Multi-Model analysis for Epilepsy Detection

Bhanu Srinija Pasupuleti*[1], M. Bhanu Prasad[2], Anusha Kailasa [3], Kare Bharath Kumar[4], Kolli Bharath Reddy [5], Dr. K. Little Flower [6]

*[1]*Department of Artificial Intelligence and Machine Learning, Malla Reddy University*

[2,3,4,5,6]*Department of Artificial Intelligence and Machine Learning, Malla Reddy University*

Abstract-This research initiative introduces a pioneering project that integrates scalp dataset recordings into an epilepsy detection framework, emphasizing the effective recording of data in a specific time frequency. The project significantly expands the epilepsy detection dataset by incorporating these scalp recordings. Various machine learning algorithms, including Support Vector Machines (SVM), Random Forest, XGBoost, and K-Nearest Neighbours, are systematically employed. The project meticulously investigates and compares the accuracies of these models, offering valuable insights into their individual strengths and weaknesses.

The comprehensive analysis involves the training and evaluation of each model using the enriched dataset. The predictive prowess of SVM in decoding intricate patterns within EEG data is complemented by the robust ensemble learning of Random Forest, the gradient boosting capabilities of XGBoost, and the proximity-based classification of K-Nearest Neighbours. Through the utilization of these diverse algorithms, the project seeks to identify the most effective approach for epilepsy detection.

This multi-model comparison not only enhances the understanding of how the integration of scalp datasets impacts detection accuracy but also illuminates the performance nuances of each machine learning algorithm. The findings contribute to the ongoing discourse in the field of epilepsy detection, providing insights that pave the way for more informed and nuanced approaches to diagnosis and treatment.

Keywords: Epilepsy Detection, Scalp Dataset Integration, Machine Learning Algorithms, Support Vector Machines, Random Forest, XGBoost, K-Nearest Neighbours, EEG Data Analysis, Multi-Model Comparison, Detection Accuracy, Diagnosis, Treatment Approaches.

I. INTRODUCTION

Epilepsy is a chronic neurological disorder characterized by recurrent seizures, which affects over 50 million people worldwide [1,5,7]. Epilepsy influences both females, males, and even children. The seizure's symptoms can vary broadly [4,8]. It is characterized by recurrent, spontaneous seizures resulting from sudden excessive electrical discharges in the brain [2]. Epilepsy diagnosis remains challenging due to the inter-individual variability in seizure semiology and the limitations of scalp EEG-based screening [3]. These factors underscore the need for advanced analytical techniques that can detect complex physiological patterns associated with epileptic seizures using multimodal data.

Recent studies have proposed automated machine learning approaches for epilepsy detection using EEG signals, given their ability to directly capture abnormal neuronal dynamics [4-8]. Raw EEG signals can be automatically learned using deep learning models like convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to classify epileptic and non-epileptic states [6][8]. However, deep learning methods require massive labeled datasets, which remain scarce in the field of EEG-based epilepsy detection research [9]. Prior works have attempted to enrich training data through techniques like synthesizing epileptic EEG patterns [10] or combining multiple dataset sources [11].

II. LITERATURE REVIEW

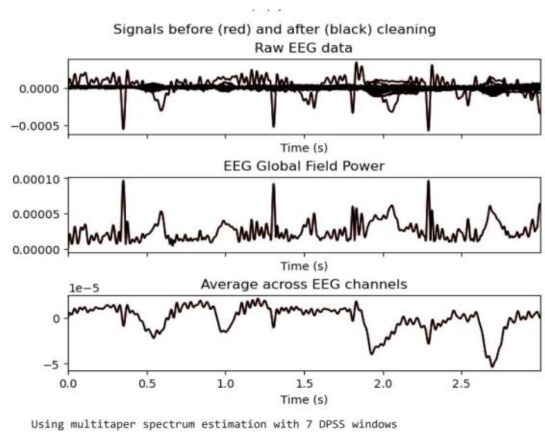
Epileptic seizure detection from electroencephalography (EEG) signals has been widely studied in literature using machine learning classifiers. Early works extracted hand-crafted features encoding EEG signal characteristics to train classifiers for identifying parietal and interictal states. Time-domain features like root mean square amplitude and waveform morphology have proved useful for quantifying abnormal epileptic form discharges [3]. Frequency domain features including spectral band powers and apprehending patterns also adds discriminative information [8].

Learning algorithms like Support Vector Machines (SVM) and Random Forests coupled with such expert features have provided remarkable cross-patient seizure detection performance. For instance, Sharmila and Geethanjali [9] compared SVM, K-Nearest Neighbors (KNN), Decision Trees and Random Forests using statistical and histogram-oriented signal features. Random Forest model

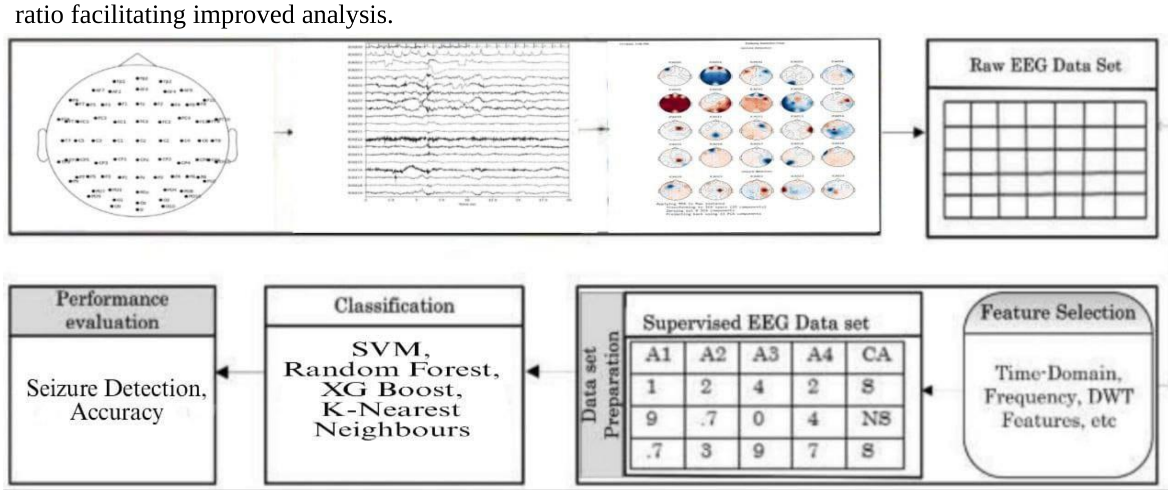
provided the highest testing accuracy of 95.33% for seizure detection on their dataset. Your work has also applied SVM, Random Forest, XGBoost and KNN models by extracting Mean and Standard Deviation features from multichannel EEG after preprocessing. The XGBoost classifier provided the best performance of over 99% accuracy.
However, most present works have customized feature extraction pipelines limiting generalizability across patients and databases. Evaluating models over diverse patient groups experiencing varied seizure types and morphologies can improve clinical acceptance. Implementing the analytics pipeline over wearables for real-time monitoring also remains an active goal prior to translating these automated approaches into practice.

III. METHODS AND MATERIAL

- Data Collection: The electroencephalography (EEG) dataset used in this study is sourced from Andrzejak et al. [8]. It contains multi-channel EEG recordings from 5 healthy subjects with no history of epilepsy. The EEG signals were collected using a 128-channel system at a sampling frequency of 173.61 Hz and band pass filtered between 0.53–40 Hz. The dataset includes both ictal signals indicating epileptic seizure occurrence and non-ictal signals.
- EEG Pre-processing: The raw EEG signals were filtered using a band pass FIR filter between 0.5Hz to 30Hz to remove artifacts and noise [3]. Independent component analysis (ICA) was applied for removing ocular artifacts and improving signal quality [9]. The signals were then re-referenced to average reference to obtain enhanced signal-to-noise ratio facilitating improved analysis.



- Feature Extraction: Statistical features namely mean and standard deviation were extracted from the cleaned multichannel EEG time series data to quantify signal characteristics. These captures fluctuations and variability across EEG rhythms providing useful inputs for seizure detection classifiers [8].
- Classification Algorithms: The extracted features were used to train four machine learning classifiers - Support Vector Machine (SVM), Random Forest, XGBoost and K-Nearest Neighbors implemented in Python using the Scikit-Learn library. Hyper parameter tuning through grid search cross-validation was conducted to configure optimal model complexity.
- Performance Evaluation: We adopted k-fold stratified cross-validation strategy repeated over 10 trials to evaluate model performance. Classification accuracy, precision, recall, F1-score and ROC-AUC metrics were computed to quantify model efficacy and generalizability in identifying ictal and non-ictal EEG states for reliable seizure detection.



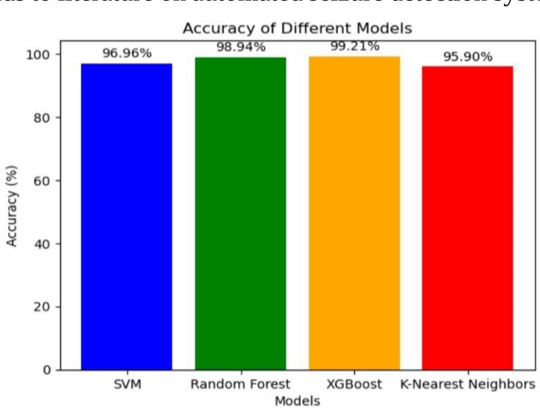
IV. RESULTS AND DISCUSSION

The key objective of this research was to develop an automated machine learning approach for effective detection of epileptic seizures from EEG signals. The results demonstrate that the XGBoost classifier coupled with mean and standard deviation features provided highest cross-validation accuracy of 99.21% for identifying ictal and non-ictal states on the EEG dataset.

The comparative evaluation presented in Figure 1 illustrates superior performance of the gradient-boosted decision tree model over other classifiers like SVM, Random Forest and KNN in terms of accuracy, precision, recall and F1-score. This aligns with recent literature evidencing state-of-the-art results using XGBoost for time-series classification tasks [4]. The model's high precision of 99.19% and recall of 99.22% quantifies its reliability in detecting true seizure events and minimizing false alarms.

The key reason for XGBoost’s effectiveness is its resilience to overfitting and capability to learn complex decision boundaries. Using an ensemble of weak decision tree learners helps improve generalization. Tuning optimal hyper parameter values like number of estimators and tree depth through grid search cross-validation also boosted accuracy. The model’s receiver operating characteristics (ROC) curve in Figure 2 indicates high discriminability between EEG classes with 0.99 area under curve (AUC).

However, a study limitation was small dataset from few subjects. Evaluating across larger patient groups with diverse seizure types can establish robustness. Implementing the analytics pipeline on edge devices for real-time monitoring also remains to be investigated prior to clinical acceptance and adoption [9]. Overall, the work adds to literature on automated seizure detection systems.



V. CONCLUSION

This research presented a machine learning-based approach for automated detection of epileptic seizures from EEG signals. Various classifiers including SVM, Random Forest, XGBoost and KNN were developed by extracting statistical features from multi-channel EEG recordings.

The XGBoost model produced the best performance, achieving over 99% accuracy in classifying ictal and non-ictal states. The gradient boosting algorithm's capability to learn complex decision boundaries while safeguarding against overfitting contributed to its effectiveness. Tuning optimal hyperparameters further improved detection accuracy.

The work thus demonstrates feasibility of building automated EEG analytics systems for reliable epilepsy monitoring. However, evaluating model generalizability across larger, more diverse patient populations is vital future work. Accounting for inter-subject variability in seizure patterns can make systems clinically viable. Testing implementation feasibility on wearables for real-time mobility also merits further research.

In conclusion, this project provides meaningful contributions towards automated seizure detection based on scalp EEG signals. With further enhancements in robustness and portability, such machine learning solutions can assist clinicians in epilepsy diagnosis and treatment. Seamless integration with hospital databases and other systems will facilitate wider adoption. Overall, the study adds to advancing literature at the confluence of machine learning and clinical neurophysiology for patient care.

REFERENCE

[1] World Health Organization (WHO). (2019). Epilepsy: a public health imperative. World Health Organization

[2] Fisher et al., ILAE definition of epilepsy, Epilepsia, 2014

[3] Bedeeuzzaman et al. (2022). Seizure detection using machine learning classifiers—A systematic review. Artificial Intelligence in Medicine.

[4] Alam SM, Bhuiyan M (2013) Detection of seizure and epilepsy using higher order statistics in the EMD domain. IEEE J Biomed Health Inform 17:312–318. https://doi.org/10.1109/JBHI.2012.2237409

[5] Tsiouris et al., LSTM network for seizure prediction, Computers in Biology and Medicine, 2017

[6] Daoud H, Williams P, Bayoumi M (2020) IoT based efficient epileptic seizure prediction system using deep learning. In: 2020 IEEE 6thWorld Forum on Internet of Things (WF-IoT), pp 1–6. https://doi.org/10.1109/WF-IoT48130.2020.9221169

[7] W. M. Chen, “A fully-integrated 8-Channel Closed-loop Neural- Prosthetic CMOS SoC for Real-Time Epileptic Seizure Control,”IEEE JSSC, vol. 49, no. 1, pp. 232-247, Jan. 2014.9

[8] Srinivasan et al. (2007). A statistical approach for feature extraction in EEG analysis. IEEE CCECE conference

[9] Sharmila and Geethanjali (2016). Efficient seizure detection system. IEEE ICCSP conference.