Real-time Epileptic Seizure Detection and Prediction Using EEG Data

By Ria Jayanti

Abstract

Epilepsy is one of the most common neurological disorders worldwide, characterized by recurrent seizures caused by abnormal and excessive electrical brain activity. These seizures can be unpredictable, leading to brain damage, physical injury, severe depression, anxiety, and even death. The condition affects approximately 50 million people globally, and while up to 70% of those affected could potentially become seizure-free with proper medical diagnosis and treatment, nearly 80% live in low- and middle-income countries. It is estimated that 75% of people with epilepsy who live in low-income countries do not receive adequate medical treatment (World Health Organization, 2024). Therefore, it is imperative that at-home seizure detection and prediction technology becomes more accessible, particularly in underserved regions. This research utilizes the CHB-MIT Scalp EEG Database, applying machine learning models (K-Nearest Neighbors, Logistic Regression, Random Forest Classifier, and Support Vector Machine) for seizure detection, along with Long Short-Term Memory (LSTM) time-series forecasting for seizure prediction. To address the significant class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was employed, increasing the percentage of seizure samples in the dataset from 0.23% to 50.00%. The Random Forest Classifier and Support Vector Machine models both achieved 100% accuracy and recall for seizure detection, and the LSTM model achieved 89.26% accuracy and 89% weighted recall for seizure prediction. This research demonstrates the potential of AI-driven methods to be integrated into at-home seizure-monitoring devices, making seizure detection and prediction technology significantly more accessible.

Introduction

Can EEG data be used to detect and predict epileptic seizures? Epilepsy is a chronic neurological disorder affecting approximately 1 in 26 Americans, and up to 40% of these patients have drug-resistant epilepsy (Tao, n.d.). Furthermore, more than 1 in 1000 adults with epilepsy unexpectedly dies each year. Many of these deaths could be prevented by increasing education on the disease (*SUDEP*, 2013). This research allows for detection of epileptic seizures, allowing patients to receive earlier treatment, which can improve long-term prognoses and decrease the risks of complications or seizure-related injuries. In addition, by predicting epileptic seizures, this research helps patients seek assistance from caregivers or medical professionals before a seizure occurs, which can be life-saving. Predicting seizures before they occur also helps patients identify seizure triggers and make lifestyle changes. This allows the condition to become more manageable and offers patients a sense of control over their condition, decreasing anxiety.

This is a classification problem, as the goal is to classify each segment of the EEG as either seizure or non-seizure, in order to detect seizure events. For prediction, the model aims to predict when a seizure might occur by identifying patterns in EEG data that are associated with seizure events. Hence, this research involves supervised learning, as labeled data (EEG recordings with identified seizure events) is used to train the model.

This research uses the 2010 CHB-MIT Scalp EEG database, which includes EEG recordings from 22 pediatric patients with intractable seizures at Boston Children's Hospital. These EEG recordings consist of numerical signals recorded from the scalp, which reveal the electrical activity in the brain.

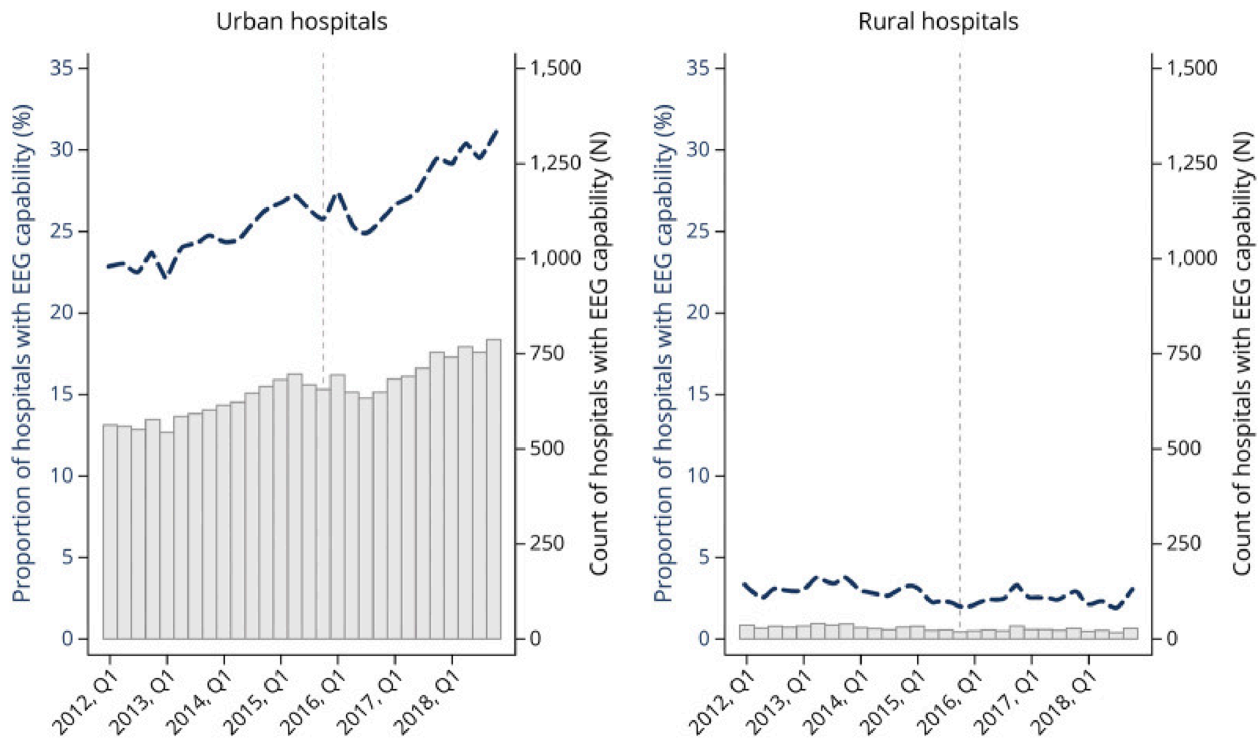
For epileptic seizure detection, I employed multiple machine learning algorithms, including K-Nearest Neighbors (KNN), Logistic Regression, Random Forest Classifier (RFC), and Support Vector Machine (SVM). For seizure prediction, I used a Long Short-Term Memory (LSTM) model. To address class imbalance, I applied the Synthetic Minority Oversampling Technique (SMOTE), increasing the proportion of seizure occurrences in the dataset.

Background

Several approaches have been used to diagnose and detect epileptic seizures using EEG data. One notable approach is from the University of Southern California (USC), where researchers developed an AI model that analyzes the positions of EEG electrodes to identify rare seizure symptoms. This model achieved a 12% higher accuracy in detecting seizures than similar technologies. A unique contribution of this approach is that the researchers aim for it to eventually be integrated with smartphone technology, allowing patients to receive real-time alerts when abnormal EEG activity is detected (Dawson, 2024).

Another approach is from Washington University, where researchers designed a device to detect seizures and pinpoint their locations within the brain with greater precision than existing technologies. This device uniquely analyzes the interactions between different brain regions and filters out background noise, such as a person scratching their arm, which often interfere with the results of EEG tests. The device's ability to localize seizures is a major advancement in seizure detection work, as it can be used to create more personalized and effective treatments (Jefferson, 2020).

Both the USC and Washington University models offer a major pro: making at-home EEG tests more common. This need for at-home tests is due to the majority of hospitals in the U.S. not offering EEG tests. In fact, in 2012, only 20.4% of hospitals nationwide were EEG-capable, and by 2018, this number had only increased to 27.3%. Moreover, 90% of these EEG-capable hospitals were in urban settings (Suen et. al., 2023).



(Suen et al., 2023).

Figure 1. Proportion of hospitals with EEG capability in urban vs. rural regions in the U.S.

Many people are hence unable to access hospital EEG tests due to their geographic location, making at-home tests an effective way to increase accessibility. However, with at-home tests comes a major con: the cost. A standard hospital EEG test in the U.S. ranges from $200 to more than $1250 (Medicare.org). This cost is already a significant financial burden for many patients, and at-home tests often cost more due to their increased convenience, advanced technology, and the potential need for a caregiver or medical professional to handle any malfunctions or concerns. Hence, both the USC and Washington University models would likely be unaffordable for many patients, especially those in underserved areas.

Therefore, despite the advancements in seizure detection models, there remains a significant need for affordable at-home EEG technologies. Moreover, the USC and Washington University models, along with similar technologies, focus on diagnosing and detecting epileptic seizures, but not predicting them. Predicting seizures and preventing them before they occur is a critical way to prevent the risks associated with epilepsy and make the condition more manageable. This research focuses on both diagnosing and predicting epileptic seizures, bridging this gap.

Furthermore, there has been limited research on epilepsy itself. Most research on at-home EEG devices has focused on applications like stress reduction and sleep quality improvement, with limited relevance in epilepsy management. This research aims to make these at-home devices more suitable for epilepsy patients. By predicting seizures before they occur and preventing potential risks or complications, this research can decrease the need for expensive hospital equipment and procedures long-term. Thus, this research increases the affordability and accessibility of treatment for epilepsy patients, especially those in rural or low-income areas.

Dataset

I used a dataset of scalp electroencephalogram (EEG) recordings from pediatric patients at Boston Children's Hospital experiencing intractable seizures and not taking anti-seizure medication. The dataset includes recordings from 23 subjects, with 664 total edf files. Of those files, 129 contain at least 1 seizure event, with 198 seizure events in total. Each case includes between 9 and 42 continuous .edf files from 23 electrodes placed around the head, capturing brain activity during both normal states and seizure events. Each recording is about one hour long. There are 18 female patients and 5 male patients in the dataset. Patients are between ages 1.5-22, and the median age is 10 years (Guttag, J). I split the dataset into training and testing sets using random sampling, allocating 80% for training and 20% for testing. Each recording is divided into 2-second epochs, allowing for the observation of specific patterns correlated with seizure occurrence (Guttag, J).

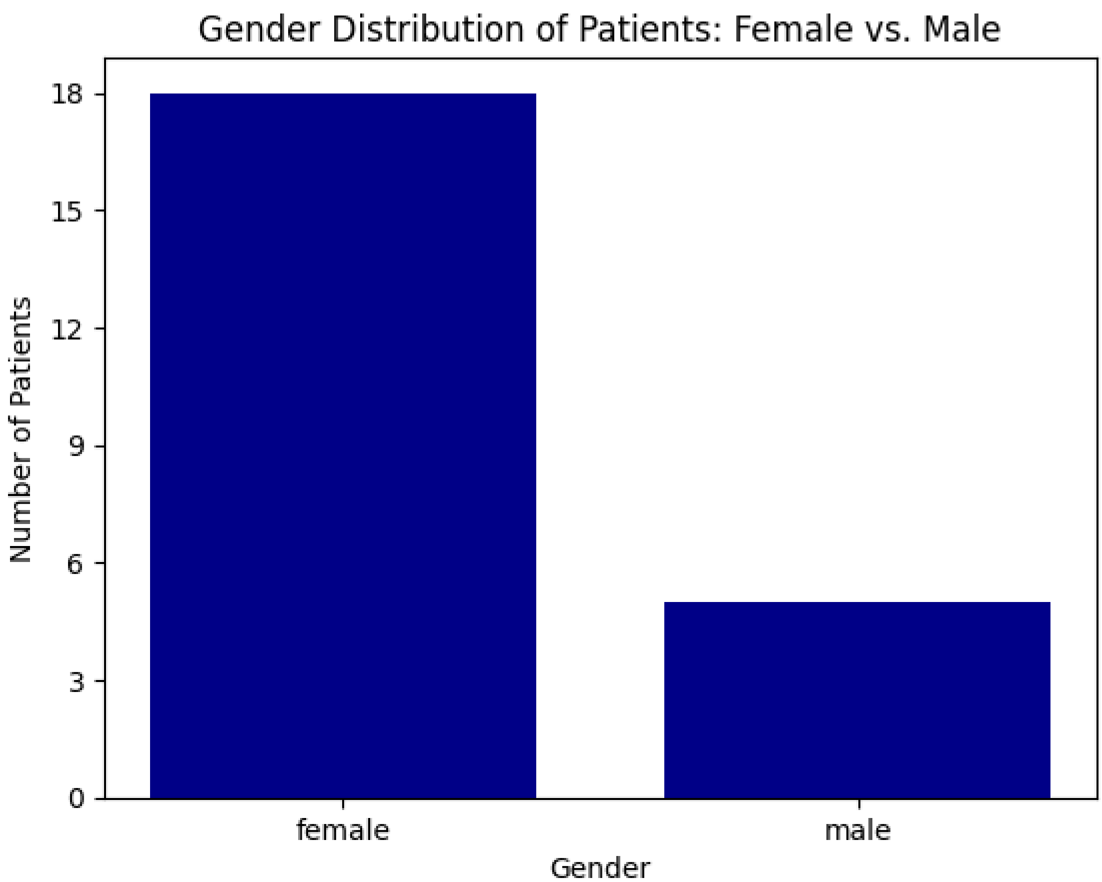


Figure 2. Bar graphs categorizing patients by gender

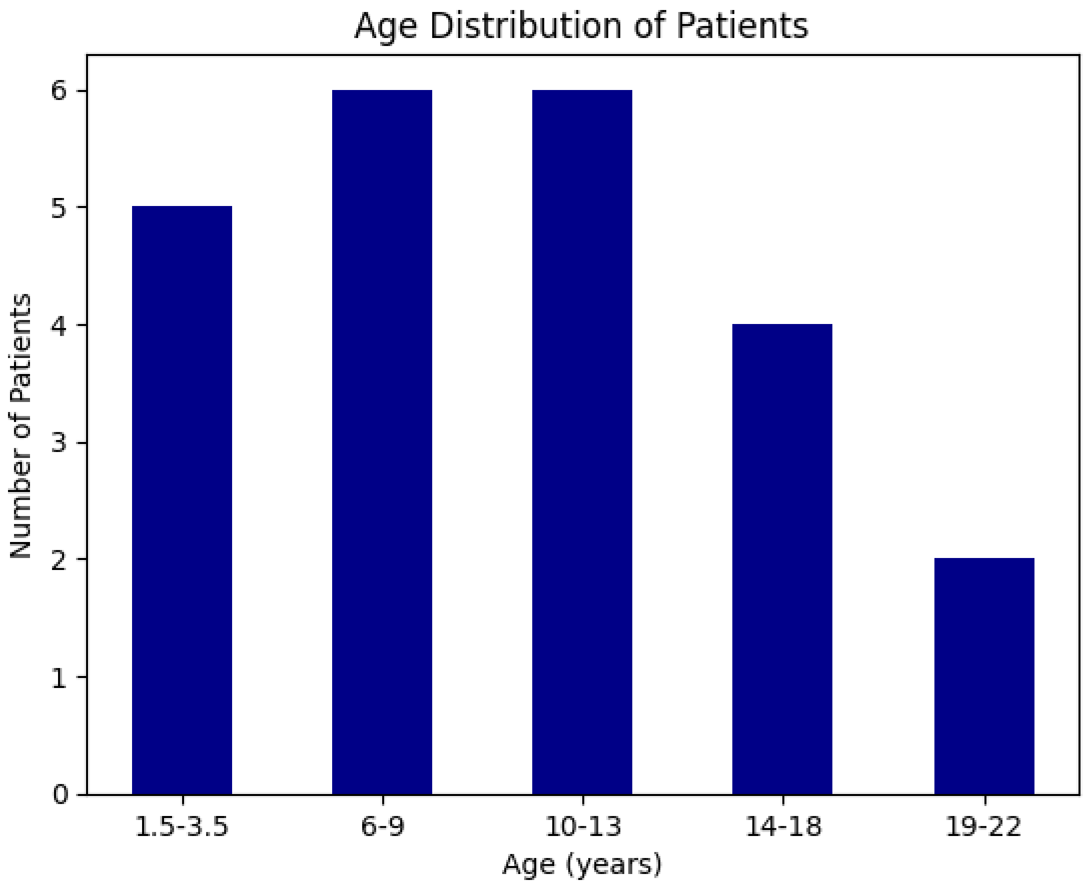


Figure 3. Bar graphs categorizing patients by age

Methodology

For epileptic seizure detection, I first split the dataset into 80% for training and 20% for testing. Several machine learning algorithms were used to train the model: K-Nearest Neighbors (KNN), Logistic Regression, Random Forest Classifier (RFC), and Support Vector Machine (SVM).

K-Nearest Neighbors (KNN) works by classifying a data point based on the majority class of its nearest neighbors. I set the value of ‘k’ to 2 for this model. The algorithm computes the distance between the test sample and all training samples, then classifies the sample according to the most common class among the nearest points. Logistic Regression is a linear model that predicts the probability of a class by finding a decision boundary that best separates the classes. This model uses a logistic function to compute the probability of a binary outcome, aiming to model the class distribution with a linear decision boundary. Random Forest Classifier (RFC) constructs multiple decision trees and combines their outputs through majority voting. Each tree is trained on a random subset of the data, which reduces overfitting compared to a single decision tree. To assess its performance, I used cross-validation, which provided a more reliable measure of how the model would perform on new data. Support Vector Machine (SVM) aims to find the optimal best line or surface, called a hyperplane, that separates data points from different classes. For this model, I used the Radial Basis Function (RBF) kernel, which transforms the data into a higher-dimensional space to make the separation easier. This kernel helped improve the performance of the model by providing more flexibility in handling complex data distributions.

To evaluate the models, I used accuracy and recall as the primary metrics. For each model, a confusion matrix was plotted to visualize its true positives, true negatives, false positives, and false negatives, in order to understand how its performance in detecting seizures.

For seizure prediction, I used a Long Short-Term Memory (LSTM) model, which is a type of recurrent neural network (RNN) designed to learn from sequences of data. LSTM is particularly effective for time-series data like EEG signals because it can retain long-term dependencies, making it ideal for detecting patterns that lead up to seizures.

The EEG data were processed using Python's MNE library, which helped convert the raw EEG signals into epochs. Each epoch was then labeled as either a non-seizure or seizure event. I applied noise reduction techniques such as Independent Component Analysis (ICA) and Signal Space Projection (SSP) to improve the signal quality.

Because the data was highly imbalanced — 9063 non-seizure samples and only 21 seizure samples — I applied the Synthetic Minority Oversampling Technique (SMOTE). SMOTE generates synthetic seizure samples to balance the dataset and reduce the bias towards the non-seizure class. It works by identifying the 'k' nearest neighbors for each sample in the minority class (seizures). For each seizure sample, SMOTE generates new synthetic data points by creating new samples along the line connecting the original sample to its nearest neighbors. After applying SMOTE, the class distribution was 9063 non-seizure samples and 4531 seizure samples, which significantly improved the recall for detecting seizures. I also extracted statistical features such as the mean, maximum, minimum, and standard deviation for each epoch before feeding the data into the LSTM model. The data was then split into training, validation, and test sets to assess the model's performance in detecting seizures.

Results and Discussion

The results from the various models showed significant differences in performance, depending on the algorithm used. The KNN classifier performed exceptionally well, achieving an accuracy of 99.98% and a perfect recall of 1.0. Similarly, the Random Forest Classifier and Support Vector Machines both reached 100% accuracy and recall, demonstrating their strong ability to correctly classify both seizure and non-seizure events. However, the Logistic Regression model showed much weaker results, with a test accuracy of only 66.68% and a recall of 0.0. Despite attempts to adjust for class imbalance by tuning the class weights, Logistic Regression struggled with detecting the minority class (seizures), which contributed to its poor recall.

The use of SMOTE proved to be crucial for improving the models' performance. Without SMOTE, the models all had a recall of 0.0, displaying an extreme bias towards the non-seizure class. By balancing the dataset, SMOTE helped the models become more sensitive to the minority class, leading to better seizure detection.

The confusion matrices for each model provide a detailed breakdown of their performance. In the confusion matrices, the top-left quadrant represents true positives (correctly identifying seizures), the top-right quadrant shows false positives (incorrectly identifying non-seizure events as seizures), the bottom-left quadrant shows false negatives (missed seizures), and the bottom-right quadrant shows true negatives (correctly identifying non-seizure events).

Below are the confusion matrices for each detection model:

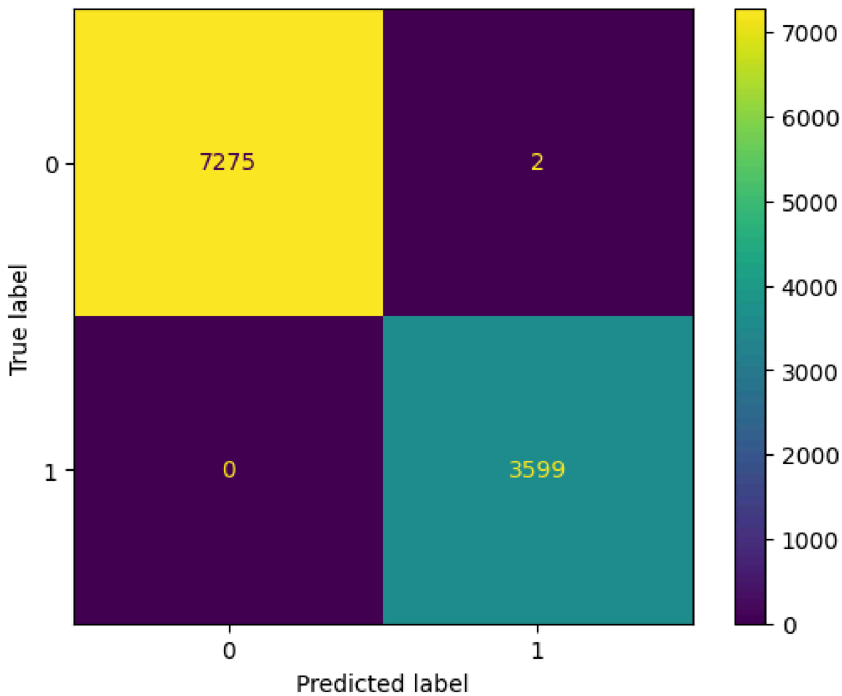


Figure 4. Confusion matrix for K-Nearest Neighbors (KNN)

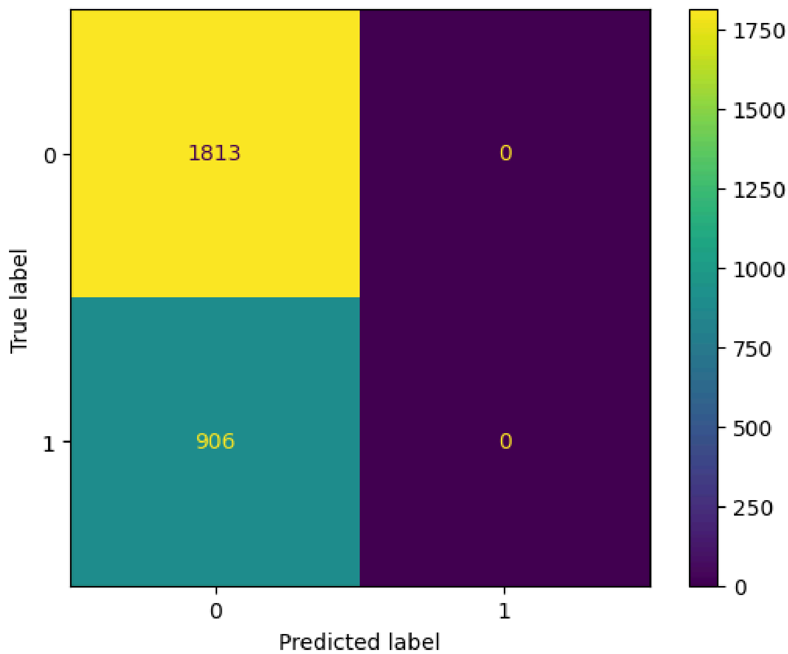


Figure 5. Confusion matrix for Logistic Regression

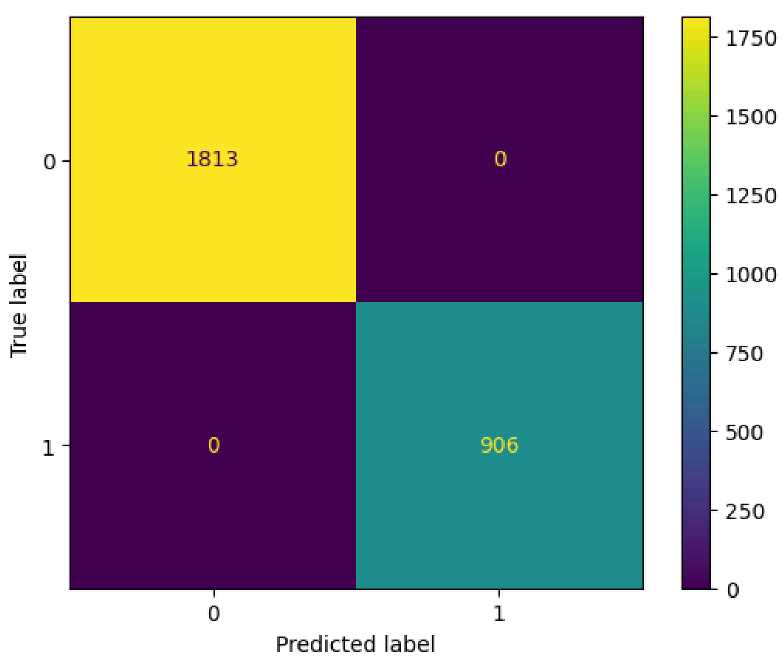


Figure 6. Confusion matrix for Random Forest Classifier (RFC)

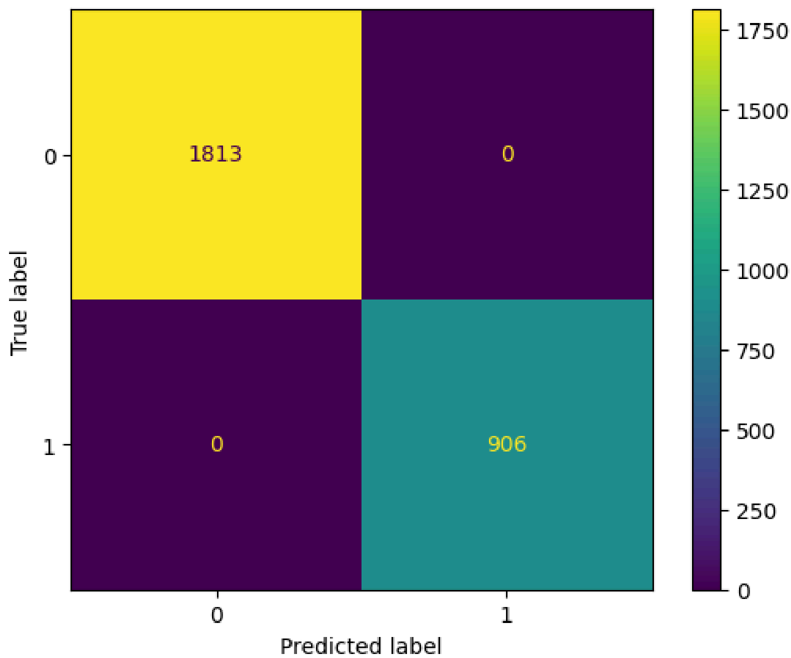


Figure 7: Confusion matrix for Support Vector Machine (SVM)

For KNN, RFC, and SVM models, the confusion matrices exhibited a high number of true positives and true negatives, with few false positives or false negatives. This indicates that these models were effective in both detecting seizures and avoiding false alarms. On the other hand, the Logistic Regression model had a significant number of false negatives, meaning it missed many seizures. This explains its low recall and suggests that the model was unable to effectively identify seizure events due to its inability to handle the highly imbalanced dataset.

For seizure prediction, the LSTM model analyzed temporal patterns in the EEG data to predict seizures before they occurred, achieving an accuracy of 89.26% and a weighted precision and recall of 89%.

Overall, these results indicate that while some models (KNN, RFC, and SVM) performed exceptionally well in terms of accuracy and recall, Logistic Regression struggled with detecting seizures. The confusion matrices were particularly helpful in visualizing these discrepancies and understanding the performance of each model, especially in terms of how well they handled the imbalance between seizure and non-seizure events. The LSTM also showed high accuracy in seizure prediction, highlighting its potential for use in at-home seizure management technology.

Overall, these results indicate that while KNN, RFC, and SVM achieved high accuracy and recall, Logistic Regression struggled with seizure detection. The confusion matrices helped visualize these differences, particularly in handling class imbalance. The LSTM model also demonstrated high accuracy, highlighting its potential for at-home seizure forecasting.

Conclusion

In this research, I developed a model that could detect and predict epileptic seizures using machine learning. For detection, I tested several algorithms: K-Nearest Neighbors (KNN), Logistic Regression, Random Forest Classifier (RFC), and Support Vector Machine (SVM). The results showed that KNN, RFC, and SVM performed exceptionally well, achieving high accuracy and recall. However, Logistic Regression struggled, likely due to the class imbalance, even after adjusting class weights. The LSTM model, designed to recognize patterns in sequential data in order to predict epileptic seizures, also showed promising results.

The model performed so well because of Synthetic Minority Over-Sampling Technique (SMOTE), which helped address the major class imbalance. SMOTE made a big difference in improving recall, helping the models become more sensitive to seizure events instead of favoring the non-seizure class. The strong performance of KNN, RFC, SVM, and LSTM suggests that machine learning has real potential for improving seizure detection.

Moving forward, the next steps for this research would involve exploring additional models, such as Gradient Boosting or more advanced deep learning architectures, to further improve performance. Expanding the dataset to include more seizure events would make the model more reliable and boost its external validity. Experimenting with different preprocessing techniques or incorporating additional EEG features could also refine the model's ability to detect seizure patterns more accurately.

Given how well KNN, RFC, SVM, and LSTM performed, this approach has real potential for being implemented in at-home seizure detection devices. For example, wearable EEG headbands or smart devices could integrate this model to monitor brain activity in real-time, sending alerts before a seizure occurs. This would be especially helpful for people with epilepsy who may not always have immediate access to medical care. Parents of children with epilepsy, for example, could receive warnings on their phones if a seizure is detected, allowing them to intervene quickly. Similarly, individuals living alone could benefit from automated emergency alerts sent to caregivers or medical professionals. With further optimization, this technology could make seizure detection more accessible, giving people with epilepsy more independence and improving their overall safety.

Acknowledgements

I would like to thank Tanish Jain, who mentored me and provided consistent support during this research.

References

Dawson, C. (2024, June 4). *New AI system detects rare epileptic seizures*. USC Viterbi School of Engineering. https://viterbischool.usc.edu/news/2024/06/ai-detects-rare-epilepsy/

Guttag, J. (2010). *CHB-MIT Scalp EEG Database (version 1.0.0)*. PhysioNet. https://doi.org/10.13026/C2K01R

Jefferson, B. (2020, November 14). *Artificial intelligence identifies, locates seizures in real-time - The Source - Washington University in St. Louis*. The Source. https://source.washu.edu/2020/06/artificial-intelligence-identifies-locates-seizures-in-real-time/

Project Overview ‹ Sleep and Epilepsy: Demystifying the thousand-year-long conundrum – MIT Media Lab. (n.d.). MIT Media Lab. https://www.media.mit.edu/projects/sleep-and-epilepsy/overview/

Seizure. (2024, July 18). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/22789-seizure

SUDEP. (2013, August 5). Epilepsy Foundation. Retrieved February 14, 2025, from https://www.epilepsy.com/complications-risks/early-death-sudep

Suen, C. G., Wood, A. J., Burke, J. F., Betjemann, J. P., & Guterman, E. L. (2023). Hospital EEG Capability and Associations With Interhospital Transfer in Status Epilepticus. *Neurology. Clinical practice*, *13*(2), e200143. https://doi.org/10.1212/CPJ.0000000000200143

Tao, J., MD PhD. (n.d.). *New treatment options for people with drug-resistant epilepsy*. https://www.uchicagomedicine.org/forefront/neurosciences-articles/new-treatment-options-for-people-with-drug-resistant-epilepsy

Too many with epilepsy are unaware of this uncommon but fatal threat. (2018, June 5). Chicago Tribune. Retrieved August 29, 2024, from https://www.chicagotribune.com/2017/04/25/too-many-with-epilepsy-are-unaware-of-this-uncommon-but-fatal-threat/

Wolf, G. (2021, March 17). *Does Medicare cover an EEG?* Medicare & Medicare Advantage Info, Help and Enrollment. https://www.medicare.org/articles/does-medicare-cover-an-eeg/

World Health Organization: WHO. (n.d.). *Epilepsy*. https://www.who.int/news-room/fact-sheets/detail/epilepsy#:~:text=People%20with%20epilepsy%20tend%20to,conditions%2C%20including%20anxiety%20and%20depression.