailored to generate graphs or make predictions on graphs. The architecture of the GGN model can vary based on the particular circumstance and data characteristics. The primary objective of this model is to enable dynamic exploration and identification of brain functional connectivity patterns. The majority of prediction algorithms rely on deep learning methods, which are difficult to explain and have great computing complexity [11].

Figure 1 shows several advances in techniques and procedures for epilepsy diagnosis. High-density EEG with a larger number of electrodes provides better spatial resolution and helps in localizing the source of epileptic activity more accurately. Advanced source localization algorithms and software tools aid in pinpointing the origin of seizures in the brain. The magnetic waves produced by brain activity are recorded using the non-invasive radiology technique known as MEG. It offers high temporal and spatial resolution and

can be valuable for localizing epileptic sources, especially in cases where EEG results are inconclusive. fMRI combined with connectivity analysis can help study the functional connectivity between different brain regions during seizures, providing insights into the epileptic network. Computational models of brain networks can be personalized for individual patients using their neuroimaging data. These models help identify epileptic foci and understand the interactions between different brain regions involved in epilepsy. Ghosh, H. et al. (2023) Presents an innovative [17] machine learning approach to assess water quality, highlighting its importance in environmental monitoring and resource management. Rahat, I.S. et al. (2023) Investigates deep learning [18] applications in brain MR imaging for glioma analysis, advancing precision medicine through enhanced tumor understanding. Ghosh, H. et al. (2023) Demonstrates [19] the use of convolutional neural networks for potato leaf disease prediction, contributing to agricultural technology and crop health. Mandava, M. et al. (2023) Integrates [20] machine and deep learning methods for cardiovascular disease prediction in Bangladesh, underscoring AI's role in healthcare analytics. Mandava, M. et al. (2023) Applies [21] deep learning for detecting yellow rust in wheat, marking a significant contribution to smart farming and disease management. Khasim, S. et al. (2023) Focuses [22] on using AI for real-time rice-leaf disease diagnosis in Bangladesh, enhancing agricultural technology and crop disease management. Khasim, S. et al. (2023) Discusses [23] the role of AI in microorganism image recognition, addressing challenges and advancements in microbiological applications. Mohanty, S.N. et al. (2023) Employs [24] advanced deep learning models for corn leaf disease classification, contributing to precision agriculture and crop health analysis. Alenezi, F. et al. (2021) Introduces [25] CNN-based underwater image dehazing for improved depth estimation, offering advancements in marine research imaging.

Haptic interactions involve the sense of touch and proprioception, and the brain regions are involved in processing haptic information. In active haptic interactions, where an individual initiates the touch or explores an object actively, there might be additional involvement of motor-related brain regions in charge of motor preparation and execution. The supplementary motor area could be more engaged in active haptic interactions compared to passive ones. In passive haptic interactions, where an individual receives haptic feedback without active movement, the focus may be more on sensory processing and perceptual integration, potentially involving areas such as the somatosensory cortex and the secondary somatosensory cortex [12]. The author suggested an innovative two-stage statistical procedure that is straightforward to understand and compute [13] to overcome these problems. Many models place a heavy emphasis on temporally static FC parameters across the course of a scan, which decreases their sensitivity to dynamic elements of brain activity. The mentioned plug-in graph neural network offers adaptability for integration into a model of primary fMRI learning, resulting in enhanced

temporal responsiveness [14]. When examining previous efforts from Kaggle, 32 and 64-node dense layers were chosen. Recognizing a potential chance to improve model performance, a 128-node dense layer is used in the architecture, intending to investigate its impact on accuracy.

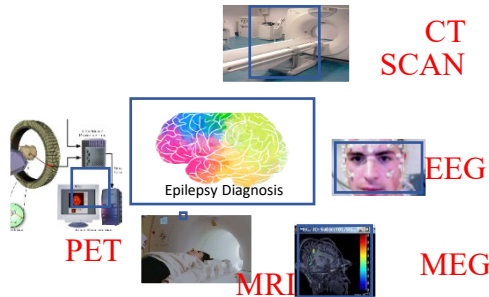


Figure 1. The most recent techniques and procedures for epilepsy diagnosis.

2. Related Work

2.1 Database

The epileptic seizure Recognition (Taken from Kaggle) is a Multivariate characteristic number of Instances 11500 Number of Attributes 179. The brain activity was monitored for 23.6 seconds in total. There are 4097 data points and 23 chunks made from the split and scrambled data. So in a sense, each segment consists of 178 data points (columns) representing one second of data. These segments are eventually partitioned into 23 smaller sections, each having 4097 points of data, also representing one second of data. We must multiply 23 by 500, which is 11500 bits of data, or rows, to determine the entire amount of information. The information about the label we will receive, which spans from 1st to 5 Qtr. categorical variables with equal partitions and is displayed in Figure 2.

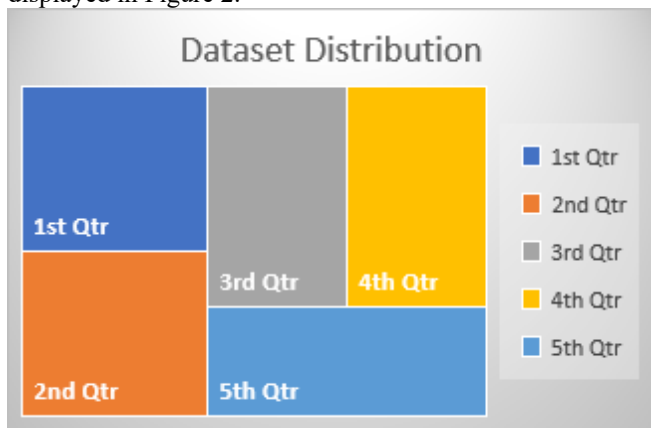


Figure 2. Dataset Distribution

The CHB-MIT dataset comprises several days' worth of continuous scalp EEG recordings. Recordings from 23 Boston Children's Hospital epileptic children. Using the global 1020 system, multi-channel EEG signals the information was captured at an acquisition rate of 256 Hz. The dataset is 23.5 GB in size. Comparison Results obtained from a number of machine learning techniques are presented in Table 1. The preictal condition was determined to be a signal that occurred 30 minutes before the seizure began, and the interictal state was established as occurring at least 4 hours after any seizure. To ensure that the preictal states last for 30 minutes, the next seizure is omitted when there are less than 30 minutes between two adjacent seizures. Patients were tested if they had at least three preictal and interictal states recorded. The explanation is that an overfitting issue would arise if there were fewer than three preictal or interictal states.

The Bonn Dataset is widely recognized and renowned in the field of EEG signals used for epileptic seizure prediction and classification. The dataset was collected at Bonn University in Germany and scientists developed and evaluated algorithms for seizure prediction. The dataset is 11.8 GB in size It contains EEG recordings from both seizure-free intervals and seizure periods. A wavelet-based approach the named flexible analytic wavelets transform. The Bonn Dataset typically includes multiple channels of EEG data, with each channel representing the electrical activity recorded from different locations on the scalp. The recordings are usually sampled at a specific frequency, such as 173.61 Hz or 200 Hz. we have observed three different techniques used for epileptic seizure prediction, each evaluated on different datasets shown in Table 1[15] [16].

Table 1. Machine learning-based comparison tables for epileptic seizure prediction.

Technique	Features Used	Dataset Name	PRE CISION (%)	Specificity (%)	Accuracy (%)
Logistic Regression, KNN[16]	Statistical Measures, Frequency	bonn	82	56	80
			92	79	93
Naïve B[16]	Statistical Measures, Time-domain		65	61	80
Logistic Regression, KNN[16]	Statistical Measures, Frequency	ChB-mit	83	78	81
			95	89	91
Naïve B[16]	Statistical Measures, Time-domain		58	82	60

XGBoost	Statistical Measures, Frequency and Time domain	Epileptic Seizure Recognition	-	-	98.7
---------	---	-------------------------------	---	---	------

2.2 Synthetic Minority Over-Sampling Technique (SMOTE)

When one class has much more than another, there is a class imbalance. A well-liked technique for addressing class imbalance in algorithmic learning samples is SMOTE., where one class has significantly fewer samples than the others SMOTE approaches will be used to address this bias and balance the number of variables in our response variable. SMOTE works by generating Artificially generated data instances belonging to the minorities in the effort to equalize the segregation of classes. Applications that use unbalanced data can be solved on two levels [17]:

Pre-processing data before learning using algorithms that under- or over-sample the majority sample, respectively (hybrid sampling).

Algorithmic: Using algorithms designed for processing learning with unbalanced data.

SMOTE is a data augmentation approach that creates synthetic samples for the underrepresented group by extrapolating between the existing samples of the minority class [18]. Figure 3 represents the data imbalance of the dataset.

The mathematical equation for generating a synthetic sample using SMOTE is as follows. Suppose we have a minority class sample "sample" with feature vector S_i and a randomly chosen k-nearest neighbor "sample_nn" with feature vector S_{nn} . Let D be the Euclidean distance between

S_i and S_{nn} . To create a synthetic sample "synthetic_sample" with feature vector $S_{synthetic}$, we calculate the synthetic feature vector as follows:

$$S_{synthetic} = s-I + rand(0,1)*(S_{nn}-S_i). \quad (1)$$

Where:

$rand(0, 1)$ is a random value between 0 and 1.

$(S_{nn} - S_i)$ represents the vector difference between the feature vectors of "sample_nn" and "sample_i".

$S_{synthetic}$ is the feature vector of the synthetic sample.

2.3 Machine Learning Models

The main goal of this effort was to use an issue with binary categorization to predict seizures. Several supervised machine learning models were used to distinguish between the states of preictal and interictal. The most effective model was XGBoost, which, due to its scalability and parallelization, is perfect for categorising massive amounts of data. [19]. The XGBoost algorithm for two groups are created from EEG information by seizures with epileptic forecasting: seizure and non-seizure circumstances. XGBoost is a robust and well-known gradient-boosting algorithm that excels in many classification problems. XGBoost delivers a feature significance score that shows Considering the significance of each characteristic in the dataset in the prediction task. Make use of this information to learn which EEG signal components are most important for seizure prediction. If the model's outcome fulfills the specified requirements, it can be used to forecast seizures in the actual world. The XGBoost algorithm provides the following advantages over other machine learning approaches. [20].

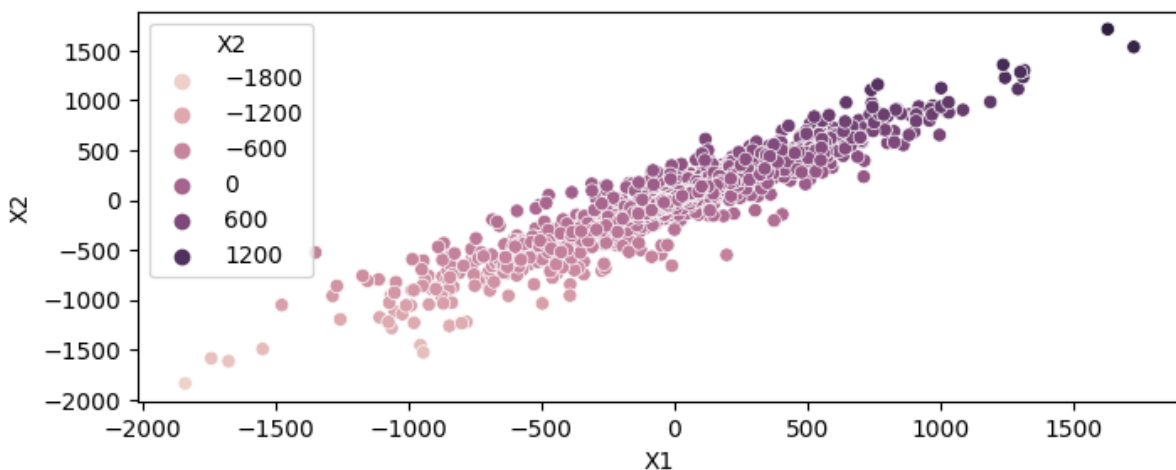


Figure 3. Data imbalance

Incredibly productive expandable

XGBoost is meant to be exceptionally fast and can handle large datasets with many elements. It employs a parallel and distributed computing framework, making it scalable and capable of using CPUs with several cores.

Regularizing approaches

XGBoost includes many regularisation techniques, including L1 (Lasso) and L2 (Ridge) periodicity, to assist in preventing excessive fitting and increasing adaptability.

Gradient boosting

It is a group learning system that uses gradient boosting as its underlying mechanism. This means that it builds several inadequate learners progressively, with each learner focusing on the mistakes of their predecessors. This cyclical technique results in a powerful overarching model.

Handling missing data

XGBoost includes the ability to handle data that is lacking. It dynamically learns the way to handle missing data through the instruction process, reducing the necessity for imputation approaches.

Feature importance ranking

It has a feature importance, a rating system that helps you decide which attributes are most crucial for making predictions. This is essential when choosing qualities while comprehending the influence of every characteristic on the model's performance.

Diversity in objective functions

Lets users design their own specific target functions, which is very beneficial when working with certain problem types or refining specific metrics.

Support for various data types

Unlike many other algorithms, XGBoost can accept both numerical and categorical features without the requirement for explicit encoding.

Figure 4 represents visualizing all channels stacked against one another is a common approach to gaining insights into multi-channel data, such as EEG signals, time series, or images. Stacking the channels allows us to observe the patterns and relationships across different channels. Stack all the channels together in a single plot, placing them one above the other. Each channel's time series will be represented as a line plot representing the time or data points. Table 2 shows the Seizure identification performance is compared between different Datasets. Different datasets are used for evaluating seizure identification algorithms. These datasets may contain EEG (Electroencephalogram) recordings from patients with and without seizures.

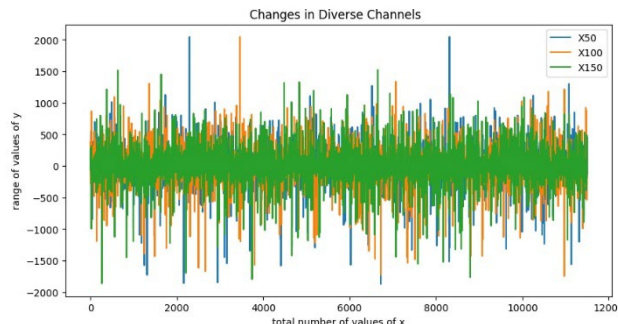


Figure 4. Channel visualization

Table 2. The performance of seizure identification is compared across Datasets.

Dataset	Method	Accuracy
CHB-MIT[21]	XGBoost	83.33
Epilepsy Ecosystem[22]	XGBoost	90.41
Epileptic Seizure Recognition	XGBoost	98.7

2.4 Model Evaluation

The metrics Accuracy represented on a scale from 0 to 1, and were taken into account while measuring performance:

$$Accuracy = ((X_p + X_n)/(X_p + X_n + Y_p + Y_n)) \quad (2)$$

The symbols X_p , X_n , Y_n , and Y_p belong to in that order, Real-Positives, True - Negatives, False-Negatives, and False-Positives.

3. Results

The ROC curve graphically illustrates the binary classifier's performance changes as the discrimination threshold is adjusted. It illustrates the connection between true positive rate (sensitivity) and false positive rate (FPR) at various threshold configurations, as shown in Figure 5. A perfect classifier will have an AUC-ROC of 1, while a random or poor classifier will have an AUC-ROC close to 0.1. The results of the suggested model's performance is better performance when compared to Existing traditional models.

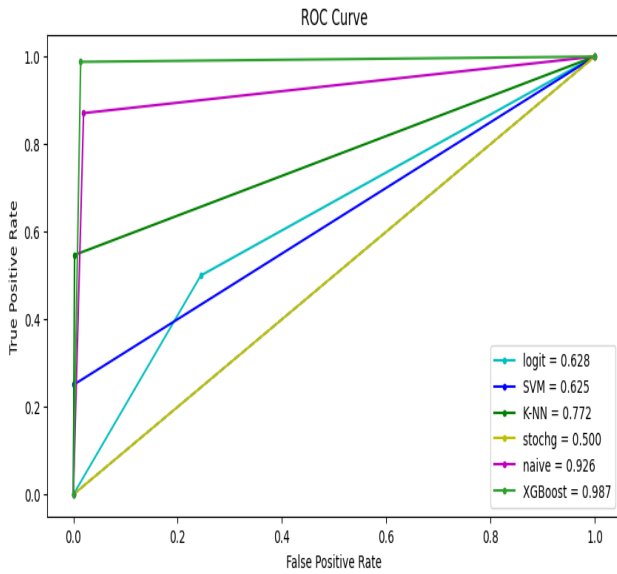


Figure 5. ROC curve

Figure 6 illustrates How the connection is linear between two variables is evaluated by their correlation coefficient in terms of both its directional nature and power. This coefficient spans from -1, denoting a complete negative linear correlation, to 1, signifying a complete positive linear correlation, while 0 signifies an ideal non-linear relationship, indicating that the variables are not correlated. Positive correlation indicates that two variables are they tend to increase together as one does. A negative correlation, on the other hand, indicates that when one variable rises, the other is more likely to fall. Correlation values close to 0 suggest no linear relationship between the variables. Correlation matrices are especially useful for feature selection and understanding multicollinearity in regression models. In machine learning, they can help identify redundant or highly related features, which can be removed to reduce model complexity and improve interpretability.

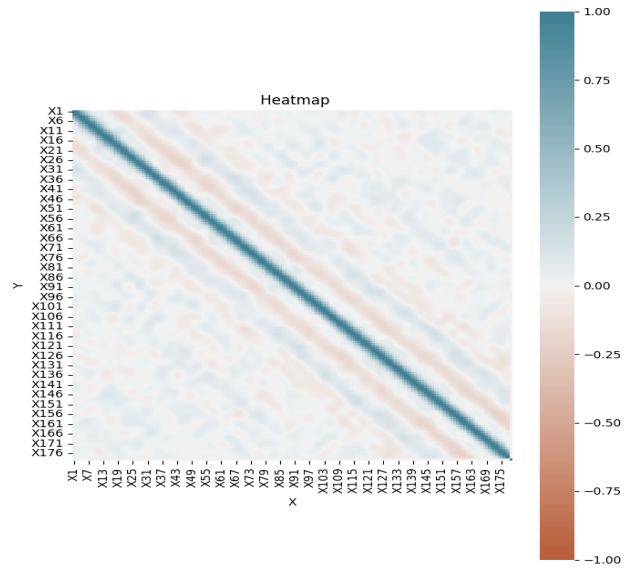


Figure 6. Color map representation of cross-correlation matrix

4. Conclusion

To boost the model's representational capacity and accuracy, With Rectified Linear Unit (ReLU) activation, a 128-node dense layer is added to the model. Thus, the suggested model can predict whether an individual is experiencing a seizure or not rather quickly. The accuracy rate for this system is up to 98.7%, demonstrating the effectiveness of the algorithm. By using more surface electrodes, expanded data gathering, and an ensemble of sophisticated algorithms, the suggested method could significantly reduce the amount of time needed to accomplish seizure prediction.

References

- [1] Hafeez A. Agboola; Colette Solebo; David S. Aribike; Afolabi E. Lesi; Alfred A. Susu; "Seizure Prediction with Adaptive Feature Representation Learning", JOURNAL OF NEUROLOGY AND NEUROSCIENCE, 2019.
- [2] Paolo Detti; Giampaolo Vatti; Garazi Zabalo Manrique de Lara; "EEG Synchronization Analysis for Seizure Prediction: A Study on Data of Noninvasive Recordings", 2020.
- [3] Khansa Rasheed; Junaid Qadir; Terence J. O'Brien; Levin Kuhlmann; Adeel Razi; "A Generative Model To Synthesize EEG Data For Epileptic Seizure Prediction", ARXIV-CS.LG, 2020.
- [4] Ercan Coşgun; Anıl Çelebi; "FPGA Based Real-time Epileptic Seizure Prediction System", BIOCYBERNETICS AND BIOMEDICAL ENGINEERING, 2021.
- [5] Ahmed I Sharaf; Mohamed Abu El-Soud; Ibrahim M El-Henawy; "An Automated Approach For Epilepsy Detection Based On Tunable Q -Wavelet And Firefly

- Feature Selection Algorithm", INTERNATIONAL JOURNAL OF BIOMEDICAL IMAGING, 2018.
- [6] Naghmeh Mahmoodian; Axel Boese; Michael Friebe; Javad Haddadnia; "Epileptic Seizure Detection Using Cross-bispectrum Of Electroencephalogram Signal", SEIZURE, 2019.
- [7] Hisham Daoud; Magdy A Bayoumi; "Efficient Epileptic Seizure Prediction Based On Deep Learning", IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS, 2019.
- [8] Saly Abd-Elateif El-Gindy; Asmaa Hamad; Walid El Shafai; Ashraf A. M. Khalaf; Sami M. El-Dolil; Taha E. Taha; Adel S. El-Fishawy; Turky N. Alotaiby; Saleh A. Alshebeili; Fathi E. Abd El-Samie; "Efficient Communication and EEG Signal Classification in Wavelet Domain for Epilepsy Patients", J. AMBIENT INTELL. HUMANIZ. COMPUT., 2021
- [9] P. Suguna; B. Kirubagari; R. Umamaheswari; "An Analysis of Epileptic Seizure Detection and Classification Using Machine Learning-Based Artificial Neural Network", 2021
- [10] S Kannan; G Premalatha; M Jamuna Rani; D Jayakumar; P Senthil; S Palanivelrajan; S Devi; Kibebe Sahile; "Effective Evaluation of Medical Images Using Artificial Intelligence Techniques", COMPUTATIONAL INTELLIGENCE AND NEUROSCIENCE, 2022.
- [11] Zhengdao Li; Kai Hwang; Keqin Li; Jie Wu; Tongkai Ji; "Graph-generative Neural Network for EEG-based Epileptic Seizure Detection Via Discovery of Dynamic Brain Functional Connectivity", SCIENTIFIC REPORTS, 2022.
- [12] Haneen Alsuradi; Wanjoon Park; Mohamad Eid; "Assessment of EEG-based Functional Connectivity in Response to Haptic Delay", FRONTIERS IN NEUROSCIENCE, 2022.
- [13] Hao Chen; Taoyun Ji; Xiang Zhan; Xiaoxin Liu; Guojing Yu; Wen Wang; Yuwu Jiang; Xiao-Hua Zhou; "An Explainable Statistical Method for Seizure Prediction Using Brain Functional Connectivity from EEG", COMPUTATIONAL INTELLIGENCE AND NEUROSCIENCE, 2022.
- [14] Irmak Sivgin; Hasan A. Bedel; Şaban Öztürk; Tolga Çukur; "A Plug-in Graph Neural Network to Boost Temporal Sensitivity in fMRI Analysis", ARXIV-EESS.SP, 2023.
- [15] Khaled Mohamad Almustafa, Classification of epileptic seizure dataset using different machine learning algorithms, Informatics in Medicine Unlocked, Volume 21, 2020, 100444, ISSN 2352-9148.
- [16] H O Lekshmyl, Dhanyalaxmi Panickar1 and Sandhya Harikumar1 2022 *J. Phys.* Comparative analysis of multiple machine learning algorithms for epileptic seizure prediction: *Conf. Ser.* 2161 012055. DOI 10.1088/1742-6596/2161/1/012055.
- [17] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4_6
- [18] Rahat IS, Ghosh H, Shaik K, Khasim S, Rajaram G. Unraveling the Heterogeneity of Lower-Grade Gliomas: Deep Learning-Assisted Flair Segmentation and Genomic Analysis of Brain MR Images. EAI Endorsed Trans Perv Health Tech [Internet]. 2023 Sep. 29 [cited 2023 Oct. 2];9. <https://doi.org/10.4108/eetpht.9.4016>
- [19] Ghosh H, Rahat IS, Shaik K, Khasim S, Yesubabu M. Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks. EAI Endorsed Scal Inf Syst [Internet]. 2023 Sep. 21 <https://doi.org/10.4108/eetsis.3937>
- [20] Mandava, S. R. Vinta, H. Ghosh, and I. S. Rahat, "An All-Inclusive Machine Learning and Deep Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population", EAI Endorsed Trans Perv Health Tech, vol. 9, Oct. 2023. <https://doi.org/10.4108/eetpht.9.4052>
- [21] Mandava, M.; Vinta, S. R.; Ghosh, H.; Rahat, I. S. Identification and Categorization of Yellow Rust Infection in Wheat through Deep Learning Techniques. EAI Endorsed Trans IoT 2023, 10. <https://doi.org/10.4108/eetiot.4603>
- [22] Khasim, I. S. Rahat, H. Ghosh, K. Shaik, and S. K. Panda, "Using Deep Learning and Machine Learning: Real-Time Discernment and Diagnostics of Rice-Leaf Diseases in Bangladesh", EAI Endorsed Trans IoT, vol. 10, Dec. 2023 <https://doi.org/10.4108/eetiot.4579>
- [23] Khasim, H. Ghosh, I. S. Rahat, K. Shaik, and M. Yesubabu, "Deciphering Microorganisms through Intelligent Image Recognition: Machine Learning and Deep Learning Approaches, Challenges, and Advancements", EAI Endorsed Trans IoT, vol. 10, Nov. 2023. <https://doi.org/10.4108/eetiot.4484>
- [24] Mohanty, S.N.; Ghosh, H.; Rahat, I.S.; Reddy, C.V.R. Advanced Deep Learning Models for Corn Leaf Disease Classification: A Field Study in Bangladesh. Eng. Proc. 2023, 59, 69. <https://doi.org/10.3390/engproc2023059069>
- [25] Alenezi, F.; Armghan, A.; Mohanty, S.N.; Jhaveri, R.H.; Tiwari, P. Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light. Water 2021, 13, 3470. <https://doi.org/10.3390/w13233470>
- [26] Werner de Vargas, V., Schneider Aranda, J.A., dos Santos Costa, R. et al. Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. Knowl Inf Syst 65, 31–57 (2023). <https://doi.org/10.1007/s10115-022-01772-8>
- [27] L. Wei and C. Mooney, "Epileptic Seizure Detection in Clinical EEGs Using an XGboost-based Method," 2020 *IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Philadelphia, PA, USA, 2020, pp. 1-6, doi: 10.1109/SPMB50085.2020.9353625.
- [28] Shafiezadeh S, Duma GM, Mento G, Danieli A, Antoniazzi L, Del Popolo Cristaldi F, Bonanni P, Testolin A. Methodological Issues in Evaluating Machine Learning Models for EEG Seizure Prediction: Good Cross-Validation Accuracy Does Not Guarantee Generalization to New Patients. *Applied Sciences*. 2023; 13(7):4262. <https://doi.org/10.3390/app13074262>
- [29] Elika Karbassiyazdi, Fatemeh Fattahi, Negin Yousefi, Amirhessam Tahmassebi, Arsia Afshar Taromi, Javad Zyaie Manzari, Amir H Gandomi, Ali Altaee, Amir Razmjou, XGBoost model as an efficient machine learning approach for PFAS removal: Effects of material characteristics and operation conditions, Environmental Research, Volume 215, Part 1, 2022, 114286, ISSN 0013-9351, <https://doi.org/10.1016/j.envres>.
- [30] Häggglund, M. "Optimization of Pre-Ictal Interval Time Period for Epileptic Seizure Prediction." *caring is sharing—exploiting the value in data for health and innovation* (2023): 232.

- [31] K. Manasvi Bhat, P. P. Anchalia, S. Yashashree, R. Sanjeetha and A. Kanavalli, "Detection and Prediction of the Preictal State of an Epileptic Seizure using Machine Learning Techniques on EEG Data," *2019 IEEE Bombay Section Signature Conference (IBSSC)*, Mumbai, India, 2019, pp. 1-5, doi: 10.1109/IBSSC47189.2019.8972992.