International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Automated Detection of Epileptic Seizures Using EEG Signals and Gradient Boosting Approach

Vadavalli Bhavani

Abstract: ***Epilepsy is a neurological disorder in which seizures recur due to abnormal electrical activity in the brain. It is a critical problem affecting millions of people worldwide, which impacts their quality of life. Electro-encephalography (EEG) is a popular approach in which brain signals can be recorded, and abnormal signals can be identified. However, analyzing EEG signals manually is a tedious job, which requires expertise from neurologists. In addition, analyzing EEG signals for a prolonged period results in a high volume of signals, which makes it difficult to analyze manually. In order to mitigate the challenges in analyzing EEG signals, many researchers have proposed machine learning-based epilepsy detection systems. Machine learning algorithms can analyze EEG signals, identify meaningful features, and classify them as seizure and non-seizure signals. In this research, an automated system to detect seizures by using EEG signals is presented. The system uses EEG signals collected in the Bonn University data set. The system is based on preprocessing EEG signals, feature extraction, and classification by using the Extreme Gradient Boosting algorithm. Nineteen features such as statistical features and frequency features, including mean, standard deviation, entropy, signal energy, and Hjorth features, are used to represent significant characteristics of EEG signals. The features extracted by the system help in training the Extreme Gradient Boosting classifier to classify seizure and normal EEG signals. The results show that the system can accurately classify EEG signals and detect seizure patterns. This system can help neurologists in diagnosing patients more efficiently.***

Keywords: EEG signals, epilepsy detection, seizure classification, machine learning, feature extraction

1. Introduction

Epilepsy is a chronic neurological disorder that affects millions of people worldwide. This neurological disorder is characterized by recurring seizures resulting from unexpected surges in brain electrical activity. According to the World Health Organization, almost 50 million people worldwide suffer from epilepsy. Accurate and timely detection of seizures is crucial in diagnosing and treating epilepsy.

Electroencephalography (EEG) is the primary technique employed in the recording of brain electrical activity. EEG signals carry significant information about brain activity. EEG signals help neurologists in identifying abnormal patterns in the brain that may lead to seizures.

Manual analysis of EEG signals is a difficult and time-consuming process because EEG signals contain a large amount of data that needs to be analyzed over a long time. In recent years, automated seizure detection techniques have gained significant research interest in the field of machine learning and artificial intelligence. Machine learning algorithms can analyze EEG signals and identify features that help in classifying EEG signals into seizure and non-seizure states.

In this study, a machine learning-based epileptic seizure detection system is proposed using the Bonn EEG dataset. Statistical and frequency domain features are used to extract features from the EEG signals, and then XGBoost classification is applied to detect epileptic seizures. The proposed system aims to enhance the accuracy of seizure detection and provide an efficient way to analyze EEG signals.

2. Related Work

2.1 Background and Importance of Automated Seizure Detection

Epilepsy is a chronic neurological disorder that causes recurrent seizures due to abnormal electrical activity in the brain. According to the World Health Organization, epilepsy affects over 50 million people worldwide, making it one of the most common neurological conditions [1]. Therefore, it is essential to detect seizure occurrences to properly diagnose, treat, and monitor patients suffering from epilepsy.

Electroencephalography (EEG) is the most widely used technique to detect the electrical activity of the brain. EEG signals are significant to understand the patterns of neural activity. EEG signals are used to detect abnormal neural activity, which causes epilepsy [2]. Therefore, it is essential to analyze EEG signals to detect abnormal neural activity. Manual analysis of EEG signals to detect abnormal neural activity is a challenging task. Neurologists have to analyze EEG recordings to detect abnormal neural activity. This process may cause significant human error.

Due to recent advances in biomedical signal processing techniques, automated seizure detection systems have become an area of interest. Automated seizure detection systems use EEG signals to detect abnormal neural activity. These systems use machine learning techniques to classify EEG signals into seizure and non-seizure categories. Automated seizure detection systems can significantly reduce the workload of neurologists [3].

2.2 Early Machine Learning Techniques in EEG Signal Classification

In the initial attempts at automating seizure detection, researchers applied conventional machine learning algorithms. The Bonn University EEG Dataset, created by the University of Bonn, is one of the most widely used datasets in EEG signal classification, especially in seizure detection [4]. The Bonn University EEG Dataset includes EEG signals representing normal brain activity, inter-ictal activity, and seizures, which can be used to evaluate machine learning algorithms.

Researchers applied conventional machine learning algorithms, including Support Vector Machine, Decision Tree, K-Nearest Neighbor, and Naive Bayes, in the initial attempts at automating seizure detection [5]. These machine learning algorithms typically use hand-crafted features, which include statistical, frequency, and complexity features. Researchers reported high accuracy using Support Vector Machine classifier algorithms, especially when hand-crafted frequency features, including wavelet coefficients, were used in the algorithm [6].

Other researchers have used ensemble learning algorithms like Random Forest and Gradient Boosting on EEG classification problems [7]. In this approach, the predictions of a set of decision trees are combined to increase the accuracy of the classifiers. The experimental results showed that the nonlinear relationships in the EEG feature set could be effectively exploited to improve the seizure detection accuracy.

2.3 Feature Extraction Techniques in EEG Signal Analysis

Feature extraction is a critical component of EEG-based seizure detection systems. EEG signals contain large amounts of raw data with a lot of noise. It is challenging for classifiers to learn meaningful patterns from raw EEG signals. To address this problem, various signal processing techniques are applied to raw EEG signals to extract meaningful features [8].

Researchers have widely applied various feature categories to EEG signal analysis. Statistical features such as mean value, variance, standard deviation, skewness, and kurtosis represent the distribution of EEG signal amplitudes. Features from the frequency domain using techniques such as the Fast Fourier Transform (FFT) represent the frequency characteristics of EEG signals. In addition, complexity measures such as entropy and Hjorth parameters represent the complexity of EEG signals.

Researchers have demonstrated that using multiple feature categories can result in higher classification accuracy. Statistical, frequency domain, and complexity measures are combined to represent diverse characteristics of EEG signals. Machine learning classifiers can learn meaningful patterns from diverse EEG signal characteristics to differentiate between seizure activity and normal brain activity.

2.4 Dealing with Imbalanced EEG Datasets

Another challenge in the field of seizure detection using machine learning algorithms is the imbalanced nature of the EEG datasets. In real-world EEG datasets, the occurrence of seizures is significantly lower compared to normal brain activity. This imbalanced nature of the dataset can lead to biased machine learning algorithms, as reported in [12].

Several methods have been proposed in the literature to balance the EEG datasets. One of the most widely used methods is the Synthetic Minority Oversampling Technique (SMOTE) proposed by Nitesh V. Chawla, et al. [13]. The main objective of using the oversampling technique is to balance the EEG datasets. By using the oversampling technique, it has been reported in [14] that the evaluation metrics, i.e., recall and F1-score, can be significantly improved. The main objective of using machine learning algorithms in the field of seizure detection is not only to achieve high accuracy on the validation set but also remain sensitive to seizure events.

2.5 Gradient Boosting and XGBoost for Biomedical Applications

In recent years, there has been significant interest in applying gradient boosting algorithms to biomedical data analysis. Among the best implementations of the gradient boosting algorithm, Extreme Gradient Boosting (XGBoost) by Tianqi Chen and Carlos Guestrin [15] stands out. XGBoost is an implementation of the gradient boosting algorithm that focuses on high performance.

The XGBoost algorithm uses decision trees to create an ensemble. Decision trees are built sequentially to correct the prediction errors of the preceding tree. The XGBoost algorithm uses a gradient-based optimization method to minimize the prediction error. XGBoost also uses regularizers to prevent overfitting. These characteristics of XGBoost make it an efficient algorithm for use with structured data that have several features.

Several studies have successfully applied XGBoost to medical classification problems. XGBoost has been successfully applied to biomedical signal processing, disease prediction, healthcare risk analysis, and other medical classification problems. XGBoost has been successfully applied to EEG-based seizure detection problems. XGBoost outperformed other classifiers due to its efficiency in handling non-linear features. XGBoost can handle several features of the data. XGBoost outperformed other classifiers in terms of accuracy.

2.6 Summary of Research Gap

There are a number of machine learning approaches that have been used for seizure detection, but there are still some challenges that need to be addressed for the creation of a more accurate system. Most of the research done previously has focused on the creation of more complex deep learning models, which require a lot of data as well as computational power [18]. On the other hand, the use of a lightweight machine learning model, along with a more efficient feature extraction method, has the potential to provide a more accurate result with reduced computational complexity.

Thus, the focus of the research will be on the creation of a more efficient system for the detection of seizures using EEG signals from the Bonn University dataset, followed by

Paper ID: SR26430095205 DOI: https://dx.doi.org/10.21275/SR26430095205

the classification of seizure and non-seizure EEG signals with the help of an XGBoost classifier.

3. Methodology

The proposed system for epileptic seizure detection employs a methodology based on a systematic approach that combines different signal processing techniques and machine learning methods to automatically detect seizure patterns in EEG signals. The proposed system aims to develop an efficient system that can differentiate seizure and non-seizure states in brain activity with high accuracy. Electroencephalography signals are complex biomedical signals that are considered complex time-series data with important patterns and background noise. Therefore, to detect seizure patterns in EEG signals, different stages are involved in the process.

The proposed methodological framework for epileptic seizure detection is based on a multi-stage system that can transform input signals into a more meaningful form that can be used by different machine learning methods. The proposed system for epileptic seizure detection involves four major stages in its methodology. The stages are data collection and preprocessing, feature extraction and engineering, model development and training, and system evaluation.

The design philosophy behind this research is to incorporate effective feature extraction methods with a robust machine learning classifier to attain effective seizure detection with efficiency. This is different from using complex deep learning models that require huge amounts of data and computational resources to train, as this approach relies on effective feature extraction methods using statistical and frequency domain features along with the XGBoost classifier.

Phase 1: Data Collection and Preprocessing

Dataset Acquisition

The dataset used in this research is the widely used Bonn University EEG Dataset, which is developed by the University of Bonn. The dataset contains EEG recordings from healthy individuals and epileptic patients. The dataset is widely used in research related to seizure detection, mainly because it is in a structured format with labeled samples.

The dataset is comprised of five subsets, labeled A, B, C, D, and E. Each subset contains 100 EEG segments, with each segment containing 4097 data points, which is approximately 23.6 seconds of EEG signal. Subsets A and B contain EEG signals from healthy individuals, while subsets C and D contain interictal EEG signals from epileptic patients. Subset E contains EEG signals from the epileptic seizure activity.

In terms of binary classification in this research, EEG segments can be classified into two classes: seizure and non-seizure signals. The seizure class includes EEG segments in subset E, while the non-seizure class includes EEG segments in subsets A-D. This classification scheme enables the machine learning model to learn patterns that can identify seizure activity from normal brain signals.

Data Pre-processing

The raw EEG signal may have noise or artifacts resulting from various sources such as muscle movements, eye blinks, or environmental interference. Such interference may have an impact on the quality of analysis or the efficiency of signal classification. Therefore, preprocessing techniques are employed to clean or normalize the EEG signal.

To start with, the EEG signal is normalized to ensure that the signal has a consistent range. This is done to guarantee that machine learning models are not affected by any feature that could influence the classification process. Moreover, statistical standardization is performed on the EEG signal. This ensures that the mean value is zero and the variance is equal to 1.

The continuous EEG signal is divided into smaller segments or windows, which are efficient for feature extraction. Each EEG signal segment represents a time window that captures the characteristics of brain activity. This process ensures that meaningful features are obtained without adding complexity. All these steps in signal processing are important in ensuring that the EEG signal is noise-reduced, normalized, and properly structured.

Feature Extraction and Feature Engineering

Feature extraction is another important step in the classification of EEG signals. EEG signals have thousands of data values; hence, it is challenging for machine learning algorithms to detect seizure patterns within the EEG signals. Therefore, meaningful features must be extracted to represent the underlying characteristics of the EEG signals.

In this research, nineteen statistical and frequency domain features are extracted from each EEG segment. The features extracted represent the important characteristics of the EEG signals.

Statistical Features

Statistical features represent the distribution of the EEG signal amplitude. Statistical features comprise the following:

- Mean
- Standard deviation
- Variance
- Skewness
- Kurtosis
- Root Mean Square (RMS)

Frequency-Domain Features

In frequency domain analysis, the spectral characteristics of EEG signals are analyzed. The Fast Fourier Transform (FFT) is applied to the EEG signal, which converts the signal from the time domain to the frequency domain. Using this approach, the frequency components of EEG signals, which cause seizures, can be identified. The features obtained in the frequency domain are as follows:

- Signal energy
- Spectral entropy
- Power spectral density

Paper ID: SR26430095205 DOI: https://dx.doi.org/10.21275/SR26430095205

Hjorth Parameters
Hjorth parameters are widely used in the analysis of EEG signals. These parameters describe the dynamic properties of EEG signals. The three Hjorth parameters used in the study are as follows:
- Activity
- Mobility
- Complexity

In the feature extraction process, a combination of statistical, frequency domain, and complexity-based features is obtained. These features act as input variables in the machine learning classifier.

Phase 2: Model Development and Training

A. Machine Learning Model
Once the feature extraction is complete, the feature vector is used to train the machine learning model. In this research, the machine learning model is Extreme Gradient Boosting, commonly known as XGBoost.

XGBoost is a sophisticated machine learning model classified as an ensemble learning model, which is a derivative of gradient boosting decision trees. The model consists of a set of decision trees, where each tree is trained to minimize the error made by the preceding tree.

XGBoost has numerous benefits, with one of the most outstanding being its ability to work with structured data with high efficiency. The model has a built-in mechanism to prevent overfitting, which is achieved by incorporating regularization techniques. The model is also able to work in parallel, which is a huge advantage in terms of reducing processing time.

B. Model Training Procedure
The data set is divided into two parts: one for training and one for testing. This is done to evaluate the performance of the developed seizure detection system. In this paper, 80% of the data set is used to train the model, and the remaining 20% is used to test the model.

During the training procedure, the XGBoost model learns how to relate the EEG feature set with their respective class labels. The performance of the model is optimized by adjusting the learning rate, maximum depth, and number of estimators.

Cross-validation techniques are employed in the model training procedure to ensure the reliability of the performance evaluation. This also prevents the model from overfitting. The model predicts the EEG signals in the testing phase.

Once the predictions have been generated, the results are evaluated and displayed through the output interface. The system displays the classification results, as well as evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics can be used to measure the efficacy of the proposed seizure detection model.

This structured workflow follows a systematic approach in processing EEG signals from raw signals to classification results. The combination of signal processing techniques and machine learning algorithms allows the system to accurately identify epileptic seizures, thus aiding healthcare professionals in neurological analysis and diagnosis.

Phase 3: System Integration and Security Implementation
The framework for the detection of seizures is implemented as an integrated analytical system that is capable of processing EEG signals and classifying the signals as either seizure or non-seizure activity. The framework is designed using a modular layered architecture that consists of a data input layer, preprocessing layer, feature extraction layer, machine learning classification layer, and result visualization layer.

The data input layer is responsible for loading EEG signal segments from the EEG dataset provided by the Bonn University EEG Dataset. The EEG signals represent the electrical activity of the human brain recorded at different intervals of time. The EEG signal segments are fed into the preprocessing layer where normalization and noise reduction operations are carried out.

The preprocessing layer is responsible for cleaning and normalizing the signal. Once the preprocessing is complete, the EEG signals are passed to the feature extraction layer. During this layer, various statistical and frequency domain features are extracted from the EEG signals. The statistical and frequency domain features that are extracted from the EEG signals include mean, variance, standard deviation, entropy, energy of the signal, skewness of the signal, kurtosis of the signal, root mean square, and Hjorth parameters.

The feature vectors are then passed to the machine learning layer. This layer incorporates the trained Extreme Gradient Boosting classifier. This classifier analyzes the extracted features. This classifier will generate a probability score that will determine whether the EEG signal belongs to epileptic seizure activity or normal brain activity. This classifier is best used for this purpose because it combines various decision tree classifiers using gradient boosting techniques. This will enhance the accuracy of the classifier while preventing overfitting.

Finally, the results are passed to the output layer. This layer will display the results of the classifier. This layer will clearly show outputs such as whether the EEG signal belongs to seizure or non-seizure activity, as well as the confidence level of the classifier. This will enhance the overall flexibility of the seizure detection system.

Working Process Flow
The proposed system for detecting an epileptic seizure works through an efficient workflow process, which aims at transforming the EEG signals into useful predictions.

The process begins with the acquisition of EEG signals from the Bonn University EEG Dataset. The signals are a representation of time series data, which show the activities of the human brain during normal and seizure states. Once the signals have been loaded into the system, preprocessing of the signals takes place, which includes normalizing the signals.

After the preprocessing of the signals, the system performs a series of tasks known as feature extraction, which aims at extracting important attributes of the EEG signals. Statistical values are calculated, which provide a description of the behavior of the signals, their amplitude, and complexity. The extracted attributes are presented as a vector, which represents the EEG signals numerically.

Finally, the system uses a classifier, known as the Extreme Gradient Boosting classifier, which works by analyzing the patterns of the vectors, determining whether the signals are related to an epileptic seizure or normal state of the brain. During this process, the classifier generates a probability value, which indicates the likelihood of a seizure occurring.

4. Results

For the experiments related to the detection of epileptic seizure, the Bonn University EEG Dataset has been used. In this dataset, EEG signals are collected from healthy people and people with epilepsy. The EEG signals are classified into different categories, which include normal brain signals and signals related to epilepsy.

For the experiments, the dataset is split into a training set and a testing set, so the models could be trained on the training set and tested on the testing set, which contains EEG signals. Several machine learning models were tested to detect the seizures, which include Logistic Regression, Random Forest, Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost).

For evaluating the models, the models were tested on the evaluation metrics, which include accuracy, precision, recall, and F1 score. Accuracy is related to the total number of correct predictions made by the models. Precision is related to the total number of predicted signals, which are actually related to seizures. Recall is related to the total number of correctly predicted seizure cases. The F1 score is related to precision and recall.

The Logistic Regression model has an accuracy of 93%. The precision, recall, and F1-score values are 93% for both seizure and non-seizure classes. This shows that the Logistic Regression model is reliable in providing consistent results for EEG signal classification.

The Random Forest model has an overall accuracy of 92%. The precision value for both classes is approximately 92%, with a recall value close to this range. The F1-score value is 92%. This shows that the Random Forest model performs well in classifying EEG signal patterns. However, the performance is not as high as that of other models.

The Support Vector Machine (SVM) model has shown the highest performance among all models. This is because the SVM model has an accuracy of 97%. The precision, recall, and F1-score values are 97% for both classes. This shows that SVM is reliable in classifying seizure and non-seizure EEG signals.

The performance of the Extreme Gradient Boosting (XGBoost) model was also impressive, with an accuracy of 98%. The precision value was 100%, and the recall value was 95% for the non-seizure class. The precision value was 95%, and the recall value was 100% for the seizure class. The F1-score values were 97% and 98%. These results show that machine learning models are effective in detecting epileptic seizure patterns using EEG signals.

The analysis of the confusion matrix shows that the models are reliable because they indicate the number of EEG signals correctly or incorrectly classified as seizure or non-seizure. From the results, most EEG signals are correctly classified as seizure or non-seizure. This shows that machine learning models are effective in classifying EEG signals as seizure or non-seizure.

5. Conclusion

This research introduced an automated system for detecting epileptic seizures based on EEG signal analysis and machine learning approaches. Accurate detection of epileptic seizures is crucial in diagnosing neurological disorders. The proposed system employed EEG data collected from the Bonn University EEG Dataset. The EEG data was analyzed using signal preprocessing techniques and machine learning approaches to detect seizures.

In this research, various machine learning algorithms such as Logistic Regression, Random Forest Classifier, Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) were employed to evaluate their performance in detecting epileptic seizures. Among these algorithms, the SVM and XGBoost algorithms were found to yield the highest classification accuracy of 97%, which is significant in differentiating seizure and non-seizure EEG signals.

The experimental results show that machine learning approaches can effectively analyze EEG signals to yield accurate seizure detection results. This automated system can also help neurologists to efficiently perform EEG signal analysis.

In conclusion, the proposed system for detecting epileptic seizures is an efficient solution in EEG-based neurological analysis. The system can also be improved in the future by integrating deep learning approaches and EEG monitoring techniques to yield accurate seizure prediction results.

References

[1] World Health Organization, *Epilepsy Fact Sheet*, 2023.
[2] U. R. Acharya et al., "Automated EEG analysis of epilepsy," *Knowledge-Based Systems*, 2013.
[3] T. Tzallas, M. Tsipouras, and D. Fotiadis, "Epileptic seizure detection in EEG signals," *IEEE Trans. Information Technology in Biomedicine*, 2009.
[4] R. G. Andrzejak et al., "Indications of nonlinear deterministic structures in EEG signals," *Physical Review E*, 2001.

[5] A. Subasi, "EEG signal classification using wavelet feature extraction and SVM," *Expert Systems with Applications*, 2007.
[6] N. Nicolaou and J. Georgiou, "Detection of epileptic EEG signals using entropy features," *Expert Systems with Applications*, 2012.
[7] U. R. Acharya et al., "Automated diagnosis of epilepsy using EEG signals," *Computer Methods and Programs in Biomedicine*,2012.
[8] S. Sanei and J. Chambers, *EEG Signal Processing*. Wiley, 2007.
[9] M. A. Hasan et al., "A review of EEG signal processing techniques," *Biomedical Signal Processing and Control*, 2016.
[10] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, 1970.
[11] U. R. Acharya et al., "Automated EEG-based diagnosis of epilepsy," *Knowledge-Based Systems*, 2013.
[12] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowledge and Data Engineering*, 2009.
[13] N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, 2002.
[14] G. Batista et al., "A study of oversampling methods for imbalanced datasets," *SIGKDD Explorations*, 2004.
[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *ACM SIGKDD*, 2016.
[16] J. Chen et al., "Machine learning for healthcare applications," *IEEE Access*, 2019.
[17] S. Roy et al., "Epileptic seizure detection using machine learning techniques," *IEEEAccess*,2020.
[18] M. Shoeibi et al., "Automatic epileptic seizure detection using deep learning," *Computers in Biology and Medicine*, 2021.