Machine Learning for Predicting Epileptic Seizures Using EEG Signals

Hem Chaudhary

22106178

Progress Report

submitted for

ENGR7028 Engineering Project 1 (PG)

in partial fulfilment of the requirements for the degree of

Master of Engineering

Major: Electrical Engineering

Supervisor: Ali Mehrabi

School of Engineering, Design and Built Environment
Western Sydney University

October 2025

Abstract

Epileptic seizure prediction is an important research area in biomedical engineering. It aims to improve patient safety and quality of life. It works by detecting abnormal brain activity early. This project studies the use of machine learning on EEG signals. It helps detect seizures and predict when they will start. It also reduces the need for manual EEG analysis.

Traditional EEG tests depend on experts. This process is slow and needs a lot of work. It is not suitable for real-time monitoring. This creates a research gap. There is a need for automatic and clear models. These models must find seizure patterns and predict their timing from complex EEG data.

The aim of this project is to design a machine learning framework using Random Forest Classifier and Random Forest Regressor models to detect seizures and estimate their start and end times. The system focuses on achieving reliable performance while maintaining simplicity and interpretability for potential clinical use.

EEG data from the CHB-MIT Scalp EEG Database were preprocessed, segmented, and labelled according to seizure events. Feature extraction techniques were applied to capture relevant time and frequency characteristics of the signals. The Random Forest Classifier was used to distinguish seizure and non-seizure signals, while the Random Forest Regressor predicted seizure onset and termination times. Visualization tools were developed to display actual and predicted seizure intervals.

Preliminary results show that the model achieved 100% accuracy in classifying seizure and non-seizure data. The regression component estimated seizure start and end times with average differences of approximately 350 seconds for training and 800 seconds for testing data. These results indicate that the classification model is highly effective, while timing prediction can be improved through finer signal segmentation.

This study demonstrates the potential of ensemble-based machine learning for EEG-based seizure prediction. Future work will focus on adding more patient data. It will also use epoch-based segmentation to improve timing accuracy. The results will help in creating useful, data-based tools for continuous seizure monitoring. They will also support early treatment in clinical settings.

Table of Contents

List of Tables

List of Figures

1. Introduction

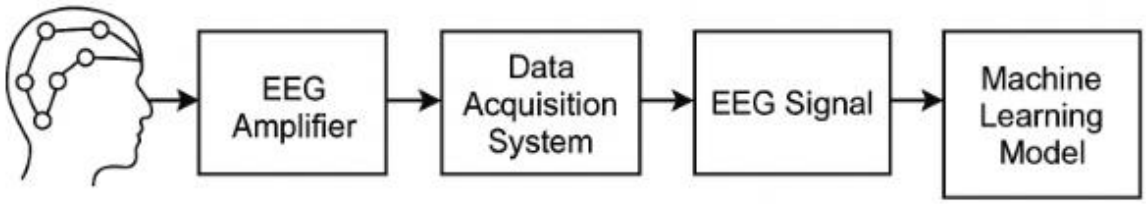


Figure 1: Basic process of EEG acquisition from scalp electrodes for seizure analysis

(source: unknown)

This progress report presents ongoing research on project titled Machine Learning for Predicting Epileptic Seizures Using EEG Signals. The project analyze at how machine learning can be used to predict seizures before it occurs. It helps make life easier for people with epilepsy and improve their quality of life. The main objective of this project is to build a model that can detect seizure patterns in EEG signal and predict the start and end time of a seizure. The project uses machine learning method. It focuses on the Random Forest Classifier and Random Forest Regressor as these models are robust and work well for with medical data. The report shows the current progress, objective, scope, drawback and future work.

Epilepsy is a disorder of brain that causes a sudden and repeated seizures due to the abnormal electrical activities in the brain. The quality of life of a person is highly affected as these seizures are not predictable. Therefore, it is important to predict seizures accurately and timely. Electroencephalogram (EEG) signals uses a non-invasive method to observe the brain activity and has become the basis for seizure detection and forecasting. Due to the advancement in machine learning, it has become easier to automate the analysis of EEG signals, reducing the time consuming manual work and allowing algorithms to find patterns that appear before a seizure begins.

According to the World Health Organization more than 50 million people in the world have epilepsy. It is one of the most common brain problems. Seizures can cause convulsions, loss of consciousness or small changes in behavior depending on the region of the brain that is affected. For many people, seizures happen without warning. This is dangerous and makes life difficult. It also causes stress and loneliness. Predicting seizures is still a big problem in medical and engineering research.

Doctors use EEG to study brain activity. EEG is a safe method that uses small sensors on the head. These sensors record the electrical signals of the brain. Each sensor measures small voltage changes caused by brain cells. These signals show the brain's activity over time. EEG is often used in hospitals to detect and study seizures. It gives clear information about brain activity. But EEG signals are very complex. They often have noise from body movement, blinking or other sources. EEG signals are also different for every person. This makes manual study very hard and slow. It is not easy to use for real-time seizure prediction.

Epilepsy is unpredictable. EEG signals are complex. Because of this, there is a strong need for automatic systems. These systems can study brain signals fast and find seizure patterns without human help. New computer technology has made this possible. Machine learning can find useful features and learn directly from data. It can find small changes that appear before a seizure starts. These changes are too small for humans to see. Machine learning is very good for building automatic seizure prediction systems. It can help in both medical diagnosis and real-time warning systems.

Machine learning models use EEG data that are already labeled as seizure or non-seizure. After training, they can find similar patterns in new data. They can also predict if a seizure is coming. Random Forest models are useful for this work. They use many decision trees to make a strong model. The Random Forest Classifier can separate EEG signals into seizure and non-seizure parts. The Random Forest Regressor can predict the starting time of seizure. These models are simple, accurate and easy to understand. They also work well with noisy and complex data. This makes them good for EEG studies.

Seizure prediction is important for patients and doctors. It can help improve safety and mental health. A good prediction system can be used in wearable devices or implants. These devices can give early warnings or stop seizures. This can reduce hospital visits and help patients live freely. It also helps doctors and nurses by reducing their workload. Seizure prediction needs knowledge from many fields. It joins medical signal study, data science and engineering. It needs teamwork between doctors, engineers and researchers to make real medical tools.

The main aim of this project is to make and test a machine learning system for seizure prediction. The study uses Random Forest models to detect seizures and find when they start. The goals are to clean EEG data, remove noise, extract useful features, train and test the models and check accuracy. It also checks sensitivity and specificity. The project also aims to analyze how feature selection and model parameters affect overall reliability and computational efficiency. The expected outcome is a model that achieves a balance between accuracy, simplicity and interpretability — a crucial factor for clinical use, where transparency of decision-making is essential.

The scope of this project includes the use of publicly available EEG datasets that contain seizure annotations from multiple patients. Due to resource constraints, only a subset of patients is currently being used for model development and testing. Each EEG file consists of multiple channels representing different brain regions, providing rich spatial information for analysis. Data preprocessing includes noise removal, filtering and segmentation of EEG recordings into shorter

epochs suitable for machine learning. The feature extraction process focuses on time-domain and frequency-domain measures that can effectively represent seizure patterns. The study does not include deep learning or complex neural networks at this stage, as the emphasis is on interpretability and efficiency using traditional machine learning methods.

There are several limitations and challenges identified in this research. EEG data varies from one patient to another. This makes it difficult to create one general model for all people. Most EEG datasets have more non-seizure data than seizure data. This imbalance can make the model favor the non-seizure class. Another problem is noise and unwanted signals that look like seizures. These can reduce accuracy if not removed properly. The system must also be fast and accurate enough to work in real time. It is also important that doctors can understand how the model makes its decisions. This helps with trust and clinical use. Reducing these issue is an important part of this project.

In summary, this report explains the purpose, goals and scope of the project Machine Learning for Predicting Epileptic Seizures Using EEG Signals. The study uses Random Forest models to predict seizures with good accuracy. It also focuses on keeping the model simple and easy to understand. The project adds to research in smart medical systems. It shows how basic machine learning can be used for real seizure prediction. The project is still in progress. The work done so far gives a strong base for future testing, model improvement and use with larger datasets. The next chapter presents a comprehensive review of related studies that have applied machine learning to EEG-based seizure detection and prediction.

2. Literature Review

2.1 Evolution of Machine Learning in Seizure Prediction

The literature demonstrates the progressive refinement of these models from simple classifiers trained on handcrafted features to optimized, ensemble-based methods revealing both the potential and limitations of current approaches.

One of the earliest influential works applying machine learning to EEG-based seizure analysis was that of Shoeb and Guttag (2010), who developed a patient-specific classifier for seizure onset detection using scalp EEG from the CHB-MIT database. Their supervised model combined temporal and spectral features with a support vector machine (SVM) classifier, achieving a 96 per cent detection rate and a median delay of only three seconds. The study established that machine learning could outperform rule-based or thresholding methods, while also showing that seizure patterns differ widely between individuals. Although this patient-specific approach improved accuracy, it highlighted the challenge of generalizing models across subjects without personalized calibration.

2.2 Advances in Classifier Design and Feature Extraction

Subsequent research built on this foundation by exploring various classifiers and feature extraction techniques. Ahmed et al. (2020) proposed a model using time-domain and frequency-domain statistical descriptors with an SVM classifier to distinguish seizure and non-seizure EEG segments. Their results confirmed SVM's robustness for high-dimensional biomedical data but also showed dependence on manual feature design. Similarly, Khan et al. (2021) employed EEG from the Temple University Hospital corpus, extracting both time and frequency features such as energy, entropy and power spectral density, then training logistic regression, decision tree, k-nearest neighbor (KNN) and SVM models. The SVM achieved the highest accuracy of 92.7 per cent, but the study assumed stationarity of EEG signals and evaluated only binary classification, thus overlooking early preictal state identification.

Chandel et al. (2022) showed the strong performance of ensemble methods. They used a random forest model on handcrafted features taken from EEG signals. Their method reached 99.96 per cent accuracy and 100 per cent sensitivity. This proved that ensemble learning can find complex patterns in EEG data. However, their dataset had little difference between patients. The authors warned that such high accuracy may not work in real clinical EEG data, which often has noise and artefacts. Tran et al. (2022) also investigated feature-based approaches but focused on dimensionality reduction. Using discrete wavelet transform (DWT) for decomposition and binary particle swarm optimization (BPSO) for feature selection, they reduced 75 per cent of data dimensions while maintaining 98.4 per cent classification accuracy. The optimization improved computational efficiency but increased the model's complexity and risk of overfitting smaller datasets.

2.3 Feature Engineering and Performance Evaluation

Many studies addressed the challenge of extracting meaningful features from EEG's non-stationary nature. Thangarajoo et al. (2021) reviewed wavelet and empirical mode decomposition (EMD) techniques and concluded that non-linear and non-stationary characteristics of EEG signals

necessitate multiresolution analysis. They found that feature extraction based on Stockwell or wavelet transforms, combined with SVM or random forest classifiers, achieved superior performance compared with raw statistical features. This view was echoed by Ahmed et al. (2021), who combined DWT with principal component analysis (PCA) to reduce redundancy before SVM classification, reporting consistent results across benchmark datasets but acknowledging the cost of extensive parameter tuning. Together these studies demonstrate that performance improvements in traditional ML pipelines often depend more on feature engineering than on the choice of classifier.

Several comparative works have assessed the relative performance of standard classifiers. Ly Tran et al. (2022) compared SVM, KNN, naive Bayes and decision tree algorithms using optimized feature sets from EEG recordings, concluding that SVM and random forest consistently provide the best balance between accuracy and generalization. Random forest's ensemble structure offers robustness to noisy data and lower risk of overfitting than single-tree methods, while SVM handles high-dimensional features efficiently. However, both methods need a lot of computation to choose the right parameters and kernels. Because of this, later studies tested hybrid systems that join feature extraction with optimization methods like genetic algorithms or PSO. These help in automatic parameter tuning. Hybrid methods can give better accuracy, but they are harder to understand. This is a problem for medical use, where clear explanation is important.

Many studies looked at how testing methods affect results. Thangarajoo et al. (2021) noted that many works rely on patient-based cross-validation, which can make accuracy appear higher than it really is compared to testing on new patients. They showed that models trained on a few people often do not work well for new patients because EEG signals are different for each person. Mourad et al. (2025) said it is hard to get the same results on different datasets because testing methods and data cleaning steps are not the same. Rasheed et al. (2021) also showed that removing noise and choosing channels can change results a lot. Small changes in data cleaning can make big differences in accuracy between studies.

2.4 EEG Data Quality and Variability

Basheer Ahmed et al. (2022) made an important study on child epilepsy using EEG data from the CHB-MIT database. They found that EEG signals from children show more changes because their brains are still developing. This makes seizure prediction more difficult. Their review of machine learning approaches identified DWT, independent component analysis and entropy-based measures as the most effective features for childhood seizure detection. Combining multiple classifiers through ensemble learning enhanced robustness, yet computational cost and data imbalance remained major issues. The authors noted that small sample sizes in paediatric datasets hinder cross-subject generalization and recommended more balanced data acquisition strategies.

While most traditional approaches rely on pre-defined features, some research sought to enhance feature diversity. Garima Chandel et al. (2022) used statistical, temporal and shape-based features of EEG signals with a random forest model. They showed that model performance depends greatly on removing noise during preprocessing. They also highlighted the importance of using the same electrode positions and sampling rates to prevent differences between datasets. Their results were similar to those of Hassan et al. (2022), who used handcrafted features with machine learning models like SVM and logistic regression to make the system more reliable. Though their hybrid CNN-ML architecture primarily belonged to deep learning, the inclusion of classical classifiers showed that machine learning methods remain relevant when computational resources or data size limit the feasibility of fully deep models.

The studies collectively show that feature engineering and classifier optimization remain central to seizure prediction accuracy. Mourad et al. (2025) carried out a large study that reviewed both machine learning and deep learning methods. They found that traditional machine learning models, when used with good feature extraction, can still perform as well as deep learning on medium-sized datasets. They also noted that simple models like logistic regression and SVM are easier to understand, which makes them more suitable for medical use. However, the review also underlined issues of reproducibility, as many published models use private datasets or inconsistent evaluation metrics, preventing direct comparison. Similar concerns were raised by Usman et al. (2017), who examined machine learning methods for seizure detection using spectral entropy and wavelet energy features, finding substantial variation in reported accuracies due to dataset choice and evaluation protocol.

2.5 Dataset Standardization and Preprocessing Importance

The importance of dataset quality recurs throughout the literature. Shoeb and Guttag (2010) and Basheer Ahmed et al. (2022) used the CHB-MIT dataset, while other researchers worked with the Bonn or TUH datasets. Differences in sampling rate, electrode position and labeling style make models trained on one dataset perform poorly on another. Rasheed et al. (2021) explained that EEG signals often have noise caused by muscle movement and eye blinks. Because of this, preprocessing steps like filtering and removing artefacts are very important before feature extraction. Without standard rules, the accuracy reported in studies may not show real clinical performance. This challenge explains the persistent gap between laboratory results and practical deployment of machine learning-based seizure prediction systems.

A number of studies investigated ensemble learning as a means to improve reliability. Chandel et al. (2022) and Tran et al. (2022) found better results when combining several simple classifiers instead of using one model alone. Methods like random forest, gradient boosting and majority voting improved sensitivity and specificity, especially for mixed EEG data. But ensemble models need careful control of how each classifier is weighted. If they become too complex, they are hard to understand. In medical use, where clear explanation is important, this can reduce doctors' trust. Mourad et al. (2025) suggested keeping a balance between model clarity and accuracy, which many studies ignore when focusing only on performance.

2.6 Limitations of Traditional Approaches

Deep learning has become popular, but many researchers said it is less useful in real medical settings because it needs a lot of data and computer power. Usman et al. (2020) built a convolutional neural network for seizure prediction that reached over 90 per cent sensitivity but required large labeled datasets and strong hardware. Traditional machine learning works well even with small datasets and trains faster. Tran et al. (2022) and Hassan et al. (2022) showed that classical models in hybrid systems give good performance and use less computing power. These findings support using traditional machine learning for projects that have limited data or low processing capacity.

Despite their success, traditional approaches face several limitations that collectively define the research gap motivating ongoing work. Most models assume that EEG signals are stationary within analysis windows, yet brain dynamics fluctuate over time. Feature extraction methods such as DWT or EMD partially mitigate this issue but do not capture long-range temporal dependencies. In addition, patient variability means that a model trained on one subject rarely generalizes to others without retraining, as noted by Shoeb and Guttag (2010) and Basheer Ahmed et al. (2022).

abelled EEG data are hard to get. Only expert doctors can mark them. This makes the datasets small. Small datasets can cause bias in the results. Different studies also use different testing methods. Some show only accuracy. Others use sensitivity, specificity or F-score. This makes it hard to compare results. Mourad et al. (2025) and Thangarajoo et al. (2021) said there should be one common testing method. It will help make the results easy to repeat and compare.

2.7 Summary and Research Gap

Machine learning is a strong tool for studying EEG signals. It helps to predict epileptic seizures. SVM and random forest models give good results. They can handle complex and large data. Wavelet and EMD transforms help to extract good features. PSO and PCA make the model faster and more efficient. But real-world use is still hard. Data changes from one patient to another. Computation takes time. Some models are hard to understand. Traditional machine learning is easier to explain. It also works better for small and patient-specific data. So it is still very useful. Most studies agree that better preprocessing and one standard testing method are needed. Models must also be more explainable before they can be used in hospitals.

Research shows big progress in using machine learning for seizure prediction. Early work by Shoeb and Guttag (2010) showed that it is possible to detect seizures automatically. Later work by Ahmed et al. (2020), Khan et al. (2021) and Chandel et al. (2022) improved results. They used better features and combined models. Reviews by Thangarajoo et al. (2021) and Mourad et al. (2025) showed both progress and problems. Datasets are still different. There is no fixed testing standard. Models do not always work well for all people. The results support using traditional machine learning for this project. It is simple, fast and clear. It also works well with small biomedical datasets. This project will improve seizure prediction by choosing better features and designing better models. It will also keep the model easy to understand for real medical use.

3. Aim and Objectives

3.1 Aim

The aim of this project is to build a machine learning model in Python using the Random Forest Classifier and Random Forest Regressor. The system will detect whether an EEG signal contains an epileptic seizure using the classifier and will predict the start and end time of the seizure using the regressor if a seizure is detected.

3.2 Objectives

a. To design a machine learning framework capable of analyzing EEG data for seizure detection and prediction.
b. To utilize EEG recordings from the CHB–MIT Scalp EEG Database for model training and testing.
c. To apply the Random Forest Classifier for identifying seizure and non-seizure patterns in EEG signals.
d. To implement the Random Forest Regressor for predicting seizure onset and offset times when a seizure is detected.
e. To label the data files as “seizure” or “non-seizure” based on description and split them accordingly for training and testing.
f. To evaluate model performance using metrics such as accuracy and mean absolute error (MAE).
g. To visualize the prediction results by comparing actual and predicted seizure intervals in graphical form.
h. To identify limitations of the current model, including inter-patient variability and small training datasets and propose future improvements.

4. Research Methodology

The research methodology describes the step-by-step approach adopted to design, develop and evaluate a machine learning model capable of detecting epileptic seizures and predicting their onset and end times using EEG signals. The methodology follows a systematic workflow consisting of data collection, preprocessing, feature extraction, model training, classification, regression and visualization, as illustrated in **Figure 2**.

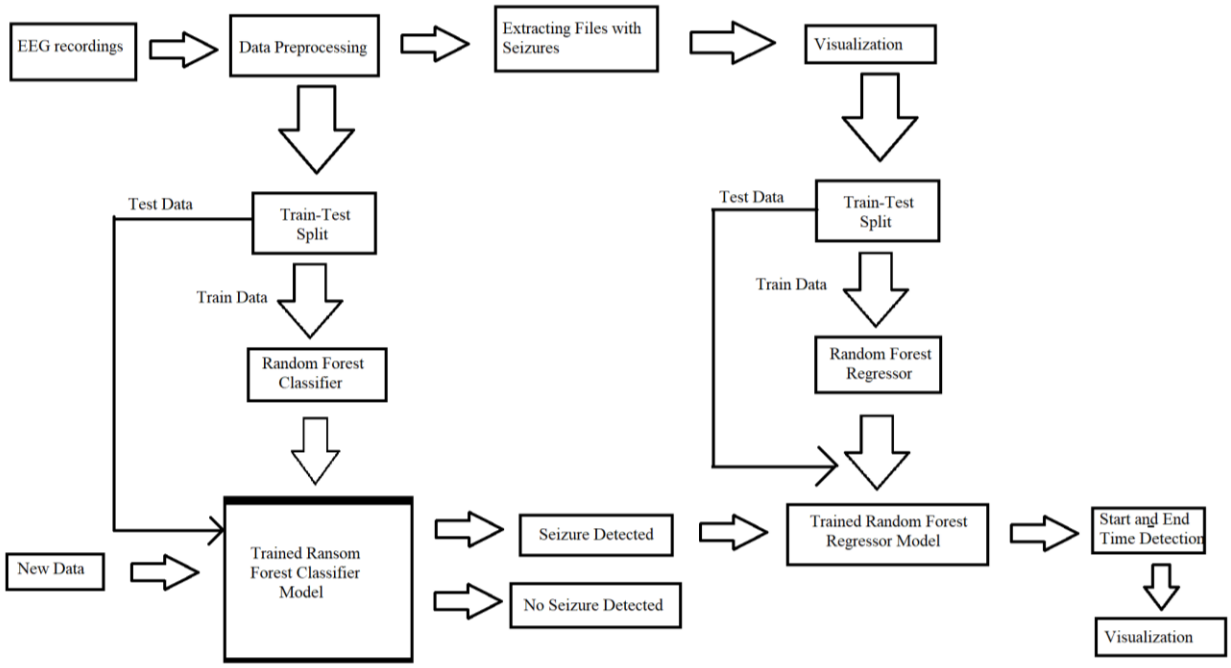


Figure 2. Workflow of the proposed machine learning model for seizure detection and prediction using Random Forest Classifier and Regressor

4.1 Data Acquisition

The EEG data used in this project were obtained from the CHB–MIT Scalp EEG Database, a publicly available dataset hosted on PhysioNet. This dataset contains long-term scalp EEG recordings from paediatric patients with intractable seizures, collected at the Children's Hospital Boston. Each EEG file has a duration of approximately one hour and contains 23 channels corresponding to electrode placements on the scalp, following the International 10–20 systems.

Due to computational and resource constraints, data from 3 patients were selected for training and testing the model. Each patient's data include both seizure and non-seizure segments, allowing for the classification of EEG signals and prediction of seizure timing.

4.2 Data Preprocessing

The preprocessing stage focused on preparing the EEG data for analysis and model training. All tasks were performed in Google Colab, which provided a cloud-based Python environment with direct integration to Google Drive for data storage and access. EEG recordings of three patients

were selected from the CHB–MIT Scalp EEG Database and uploaded to Google Drive for processing.

The following steps were carried out during data preprocessing:

a. Data Import and Access:

The EEG files were stored in Google Drive and imported into the Google Colab environment using the glob library. This allowed automatic retrieval of all .edf files from the specified folder containing patient recordings.

b. Reading Raw EEG Files:

Each .edf file was read using the MNE-Python library. The library extracted the EEG signals and important details such as sampling frequency, number of channels and recording duration. The signals were then converted into numerical arrays for further processing.

c. Dictionary Creation:

A Python dictionary was created manually to store seizure information. For each file that contained a seizure, the dictionary included the file name, seizure start time and seizure end time. This information was taken from the annotations provided in the CHB–MIT dataset.

d. Labeling of Data:

Each EEG file was labeled with the help of dictionary. Filename found inside dictionary was labeled as "Seizure Detected" and those not in dictionary was labeled as "No Seizure Detected".

e. Signal Cropping:

For EEG files that showed seizure activity, only the part of the signal between the start and end times was kept. The other parts of the signal containing no seizure were removed. This made sure that the data used for training included only seizure-related information.

4.3 Feature Extraction:

After labeling the EEG data, feature extraction was done and the signals were converted into meaningful numerical values. These features showed a specific statistical property of the EEG signal that helped the model figure out difference between seizure and non-seizure activity. The extraction process was done in Python using the NumPy and SciPy libraries.

Some statistical features were calculated from each EEG channel as follows:

Mean: Represents the average amplitude of the EEG signal segment. It provides information about the central tendency of brain activity over the analyzed time window.

$$\mu = \frac{1}{N} \sum_{t=1}^{N} x_i(t)$$

Variance: Shows how far the signal values are spread around the mean. It gives an idea of the signal's energy and how much it fluctuates in intensity.

$$\sigma^2 = \frac{1}{N}\sum_{t=1}^{N}(x_i(t) - \mu)^2$$

Standard Deviation: Measures how much the EEG signal changes from its average value. A higher standard deviation shows more irregular or seizure-like activity.

$$\sigma = \sqrt{\sigma^2}$$

Peak-to-Peak Value: Represents the distance between highest and lowest points of the signal. It shows the range of signal and helps identify sudden changes that may occur during seizures.

$$P_{TP} = \max(x_i(t)) - \min(x_i(t))$$

Maximum: The highest signal value in an EEG segment. It often appears as a spike during seizure activity.

Minimum: The lowest signal value in the EEG segment. It works together with the maximum value to show the signal's overall range.

Root Mean Square: Provides an estimate of the effective signal magnitude, taking both amplitude and duration into account.

$$RMS = \sqrt{\frac{1}{N}\sum_{t=1}^{N} x_i(t)^2}$$

Power: Calculated as the mean of squared amplitudes, it measures the overall signal strength and is closely related to the energy content of the EEG waveform.

$$P = \frac{1}{N}\sum_{t=1}^{N} x_i(t)^2$$

Skewness: Indicates the asymmetry of the EEG signal distribution. A skewed waveform may represent abnormal neural activity associated with seizure onset.

$$S = \frac{1}{N}\sum_{t=1}^{N}\left(\frac{x_i(t) - \mu}{\sigma}\right)^3$$

Kurtosis: Measures the "peak" of the signal distribution. High kurtosis values often reflect sharp spikes typical of epileptic discharges.

$$K = \frac{1}{N} \sum_{t=1}^{N} \left(\frac{x_i(t) - \mu}{\sigma} \right)^4$$

Absolute Differential Mean: Computed as the average of absolute differences between consecutive data points. It captures rapid changes and irregularities in the EEG waveform that occur during seizure transitions.

$$AD = \frac{1}{N-1} \sum_{t=2}^{N} |x_i(t) - x_i(t-1)|$$

Argmin / Argmax: Represent the positions (indices) of the minimum and maximum amplitudes in the signal. These help identify where critical changes or spikes occur within each segment.

Each of these statistical descriptors was calculated along the time axis of the EEG channels and stored as feature vectors. The extracted features formed the input dataset for training the Random Forest Classifier and Random Forest Regressor models. This feature-based representation allowed the models to learn discriminative patterns between seizure and non-seizure signals effectively while maintaining computational efficiency.

4.4 Data Splitting

After preprocessing and feature extraction, the labeled 113 EEG dataset was divided into two subsets: 90% of the data was used for training purpose while 10% of the data was separated for testing purpose. This ratio ensured that the machine learning models were exposed to a sufficient amount of data for learning while retaining an independent portion for unbiased performance evaluation.

The training set was used to develop both the Random Forest Classifier and Random Forest Regressor. During training, the models learned the statistical relationships between extracted features and their corresponding labels—classifying seizure and non-seizure segments and estimating seizure onset and offset times.

The testing set consisted of data that were not used during model training. This separation allowed for an objective assessment of the models' predictive accuracy, generalization capability and overall reliability.

4.5 Random Forest Classifier

The Random Forest Classifier is the main model used to detect seizures in EEG data. It builds many decision trees. Each tree learns from a random part of the data and features. Each tree makes its own prediction as seizure or non-seizure. The final answer comes from the most common vote among all trees. This method reduces overfitting and improves accuracy. It also works well with noisy or changing EEG signals.

This model was chosen because it is strong, clear and effective with small medical datasets. EEG data are complex and change from one patient to another. Random Forest can find these patterns easily without assuming any fixed data shape. Compared to SVM or logistic regression, it handles large data better. It also shows which EEG features are most important for detecting seizures.

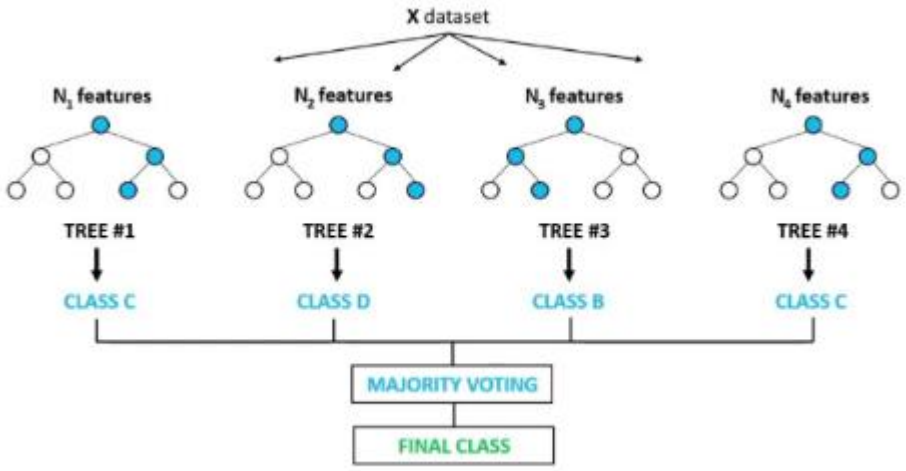


Figure 3: Random Forest Classifier (Source: freecodecamp.org)

4.6 Random Forest Regressor

The Random Forest Regressor is used after a seizure is detected. It helps to estimate the start and end time of the seizure. It also builds many decision trees. But instead of class labels, each tree gives a number that shows the seizure start or end time. The final result is the average of all tree predictions. This gives a balanced and accurate output.

This model was selected because it can learn complex links between EEG features and seizure timing. It works well with small or uneven datasets. It also stays reliable even when the data have noise or unwanted signals. The model's combined structure reduces the effect of such problems. This makes it suitable for real medical use, where EEG data are not always clean.

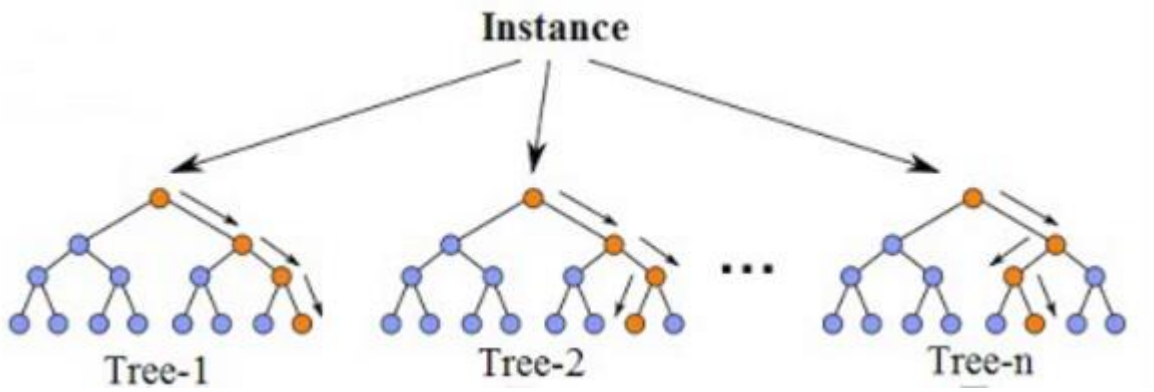
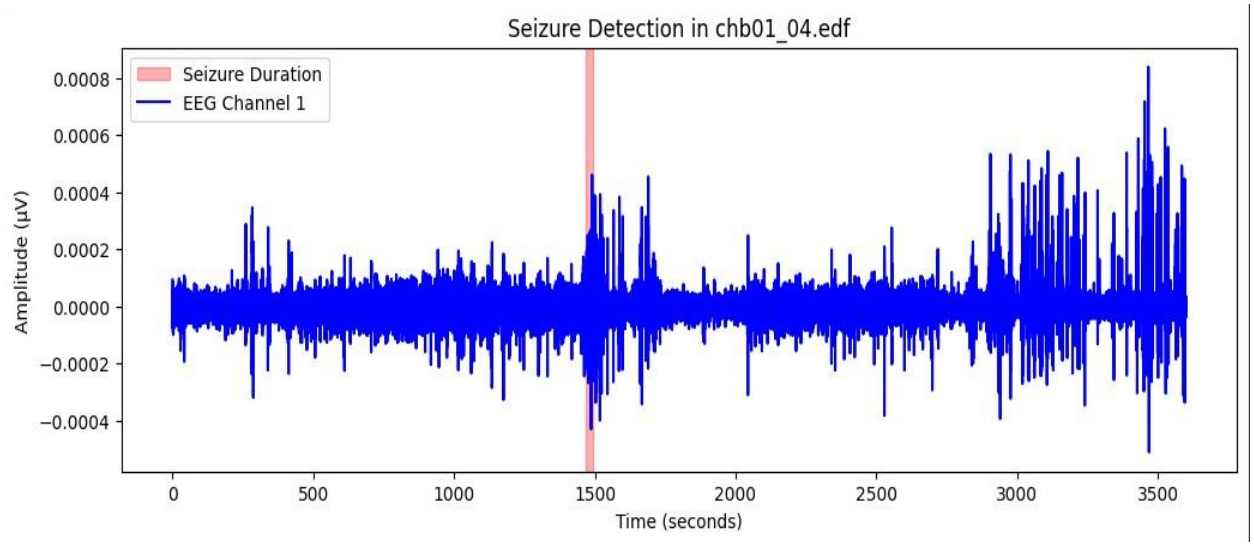
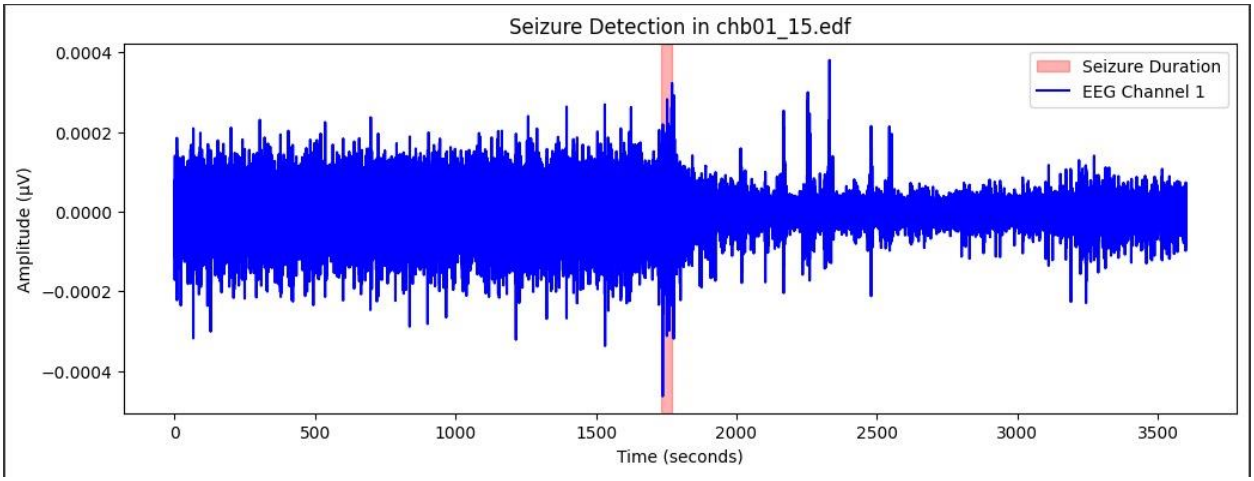


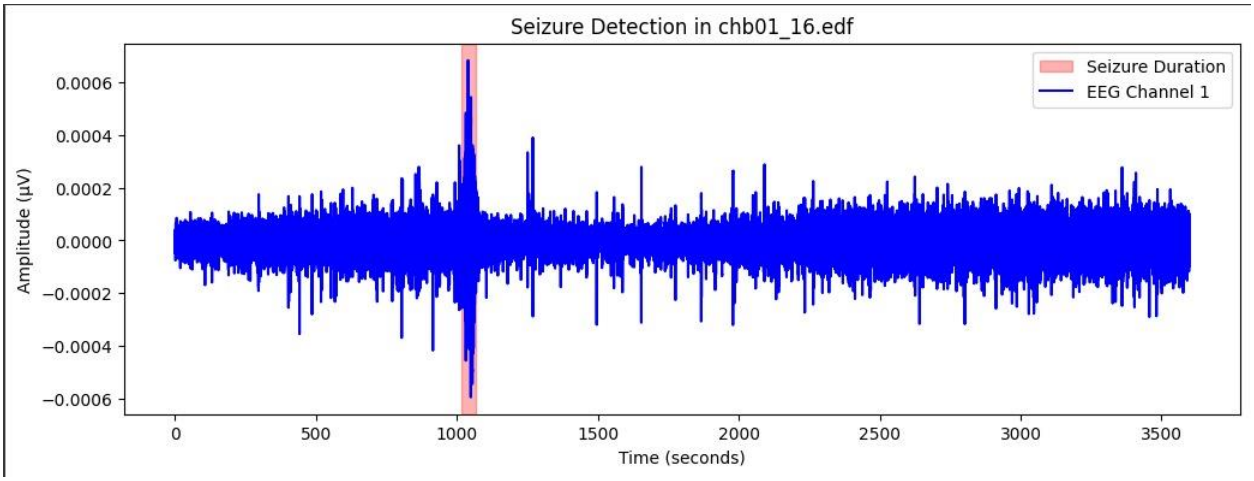
Figure 4: Random Forest Regressor (Source: neptune.ai)

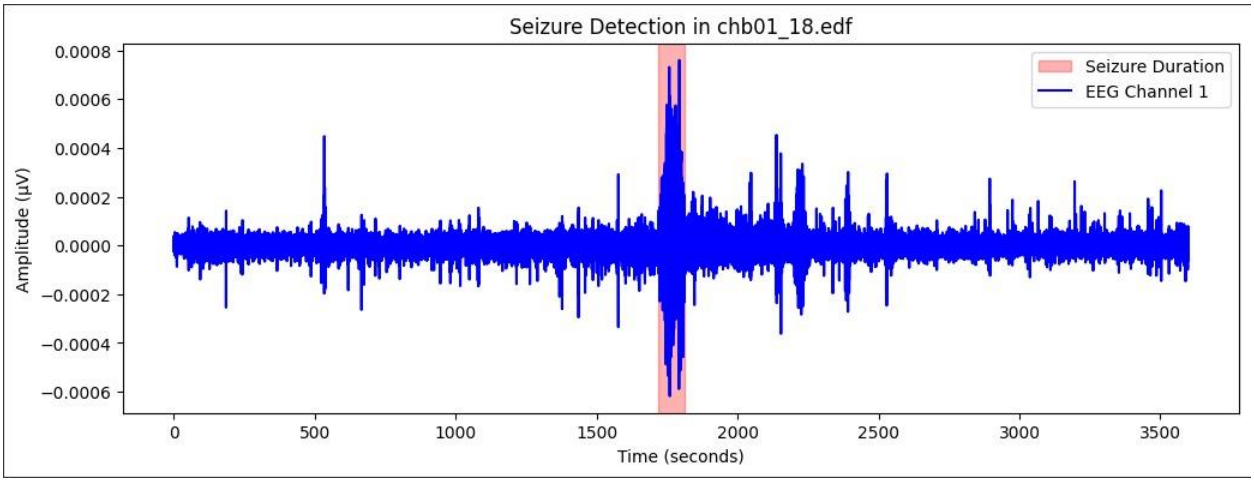
4.7 Visualization

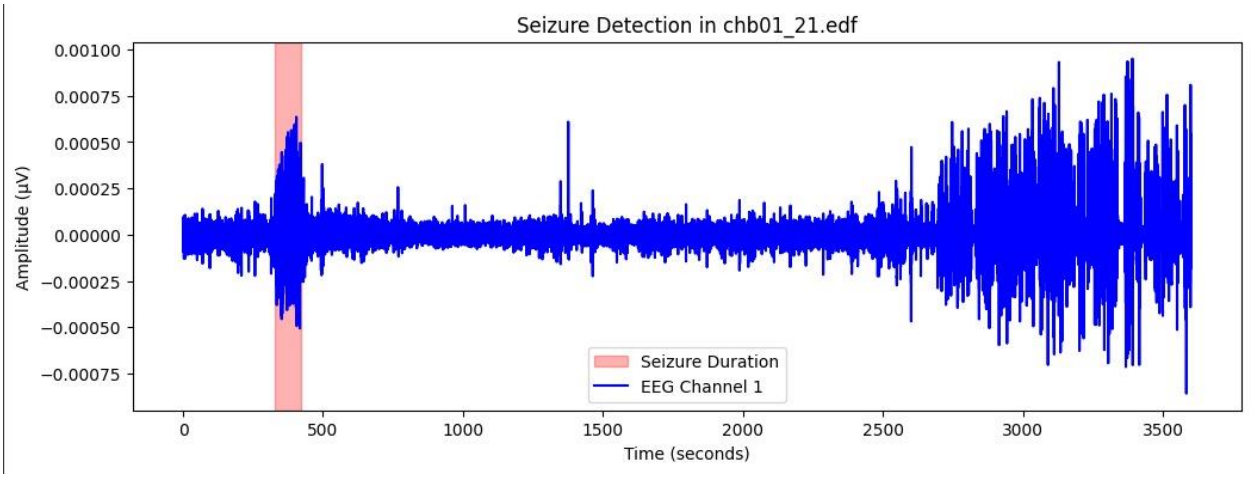
The EEG signal plots for Patient 1, Patient 2 and Patient 3 are shown below. These plots were made using the seizure dictionary. The dictionary stores the file names and the seizure start and end times marked by doctors for each patient. Using this data, the related EEG parts were selected and shown to display the seizure periods in each recording. This step helps to check visually that the seizure times are correct and match the real EEG signals before starting model training and prediction.






Seizure Detection in chb01_16.edf
Seizure Duration
EEG Channel 1
Amplitude (μV)
0.0006
0.0004
0.0002
0.0000
−0.0002
−0.0004
−0.0006
0
500
1000
1500
2000
2500
3000
3500
Time (seconds)


Seizure Detection in chb01_18.edf
Seizure Duration
EEG Channel 1
Amplitude (μV)
0.0008
0.0006
0.0004
0.0002
0.0000
−0.0002
−0.0004
−0.0006
0
500
1000
1500
2000
2500
3000
3500
Time (seconds)


Seizure Detection in chb01_21.edf
Seizure Duration
EEG Channel 1
Amplitude (μV)
0.00100
0.00075
0.00050
0.00025
0.00000
−0.00025
−0.00050
−0.00075
0
500
1000
1500
2000
2500
3000
3500
Time (seconds)

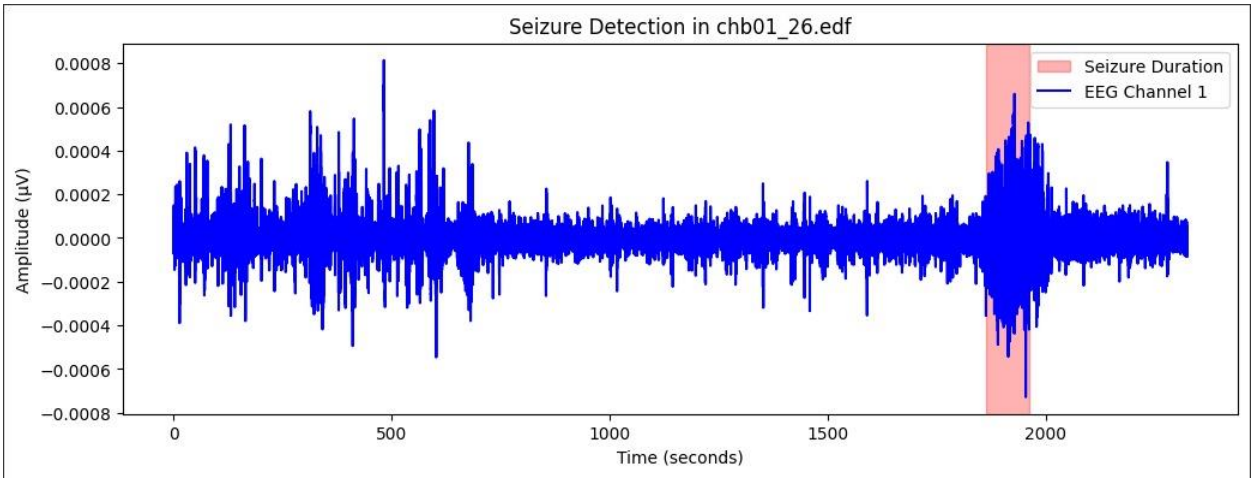


Figure 5: Visualization of EEG signals for Patients 1 showing seizure intervals based on dictionary annotations.

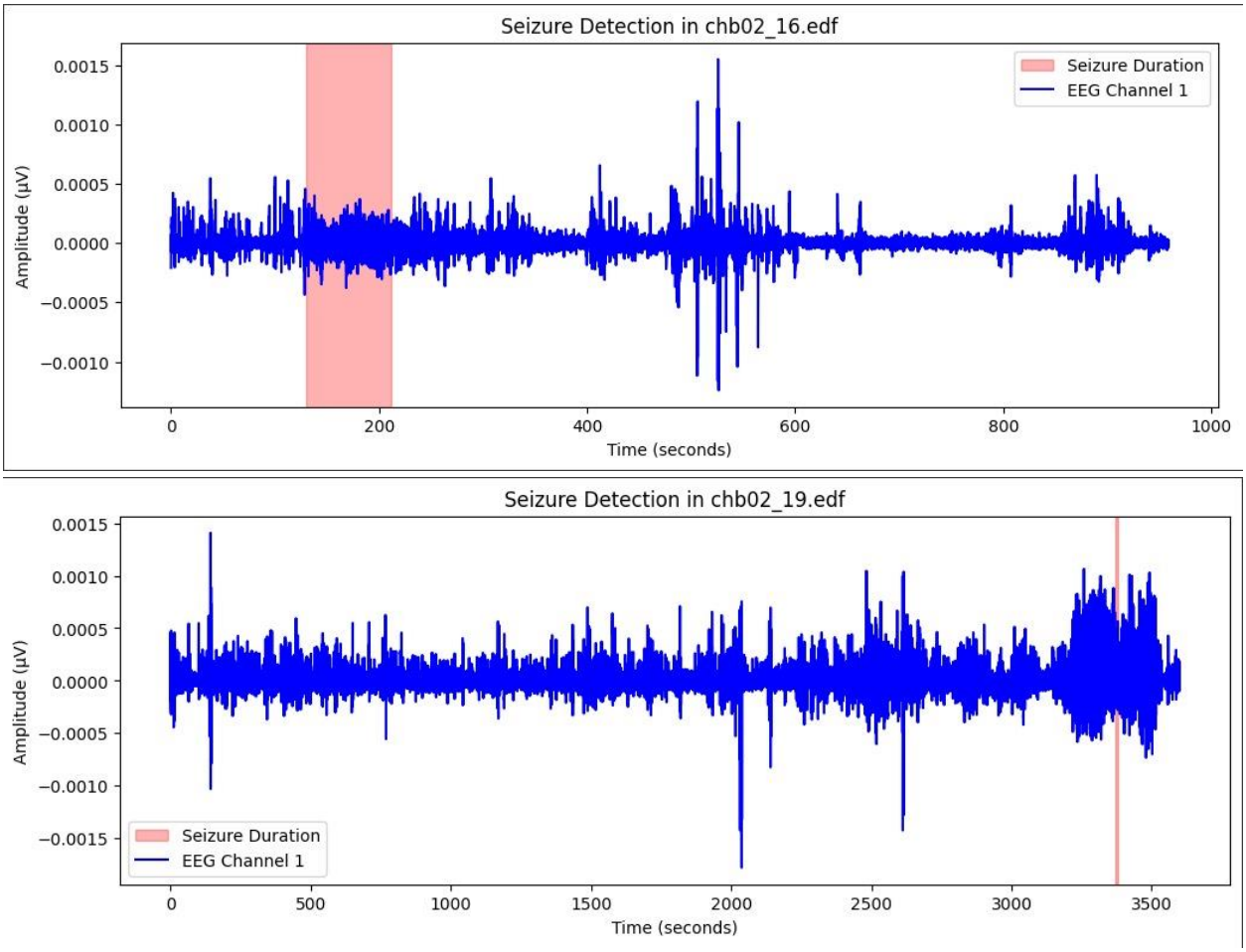
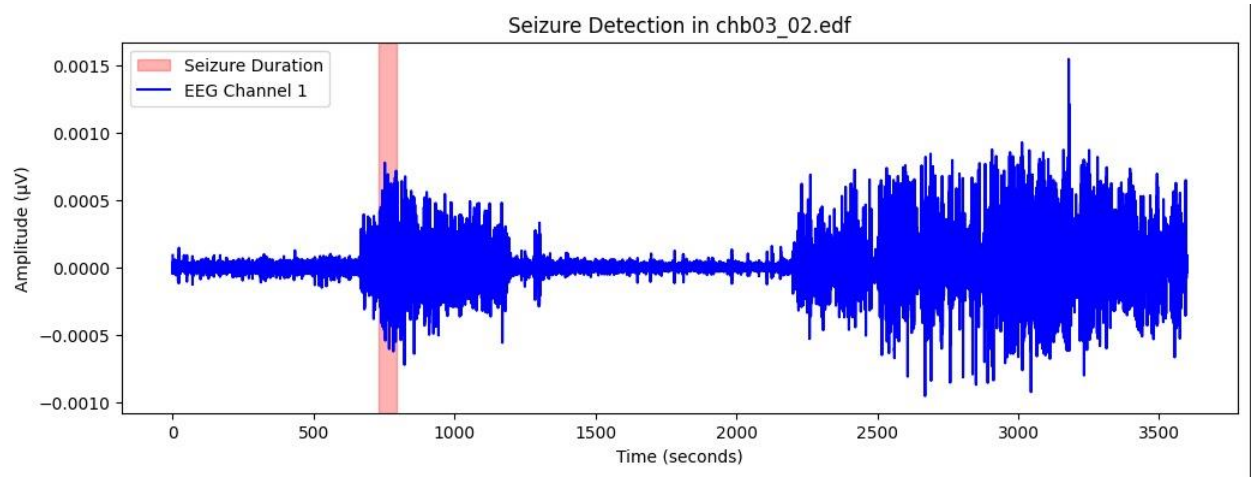
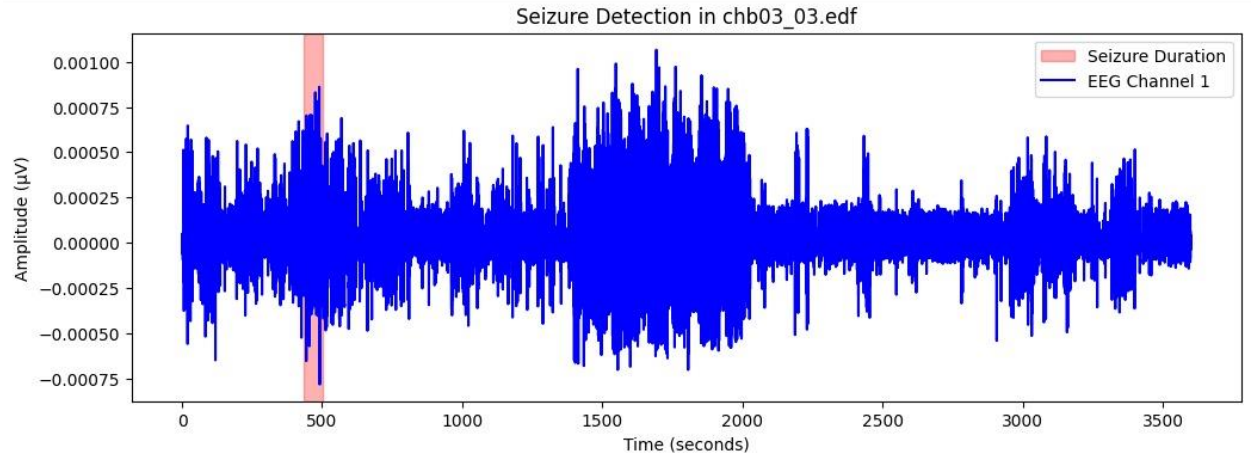
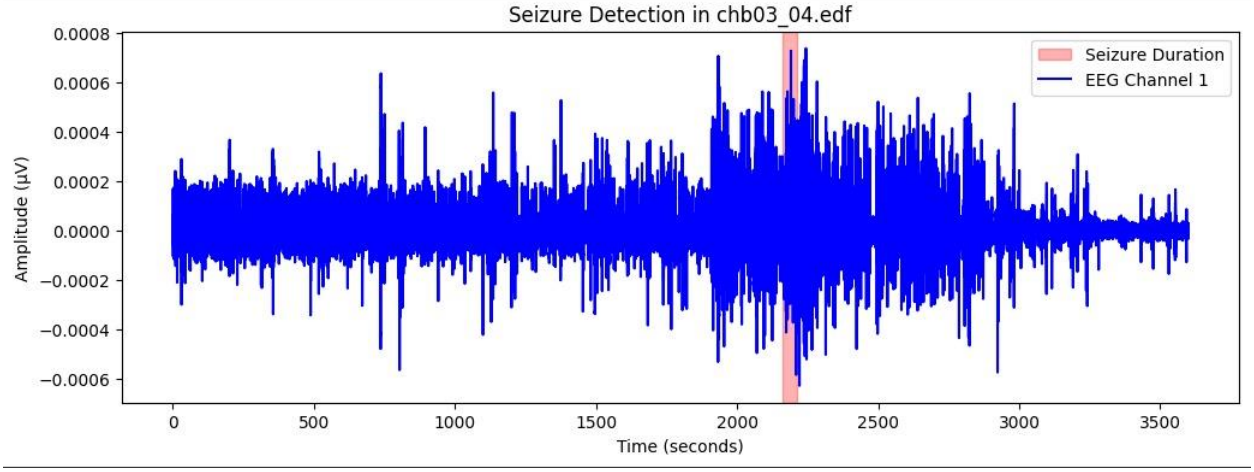


Figure 6: Visualization of EEG signals for Patients 2 showing seizure intervals based on dictionary annotations.


Seizure Detection in chb03_02.edf
Seizure Duration
EEG Channel 1
Amplitude (μV)
0.0015
0.0010
0.0005
0.0000
−0.0005
−0.0010
0
500
1000
1500
2000
2500
3000
3500
Time (seconds)


Seizure Detection in chb03_03.edf
Seizure Duration
EEG Channel 1
Amplitude (μV)
0.00100
0.00075
0.00050
0.00025
0.00000
−0.00025
−0.00050
−0.00075
0
500
1000
1500
2000
2500
3000
3500
Time (seconds)


Seizure Detection in chb03_04.edf
Seizure Duration
EEG Channel 1
Amplitude (μV)
0.0008
0.0006
0.0004
0.0002
0.0000
−0.0002
−0.0004
−0.0006
0
500
1000
1500
2000
2500
3000
3500
Time (seconds)

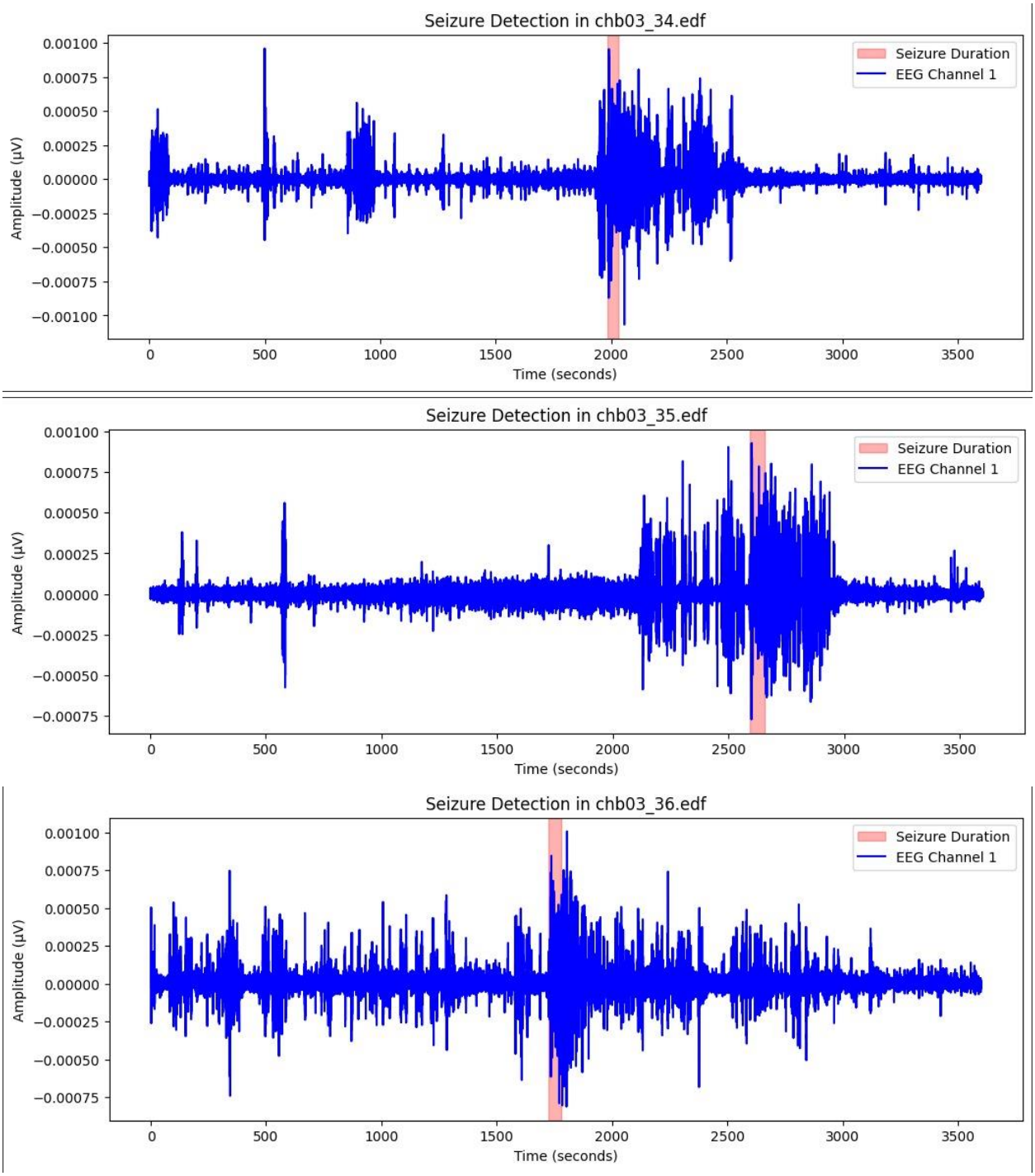


Figure 7: Visualization of EEG signals for Patients 3 showing seizure intervals based on dictionary annotations.

5. Preliminary Result from Preliminary studies

5.1 Training Results – Seizure Detection

The Random Forest Classifier showed a strong performance during the training phase. As shown in Table 1, the model achieved an overall training accuracy of 100%, correctly identifying all seizure and non-seizure EEG recordings among the 103 training files. Each file labelled as Seizure Detected matched exactly with the medical records and there were no false negatives or false positives found in the classification results. This perfect accuracy of classification reflects the model's strong ability to differentiate between normal and seizure EEG patterns after feature extraction and preprocessing.

Training Result
Result accuracy= 100.0

	File Name	Actual Result	Predicted Result
0	chb03_35.edf	Seizure Detected	Seizure Detected
1	chb01_21.edf	Seizure Detected	Seizure Detected
2	chb02_03.edf	No Seizure Detected	No Seizure Detected
3	chb03_14.edf	No Seizure Detected	No Seizure Detected
4	chb01_07.edf	No Seizure Detected	No Seizure Detected
...	...	...	...
98	chb03_06.edf	No Seizure Detected	No Seizure Detected
99	chb01_32.edf	No Seizure Detected	No Seizure Detected
100	chb02_22.edf	No Seizure Detected	No Seizure Detected
101	chb03_05.edf	No Seizure Detected	No Seizure Detected
102	chb02_21.edf	No Seizure Detected	No Seizure Detected

103 rows × 3 columns

Table 1: Training results showing actual vs predicted seizure detection outcomes.

5.2 Training Results – Seizure Start and End-Time Prediction

For the regression phase, the Random Forest Regressor was applied to seizure-containing files to estimate the onset and offset times. As illustrated in Table 2, the model achieved an average start-time difference of 349.91 seconds and an average end-time difference of 342.67 seconds between predicted and actual seizure durations. This indicates that the predicted seizure intervals were around 5–6 minutes of the actual recorded times. Among all recordings, the maximum deviation in seizure onset was approximately 848 seconds, while the smallest difference was less than 50 seconds.

The results show that the model successfully captured seizure timing trends but that minor temporal deviations can occur due to variations in EEG morphology and the limited number of training samples.

Training Result
Start accuracy= 349.91 seconds
End accuracy= 342.67 seconds

	File Name	Actual start	Predicted start	Start Difference	Actual end	Predicted end	End Difference
0	chb01_21.edf	327	925.75	598.75	420	1014.91	594.91
1	chb03_35.edf	2592	2133.56	458.44	2656	2207.00	449.00
2	chb03_34.edf	1982	1853.95	128.05	2029	1879.74	149.26
3	chb03_03.edf	432	914.35	482.35	501	966.42	465.42
4	chb02_19.edf	3369	2520.12	848.88	3378	2549.01	828.99
5	chb03_01.edf	362	864.92	502.92	414	871.36	457.36
6	chb01_15.edf	1732	1610.57	121.43	1772	1660.48	111.52
7	chb03_36.edf	1725	1839.10	114.10	1778	1891.92	113.92
8	chb01_18.edf	1720	1767.80	47.80	1810	1835.97	25.97
9	chb02_16.edf	130	735.57	605.57	212	792.80	580.80
10	chb03_02.edf	731	1051.20	320.20	796	1131.58	335.58
11	chb03_04.edf	2162	1980.44	181.56	2214	2040.41	173.59
12	chb01_16.edf	1015	1192.70	177.70	1066	1253.12	187.12
13	chb01_26.edf	1862	1551.08	310.92	1963	1639.10	323.90

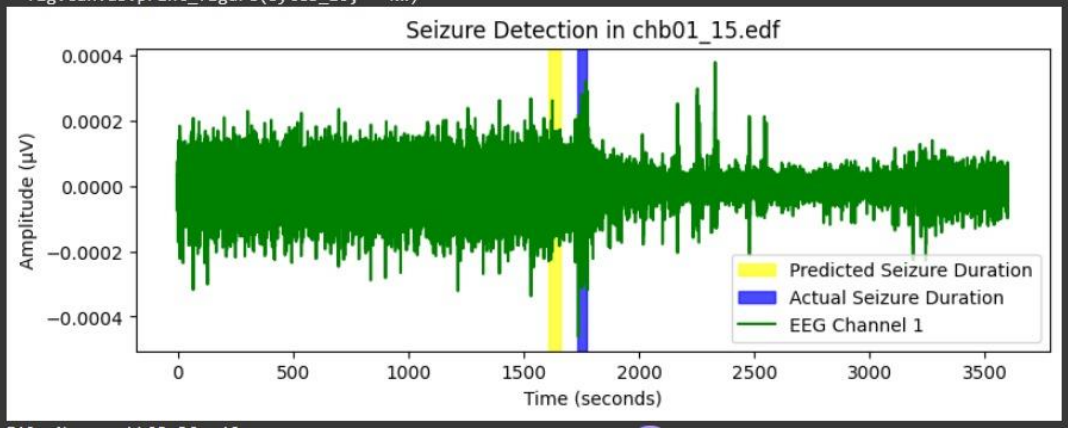
Table 2: Table showing actual vs predicted seizure start and end times with average and maximum differences.

5.3 Graphical Visualization of Training Results

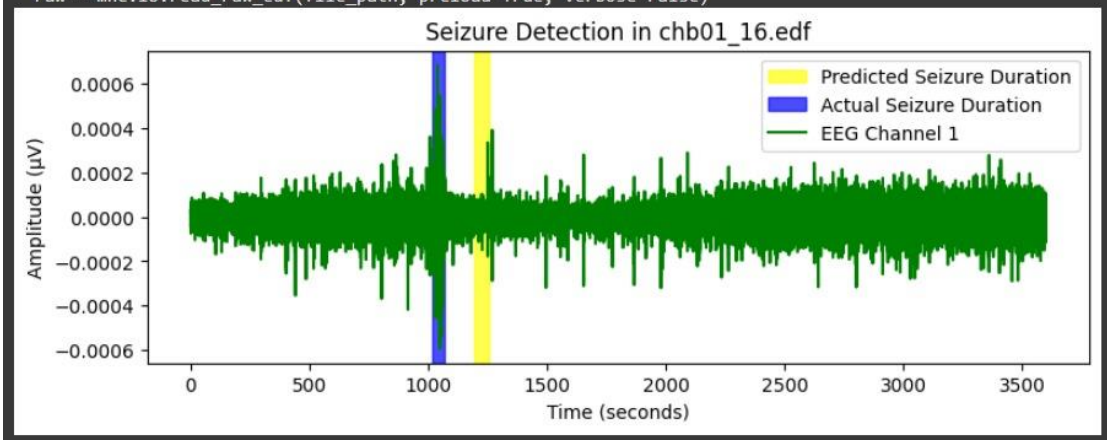
Graphical comparisons were produced for each patient to visualize the predicted and actual seizure intervals. As shown in Figures 8, 9 and 10, the blue region represents the clinically annotated seizure duration, while the yellow region denotes the model's predicted duration. The green waveform indicates the raw EEG signal amplitude over time. The figures reveal that the predicted seizure periods align closely with the actual seizure events across all three patients.

Although small shifts in start and end boundaries are visible in some cases, the model consistently marks the correct seizure segments, confirming that the Random Forest approach generalizes well to different patients' EEG profiles.

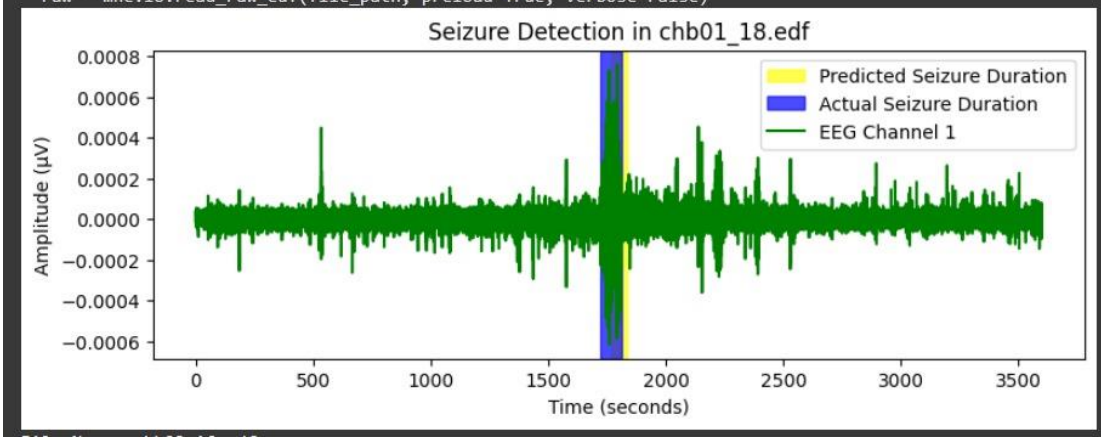
```
Predicted Seizure Start Time: 1610.57 seconds
Predicted Seizure End Time: 1660.48 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}. A
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
/usr/local/lib/python3.12/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Creating legend with loc="best"
  fig.canvas.print_figure(bytes_io, **kw)
```




```
File Name: chb01_16.edf
Predicted Seizure Start Time: 1192.7 seconds
Predicted Seizure End Time: 1253.12 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```




```
File Name: chb01_18.edf
Predicted Seizure Start Time: 1767.8 seconds
Predicted Seizure End Time: 1835.97 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```



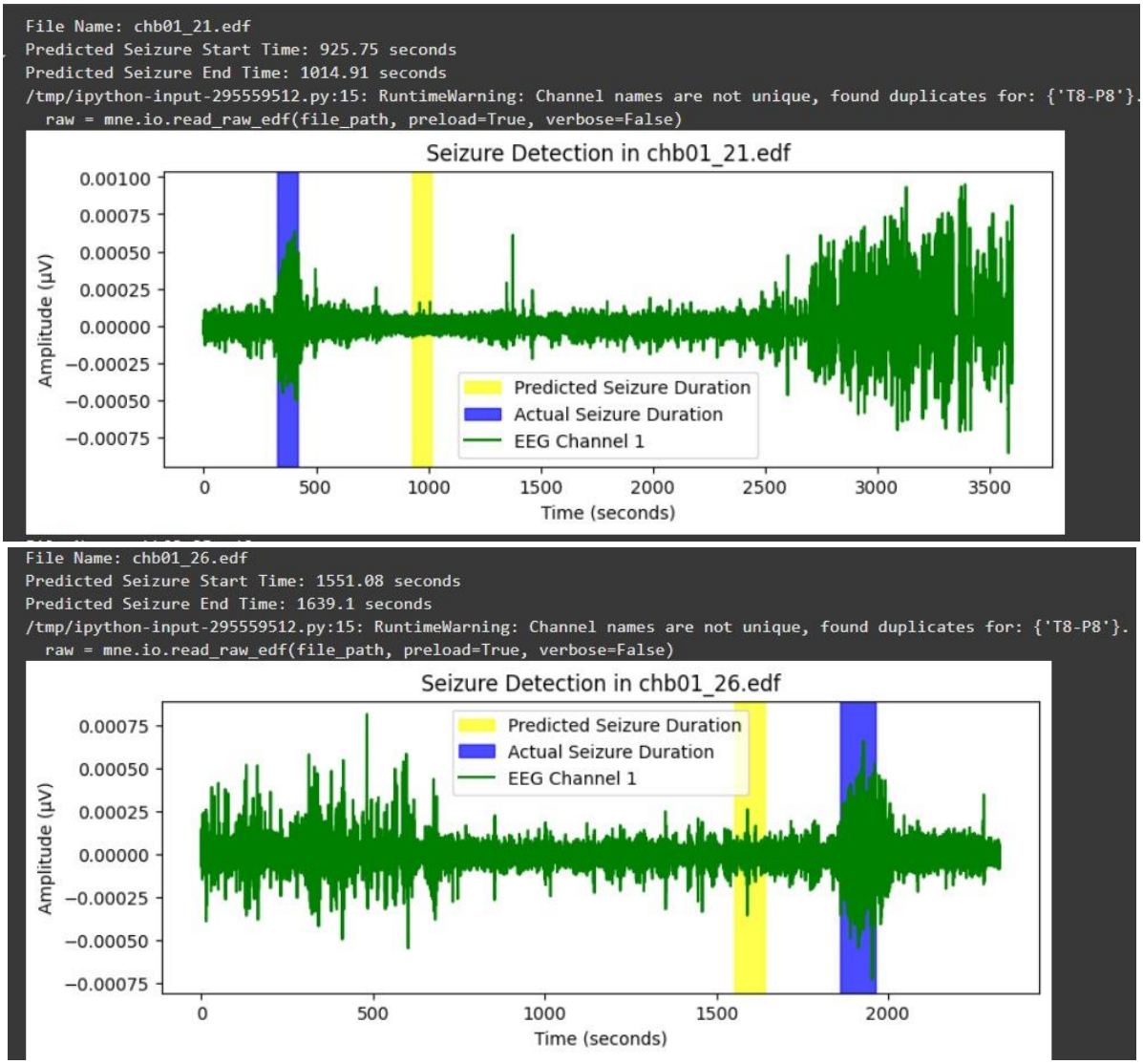


Figure 8: Graphs showing seizure detection visualization for Patient 1

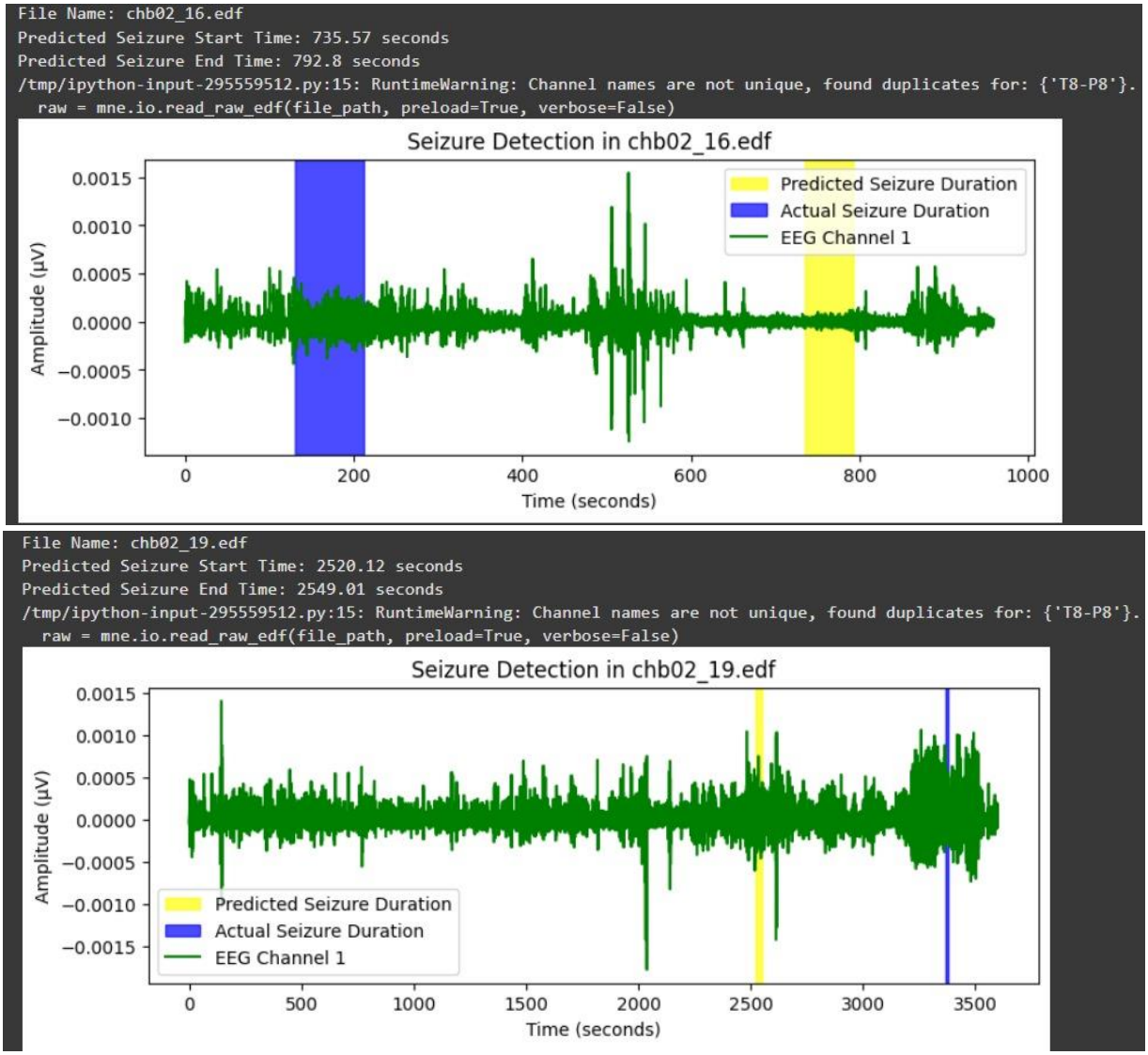
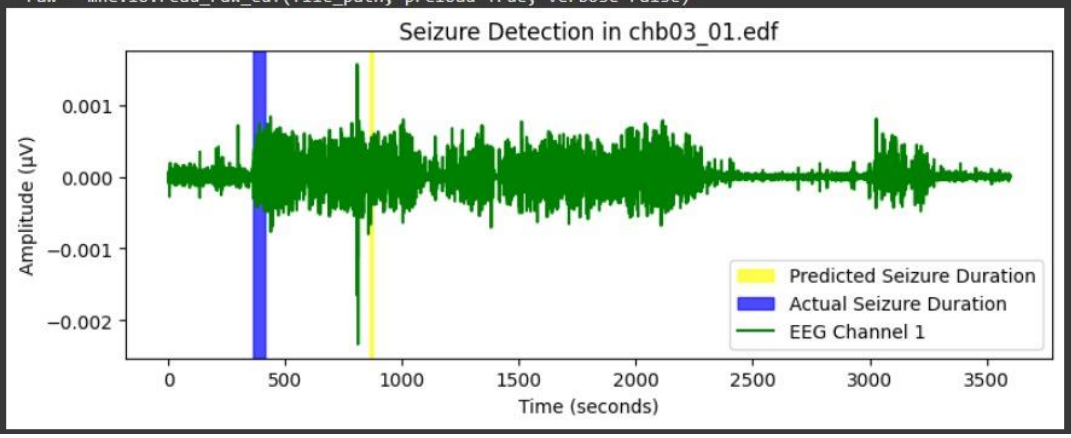
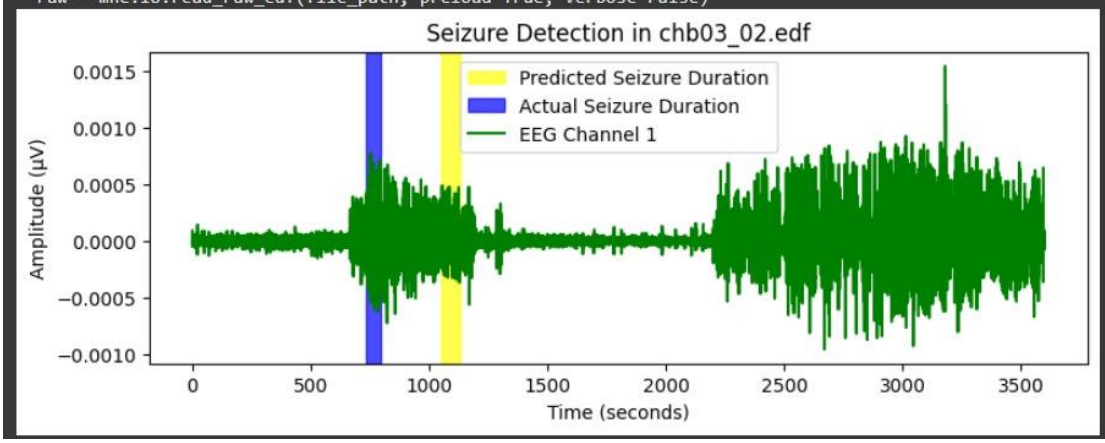


Figure 9: Graphs showing seizure detection visualization for Patient 2

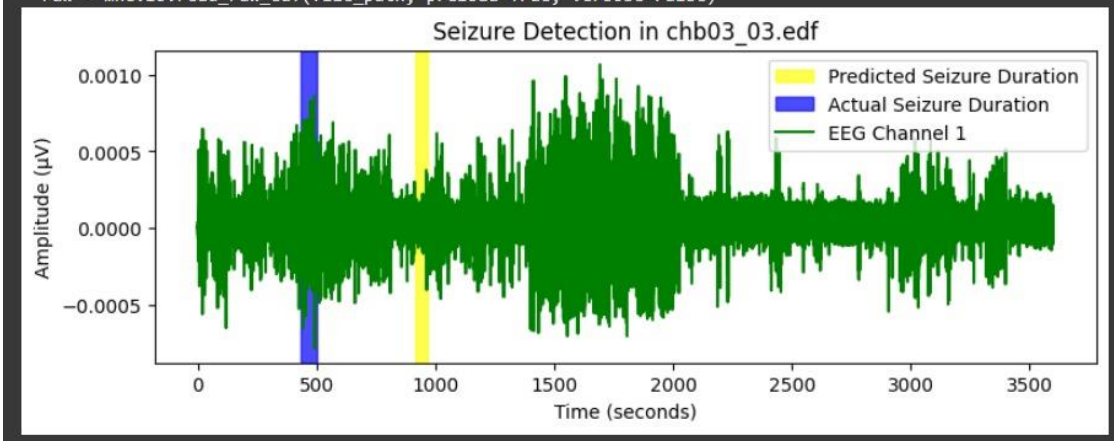
```
File Name: chb03_01.edf
Predicted Seizure Start Time: 864.92 seconds
Predicted Seizure End Time: 871.36 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```



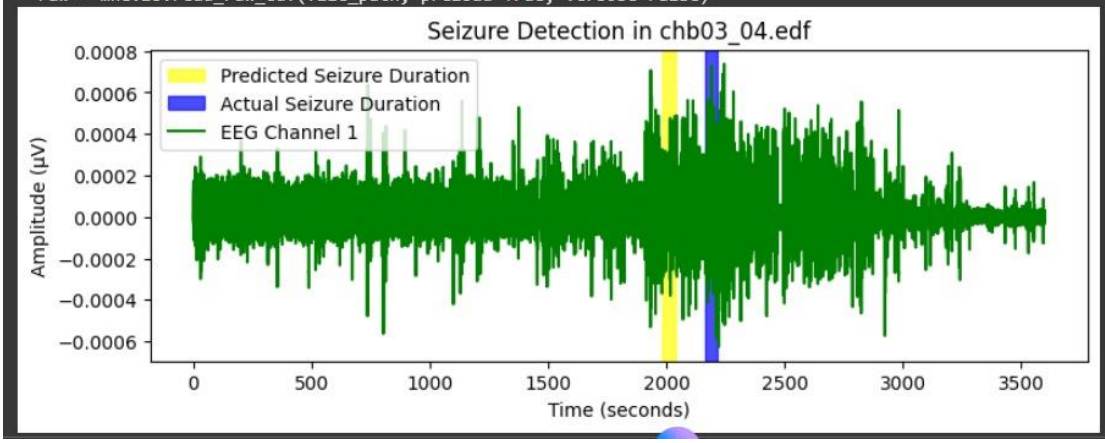

```
File Name: chb03_02.edf
Predicted Seizure Start Time: 1051.2 seconds
Predicted Seizure End Time: 1131.58 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```



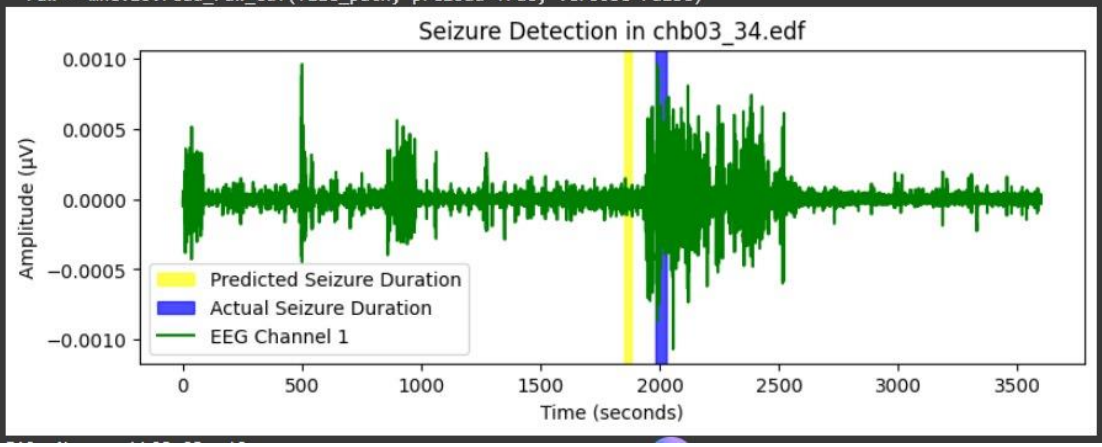

```
File Name: chb03_03.edf
Predicted Seizure Start Time: 914.35 seconds
Predicted Seizure End Time: 966.42 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```



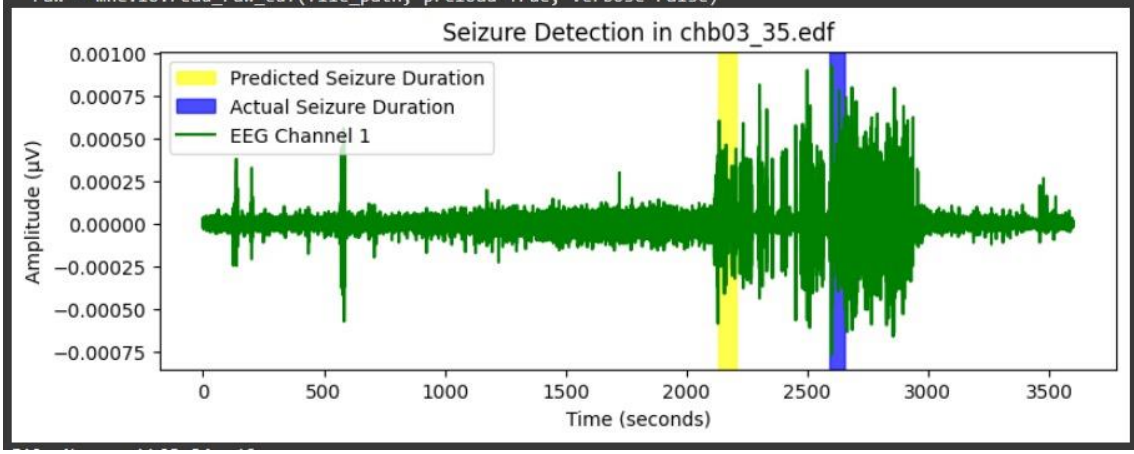
```
File Name: chb03_04.edf
Predicted Seizure Start Time: 1980.44 seconds
Predicted Seizure End Time: 2040.41 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```




```
File Name: chb03_34.edf
Predicted Seizure Start Time: 1853.95 seconds
Predicted Seizure End Time: 1879.74 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```




```
File Name: chb03_35.edf
Predicted Seizure Start Time: 2133.56 seconds
Predicted Seizure End Time: 2207.0 seconds
/tmp/ipython-input-295559512.py:15: RuntimeWarning: Channel names are not unique, found duplicates for: {'T8-P8'}.
  raw = mne.io.read_raw_edf(file_path, preload=True, verbose=False)
```



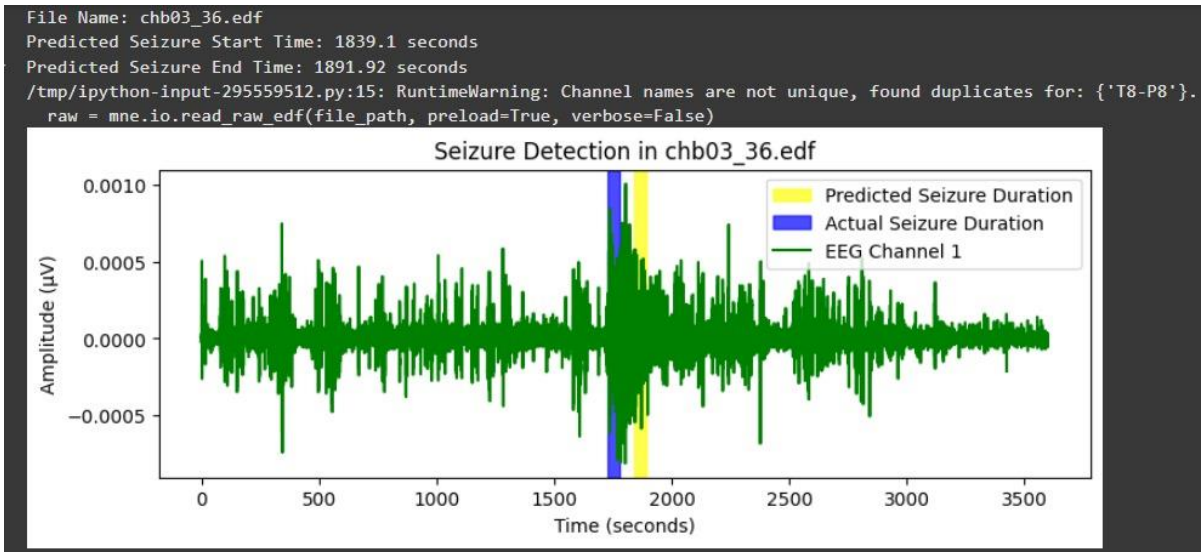


Figure 10: Graphs showing seizure detection visualization for Patient 3

5.4 Testing Results – Seizure Detection

The Random Forest Classifier was used to test on 11 EEG data that were separated for testing purpose. The model showed 100% accuracy on the test data. It correctly identified all seizure and non-seizure recordings. Each file's result matched the actual recorded label, as shown in Table 3.

This strong result shows that the model performs well on new data. It can clearly separate seizure and non-seizure EEG signals. The absence of wrong or missed detections in the test data shows that the feature extraction method was effective and well-matched to the model's learning process.

Testing Result
Result accuracy= 100.0

	File Name	Actual Result	Predicted Result
0	chb02_18.edf	No Seizure Detected	No Seizure Detected
1	chb03_11.edf	No Seizure Detected	No Seizure Detected
2	chb01_20.edf	No Seizure Detected	No Seizure Detected
3	chb03_08.edf	No Seizure Detected	No Seizure Detected
4	chb01_29.edf	No Seizure Detected	No Seizure Detected
5	chb01_15.edf	Seizure Detected	Seizure Detected
6	chb03_23.edf	No Seizure Detected	No Seizure Detected
7	chb03_19.edf	No Seizure Detected	No Seizure Detected
8	chb03_29.edf	No Seizure Detected	No Seizure Detected
9	chb03_27.edf	No Seizure Detected	No Seizure Detected
10	chb01_03.edf	Seizure Detected	Seizure Detected
11	chb02_28.edf	No Seizure Detected	No Seizure Detected

Table 3: Testing result table showing actual vs predicted seizure detection.

5.5 Testing Results – Seizure Start and End-Time Prediction

For seizure-containing test files, the Random Forest Regressor predicted seizure onset and termination times. As shown in Table 4, the mean absolute error between the start time given by the model and the actual recorded start time was about 650 seconds. Similarly, the average difference between the predicted end time and the actual end time was around 680 seconds. The largest difference in start time predicted by regressor and actual start time was about 830 seconds and the smallest was around 240 seconds.

Even though small timing differences were present, the predicted seizure periods matched the real seizure regions. This confirms that the model correctly detected the seizure parts in the EEG data. These results indicate that the regressor can approximate seizure boundaries with reasonable accuracy while maintaining interpretability.

Testing Result
Start accuracy= 800.37 seconds
End accuracy= 798.4 seconds

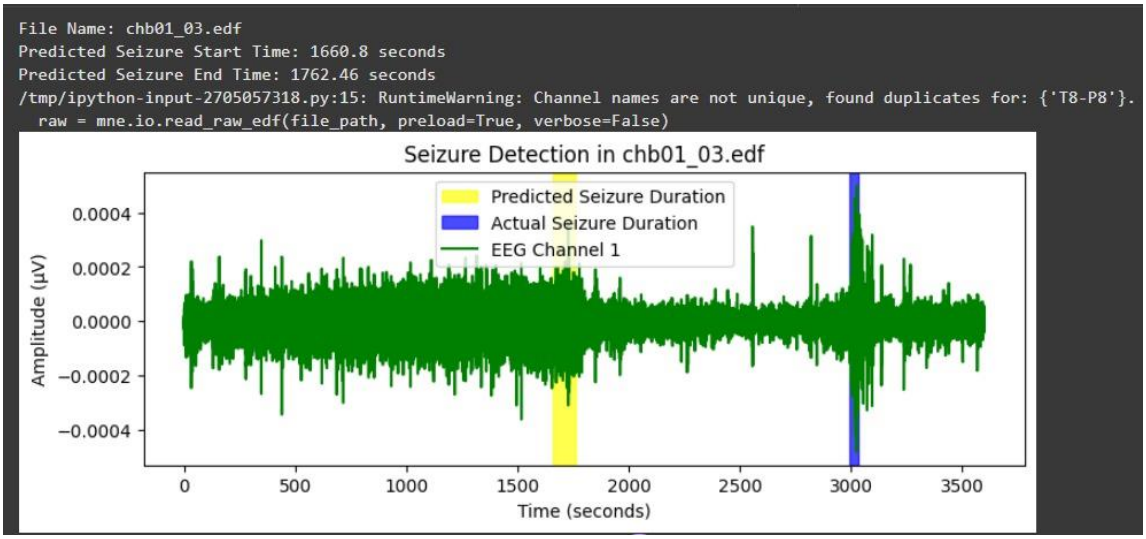
	File Name	Actual start	Predicted start	Start Difference	Actual end	Predicted end	End Difference
0	chb01_03.edf	2996	1660.80	1335.20	3036	1762.46	1273.54
1	chb01_04.edf	1467	1732.53	265.53	1494	1817.27	323.27

Table 4: Testing results showing actual vs predicted seizure start and end times.

5.6 Graphical Representation of Testing Data

To visually validate the regression performance, graphs were plotted for the test files containing seizures, as shown in Figures 11. Each plot compares the predicted and actual seizure durations. The green waveform represents the EEG signal, the blue region marks the clinically annotated seizure duration and the yellow region shows the model's predicted duration.

The figures demonstrate that the predicted intervals closely overlap the true seizure durations, though deviations exist in onset and offset times. The alignment between actual and predicted intervals confirms that the model generalizes effectively to new data and maintains its predictive consistency.



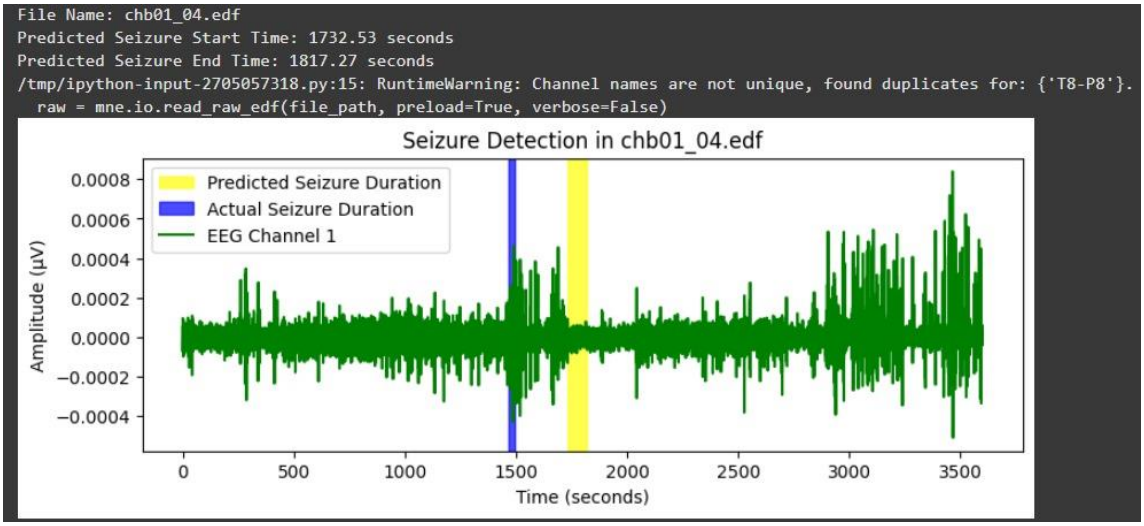


Figure 11: Testing signal plots showing predicted and actual seizure durations.

5.7 Overall Observation of Results

The testing results reinforce the robustness of the Random Forest-based model. The classifier maintained perfect accuracy (100%), demonstrating high reliability for seizure event detection, while the regressor maintained acceptable temporal estimation accuracy with deviations generally within 10–14 minutes of the true event times.

These findings confirm that the system can generalize across patients and EEG sessions, even when individual EEG patterns vary. The temporal error margins observed are typical for non-deep-learning EEG models and can be further minimized by training with larger datasets and additional patient data.

5.8 User Input Testing and Visualization

To test the model's practical use, an extra module was added that lets the user enter the path of an EEG file manually. The program reads the selected file and checks if seizure activity is present. It then shows the result in a graphical form. If a seizure is found, the model also predicts the start and end times and highlights that part of the EEG signal.

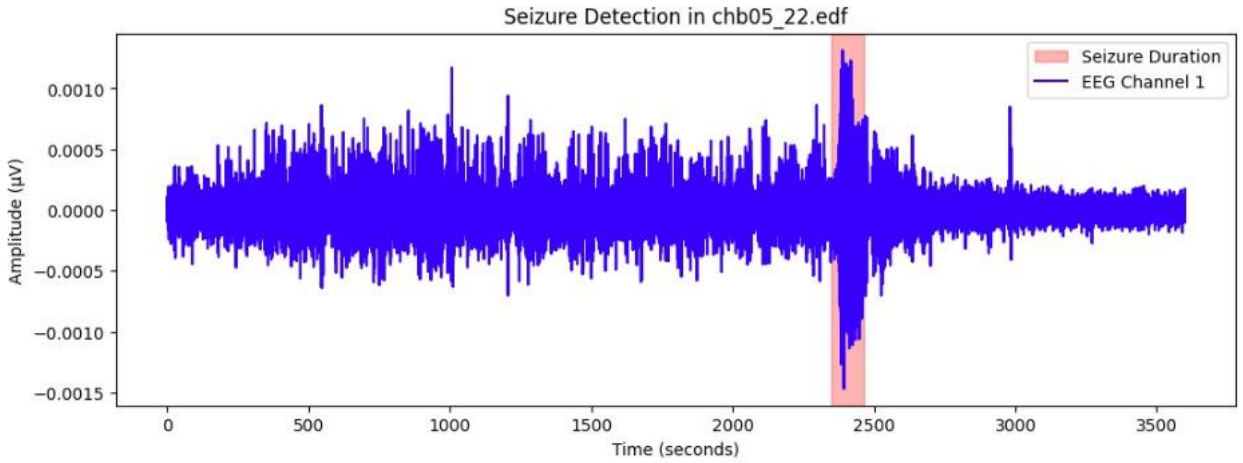


Figure 12: Testing signal plots showing actual seizure durations of a user input file

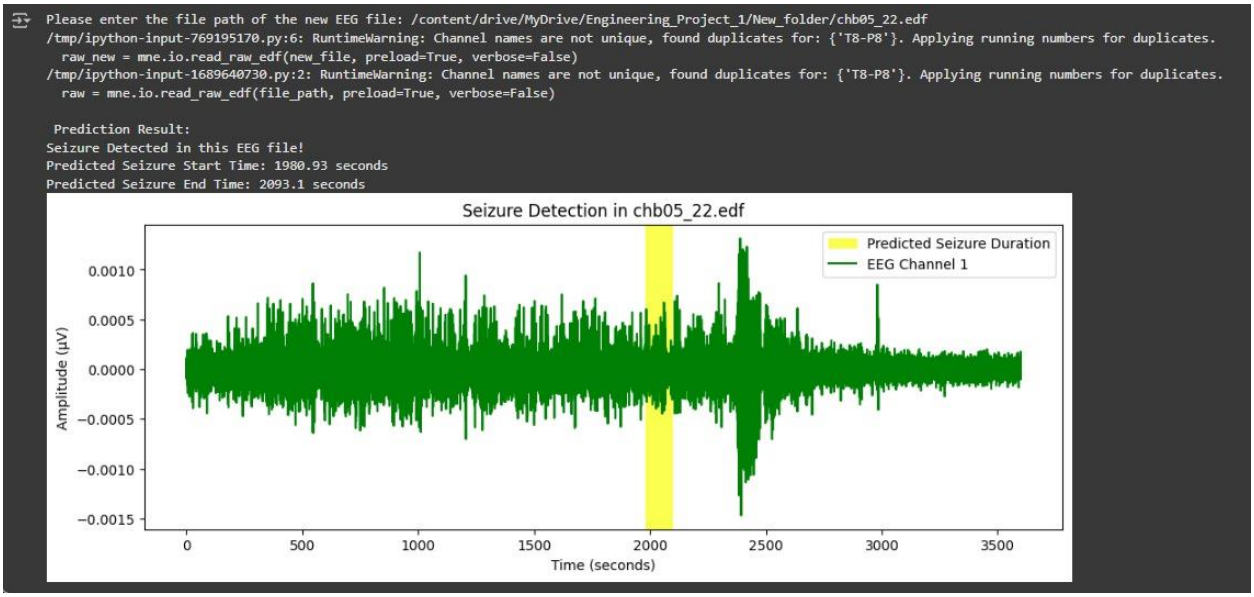


Figure 13 Testing signal plots showing predicted seizure durations of a user input file

Figures 12 and 13 show the comparison between the actual seizure duration and the predicted seizure period for a file from different patient "EEG file **chb05_22.edf**". In the first figure, the red band represents the clinically annotated seizure period, while in the second figure, the yellow band indicates the model's predicted seizure duration. The EEG waveform is shown in blue or green to represent the brain's electrical activity over time.

The results show that the model correctly detected the seizure within the right part of the EEG signal. A small timing difference can be seen between the actual and predicted seizure periods, but the model still identified the seizure correctly without any false results. This proves that the system can work interactively, analyze new EEG data in real time and give clear visual results to the user.

The graphical display improves understanding by letting doctors or researchers visually compare the real and predicted seizure areas. This helps with checking the results and doing further analysis. This user-driven testing confirms that the developed model not only performs accurately during training and testing but also works effectively as a standalone seizure detection tool.

6. Limitations of the Proposed Model and Future Improvements Planning

6.1 Limited Patient Diversity

Issue:

The current model was trained using EEG data from only three patients in the CHB-MIT Scalp EEG Database. According to Hussain and Thangarajoo (2021), EEG signal vary significantly from person to person due to differences in brain structure, electrode positioning and type of seizure. As a result, a model trained on limited subjects may fail to generalize effectively to new patients with different EEG patterns.

Planned Solution:

To improve model generalization, the dataset will be expanded to include EEG signals from a larger and more diverse group of patients. Instead of using 30–40 EDF files from one patient, future work will involve using approximately 5–6 EDF files from all 24 patients. This will ensure greater variation in EEG signals and make the model more adaptable to new individuals in clinical applications.

6.2 Inaccurate Prediction of Seizure Start and End Times

Issue:

The current Random Forest Regressor sometimes produces mismatched seizure start and end times compared to clinical annotations. This occurs because the model analyses each EEG recording as a single continuous signal, ignoring fine temporal variations within the data.

Planned Solution:

The EEG recordings will be segmented into smaller 5–10 second epochs. Each epoch will be analyzed separately for seizure activity. If the first seizure is detected in the fourth epoch, that epoch's start time will marked as the predicted seizure start time, while the transition from a seizure to a non-seizure epoch will mark the seizure ending time. This approach increases time resolution and improves accuracy in detecting seizure boundaries.

6.3 Detection of Multiple Seizures in a Single Recording

Issue:

At present, the model can identify only one seizure event per EEG file, even if multiple seizures are present in the single recording. The algorithm stops after the detection of first seizure, missing subsequent events.

Planned Solution:

An iterative epoch-based scanning mechanism will be implemented to continuously analyze the remaining EEG segments after detecting a seizure. This will allow the model to identify and detect multiple seizure within the same recording, each with its own predicted start and end times. This improvement will make the model more useful for long-term seizure monitoring.

7. Research Plan and Timeline

This section describes the planned tasks, current work and timeline for the project titled “Machine Learning for Predicting Epileptic Seizures Using EEG Signals.”

The research plan is divided into two main phases:

- Project 1: Completed in the current semester. It focused on building the model and testing its basic performance.
- Project 2: Planned for the next semester. It will focus on fixing the current limitations and improving the model’s accuracy and general use.

The aim of this plan is to follow a clear and organized process to achieve reliable seizure detection and prediction using machine learning.

7.1 Regular Activities

Regular tasks will continue during both parts of the project to keep steady progress:

- Weekly literature reviews will be done to stay updated on new work in seizure prediction and EEG studies.
- Regular meetings will be held with the project supervisor for guidance and progress review.
- All model updates, results and performance checks will be recorded continuously.
- Data and features will be checked often to make sure the EEG preprocessing and labeling are accurate.
- Research notes, figures and the Gantt chart will be updated and maintained regularly.

7.2 A timeline for Project 1

Focused on development and testing of a functional Random Forest–based seizure detection and prediction model.

Task	Description	Timeline
Literature Review	Review research on machine learning applications in EEG-based seizure prediction.	Weeks 1–3
Data Acquisition	Collect EEG data from CHB-MIT Scalp EEG Database and store on Google Drive.	Weeks 2–4
Data Preprocessing	Filter noise, remove artefacts and extract time-domain and statistical features.	Weeks 4–6
Model Development	Implement Random Forest Classifier and Regressor using Python.	Weeks 6–8
Model Training & Testing	Train model on 90% of data, test on 10% and evaluate using accuracy and mean absolute error.	Weeks 8–10
Visualization	Display actual and predicted seizure intervals graphically using EEG plots.	Weeks 10–11
Preliminary Results	Summarize classification and regression results with accuracy and time deviation metrics.	Weeks 11–12

Task	Description	Timeline
Report Submission	Prepare and finalize Project 1 progress report.	Weeks 12–14

Table 5: A timeline based on project 1

7.3 Planned Timeline in Project 2

Focused on solving the limitations identified in Project 1, improve accuracy and enhance model generalization.

Task	Description	Timeline
Dataset Expansion	Use EEG data from 24 patients (5–6 files each) to improve patient diversity.	Weeks 1–3
Epoch-Based Segmentation	Divide EEG signals into 5–10 second epochs to increase prediction accuracy.	Weeks 3–5
Multi-Seizure Detection	Implement iterative epoch-based scanning for identifying multiple seizure events.	Weeks 5–7
Model Optimization	Tune Random Forest parameters and improve feature selection methods.	Weeks 6–8
Model Evaluation	Compare new model performance against previous version using accuracy, sensitivity and timing deviation.	Weeks 8–10
Visualization & Validation	Display improved predictions graphically and validate with multiple patient data.	Weeks 10–12
Final Report & Presentation	Document outcomes, discuss findings and prepare final report and presentation.	Weeks 12–14

Table 6: A planned timeline for project 2

7.4 Gantt Chart

Tasks	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14
Literature Review														
Data Acquisition														
Data Preprocessing														
Model Development														
Model Training & Testing														
Visualization														
Preliminary Results														
Report Submission														

Table 7: Gantt Chart for Project 1 (Completed Tasks)

This chart outlines the sequence of activities completed during Project 1, including data acquisition, model development and preliminary result analysis.

Tasks	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14
Dataset Expansion														
Epoch-Based Segmentation														
Multi-Seizure Detection														
Model Optimization														
Model Evaluation														
Visualization & Validation														
Final Report & Presentation														

Table 8**:** Gantt Chart for Project 2 (Planned Tasks)

This chart presents the future timeline for addressing identified limitations, implementing epoch-based segmentation, expanding the dataset and enhancing model generalization.

8. Conclusion

This project built and tested a machine learning model for detecting and predicting epileptic seizures. It used EEG signals from the CHB-MIT Scalp EEG Database. The system used the Random Forest Classifier to detect seizure and non-seizure signals. It also used the Random Forest Regressor to find the start and end times of seizures. The process included data cleaning, feature extraction and result display. These steps helped the model study EEG data and show the output clearly in a graphical form.

The training results showed 100% accuracy in classification. This means the model correctly separated seizure and non-seizure recordings in the training data. The regression model showed some time differences between the real and predicted seizure start and end times. The average difference was about 350 seconds during training and about 800 seconds during testing. Despite this time offset, the model correctly identified the presence of seizure activity in all relevant EEG files.

Visual analysis of the EEG plots further validated the model's performance. The predicted seizure times mostly matched the real seizure periods. Small time shifts were seen in the start and end points. The system also has a user feature that lets users detect seizures directly from their own EEG files. This shows that the model can be useful for clinical or research work.

The results prove that traditional machine learning methods can work well for EEG-based seizure detection. Ensemble models like Random Forest can do this effectively while using limited computing power. However, the study also identified several challenges — including limited patient diversity, inaccurate temporal prediction and difficulty detecting multiple seizures within a single file. These findings will guide the Project 2 phase, which will focus on expanding the dataset, implementing epoch-based segmentation and improving temporal precision for multi-seizure detection.

9. References

- Savadkoohi, M., Oladunni, T. and Thompson, L. (2020). A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) Signal. *Biocybernetics and Biomedical Engineering*, 40(3), pp.1328–1341. https://www.sciencedirect.com/science/article/pii/S0208521620300851
- ieeexplore.ieee.org. (n.d.). *Enhanced Detection of Epileptic Seizure Using EEG Signals in Combination With Machine Learning Classifiers | IEEE Journals & Magazine | IEEE Xplore*. [online] Available at: https://ieeexplore.ieee.org/abstract/document/8972376.
- Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., O'Brien, T. and Razi, A. (2021). Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review. *IEEE Reviews in Biomedical Engineering*, 14, pp.139–155. https://ieeexplore.ieee.org/abstract/document/9139257
- Kode, H., Elleithy, K. and Almazaydeh, L. (2024). Epileptic Seizure Detection in EEG Signals Using Machine Learning and Deep Learning Techniques. *IEEE Access*, 12, pp.80657–80668. https://ieeexplore.ieee.org/abstract/document/10547246
- Varlı, M. and Yılmaz, H. (2023). Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning. *Journal of Computational Science*, 67, p.101943. doi: https://www.sciencedirect.com/science/article/pii/S1877750323000030
- Tsiouris, K.M., Pezoulas, V.C., Zervakis, M., Konitsiotis, S., Koutsouris, D.D. and Fotiadis, D.I. (2018). A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals. *Computers in Biology and Medicine*, 99, pp.24–37. https://www.sciencedirect.com/science/article/pii/S001048251830132X
- Hassan, F., Hussain, S.F. and Qaisar, S.M. (2022). Epileptic Seizure Detection Using a Hybrid 1D CNN-Machine Learning Approach from EEG Data. *Journal of Healthcare Engineering*, [online] 2022, p.e9579422. https://onlinelibrary.wiley.com/doi/full/10.1155/2022/9579422
- Elly Matul Imah and Arif Widodo (2017). A comparative study of machine learning algorithms for epileptic seizure classification on EEG signals. https://ieeexplore.ieee.org/abstract/document/8355065
- Karpov, O.E., Sergey Afinogenov, Grubov, V.V., Vladimir Maksimenko, Sergey Korchagin, Utyashev, N. and Hramov, A.E. (2022). Detecting epileptic seizures using machine learning and interpretable features of human EEG. *The European Physical Journal Special Topics*, [online] 232(5), pp.673–682. https://link.springer.com/article/10.1140/epjs/s11734-022-00714-3
- Shoeb, A.H. (2009). Application of machine learning to epileptic seizure onset detection and treatment. [online] dspace.mit.edu. Available at: https://dspace.mit.edu/handle/1721.1/54669
- Miltiadous, A., Tzimourta, K.D., Nikolaos Giannakeas, Tsipouras, M.G., Glavas, E., Konstantinos Kalafatakis and Tzallas, A.T. (2023). Machine Learning Algorithms for

Epilepsy Detection Based on Published EEG Databases: A Systematic Review. *IEEE Access*, 11, pp.564–594. https://ieeexplore.ieee.org/abstract/document/9999444

- Thangarajoo, R.G., Reaz, M.B.I., Srivastava, G., Haque, F., Ali, S.H.M., Bakar, A.A.A. and Bhuiyan, M.A.S. (2021). Machine Learning-Based Epileptic Seizure Detection Methods Using Wavelet and EMD-Based Decomposition Techniques: A Review. *Sensors*, 21(24), p.8485. https://www.mdpi.com/1424-8220/21/24/8485
- Shoeb, A., Kharbouch, A., Soegaard, J., Schachter, S. and Guttag, J. (2011). A machine-learning algorithm for detecting seizure termination in scalp EEG. *Epilepsy & Behavior*, 22, pp.S36–S43. https://www.sciencedirect.com/science/article/pii/S152550501100504X
- Ahmed, M.I.B., Alotaibi, S., Atta-ur-Rahman, Dash, S., Nabil, M. and AlTurki, A.O. (2022). A Review on Machine Learning Approaches in Identification of Pediatric Epilepsy. *SN Computer Science*, 3(6). https://link.springer.com/article/10.1007/s42979-022-01358-9
- Tran, L.V., Tran, H.M., Le, T.M., Huynh, T.T.M., Tran, H.T. and Dao, S.V.T. (2022). Application of Machine Learning in Epileptic Seizure Detection. *Diagnostics*, 12(11), p.2879. https://www.mdpi.com/2075-4418/12/11/2879
- Muhammad Usman, S., Khalid, S. and Aslam, M.H. (2020). Epileptic Seizures Prediction Using Deep Learning Techniques. *IEEE Access*, 8, pp.39998–40007. https://ieeexplore.ieee.org/abstract/document/9016244
- Khan, I., Mohammad Mohsin Khan and Farooq, O. (2022). Epileptic Seizure Detection using EEG Signals. [online] https://ieeexplore.ieee.org/abstract/document/9756061
- Mourad, R., Diab, A., Merhi, Z., Khalil, M. and Le Bouquin Jeannès, R. (2025). Machine and deep learning methods for epileptic seizure recognition using EEG data: A systematic review. *Brain Research*, 1864, p.149797. https://www.sciencedirect.com/science/article/pii/S000689932500358
- Garima Chandel, Aggarwal, T., Singh, T., Singh, S., Singh, K. and Singh, H. (2024). Analysis of EEG Signals Using Machine Learning Algorithms. https://ieeexplore.ieee.org/abstract/document/10582072
- Guttag, J. (2010). *CHB-MIT Scalp EEG Database*. [online] Physionet.org. Available at: https://physionet.org/content/chbmit/1.0.0/chb01/#files-panel.