Epileptic Seizure EEG Analysis & Prediction

1. Introduction Of Epileptic

Epilepsy is a serious brain illness that is an endemic neurological disorder all over the world. It is a clinical result that occurs with abnormal neurological electrical discharging of the brain. Epileptic seizures represent the most common positive signs and symptoms of brain disturbance, and epilepsy is one of the most common primary brain disorders .

The evaluation and treatment of neurophysiologic disorders are diagnosed with the electroencephalogram [EEG]. EEG is crucial for accurate classification of different forms of epilepsy .

EEG Data Recording

EEG signals were analyzed in four frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). The dataset consists of EEG recordings from **400 subjects**, including **200 epileptic patients** and **200 healthy individuals**, collected at Dicle University Medical Faculty Hospital.

Data acquisition was performed using a **PCI-MIO 16E DAQ card** with **LabVIEW** software, following the **International 10–20 electrode placement system**. Signals were sampled at **173 Hz**, digitized using **12-bit A/D conversion**, and filtered with a **0.53–40 Hz band-pass filter**.

EEG recordings were obtained from long-term (24-hour) monitoring sessions. For analysis, **23.6-second EEG segments** from four bipolar channels (F7–C3, F8–C4, T5–O1, and T6–O2) were selected. A total of **500 EEG segments**, including epileptic patterns, normal EEG, and artifacts, were used to evaluate classifier performance.

Attribute Information

The original EEG dataset consists of **5 folders**, each containing **100 files**, where each file represents EEG data from one subject. Each recording is **23.6 seconds long** and sampled into **4097 data points**, resulting in a total of **500 subjects**.

To prepare the data for analysis, each 4097-point signal was shuffled and divided into **23 segments**, with each segment containing **178 data points (1 second)**. This resulted in **11,500 samples (23 × 500)**. Each sample has **178 explanatory features (X_1–X_{178})** and **one response variable y**.

The response variable **y (column 179)** represents one of five classes:

1: Seizure activity
2: EEG from tumor region
3: EEG from healthy brain region

4: Eyes closed
5: Eyes open

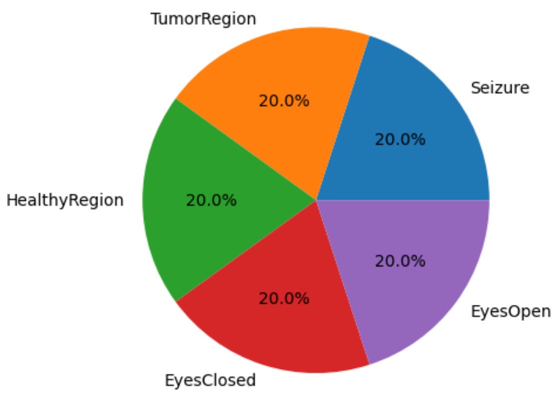
Only **class 1** corresponds to epileptic seizure, while classes **2–5** represent non-epileptic conditions. Although the dataset contains five classes, most studies, including this project, focus on **binary classification**, distinguishing **seizure (class 1)** from **non-seizure (classes 2–5)**.

2. Data Loading & Quick Overview

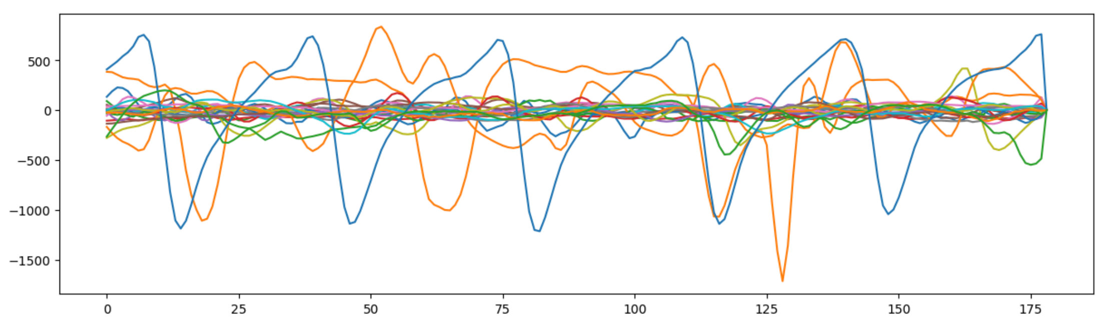
- Label column : y
- Number feature per window : 178
- Shape(11500,180)
- The number of trials for the non-seizure class is: 9200
- The number of trials for the seizure class is : 2300

- No null values

3. Exploratory Data Analysis (EDA)



- Data is evenly distributed.



- It shows epileptic signals seem to be a lot more deviating than non-epileptic signals.

Total Mean VALUE for NON-Epileptic: 1259.5026263436266

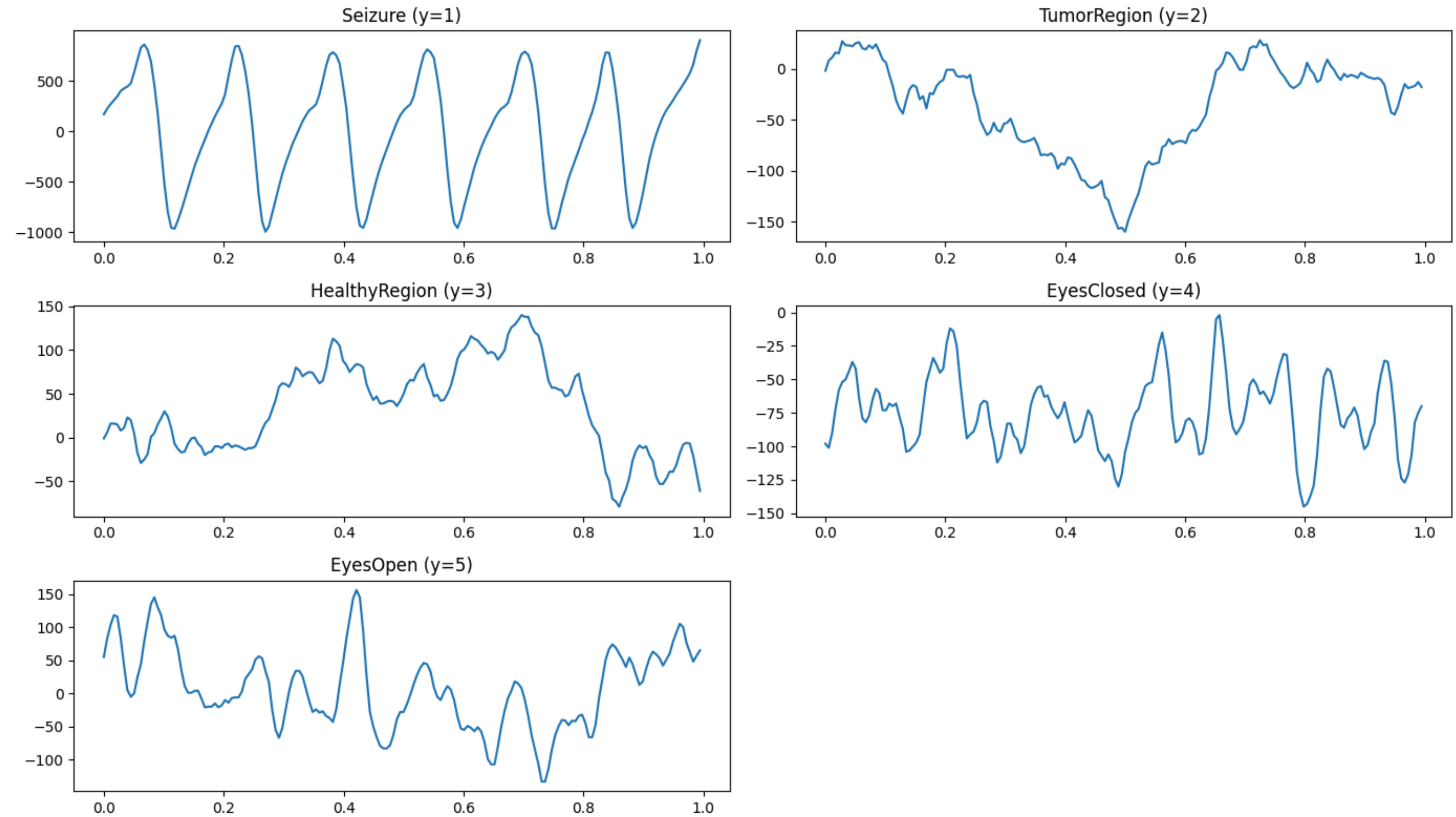
Total Std VALUE for NON-Epileptic: 15.558473767994222

Total Mean VALUE for Epileptic: 290.118920064668

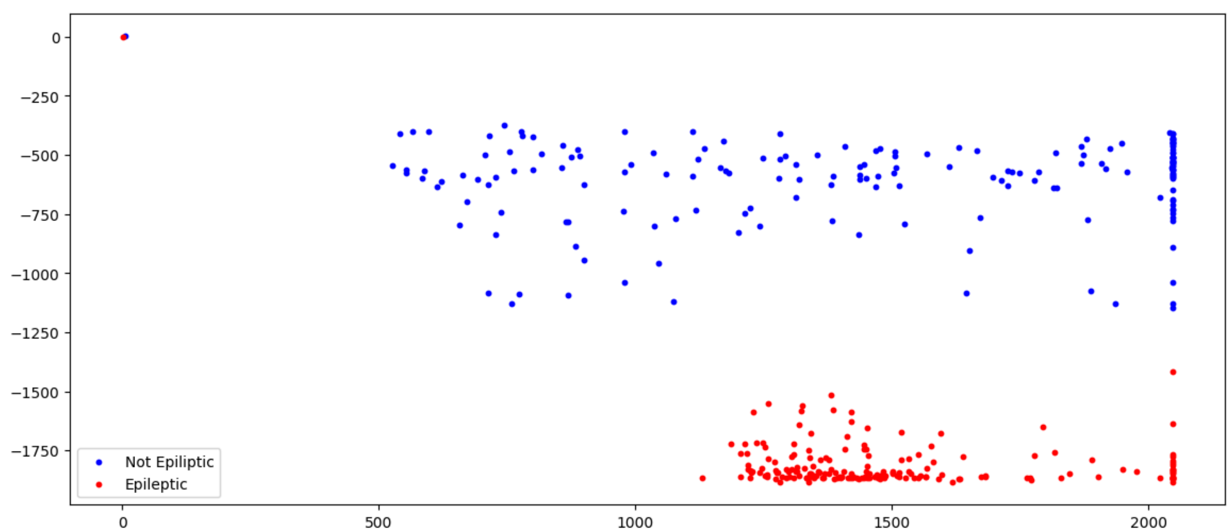
Total Std VALUE for Epileptic: 61.26568866204571

We see a quite big difference in Mean values between Epileptic and non-epilectic, we will have to normalize/scale data

---

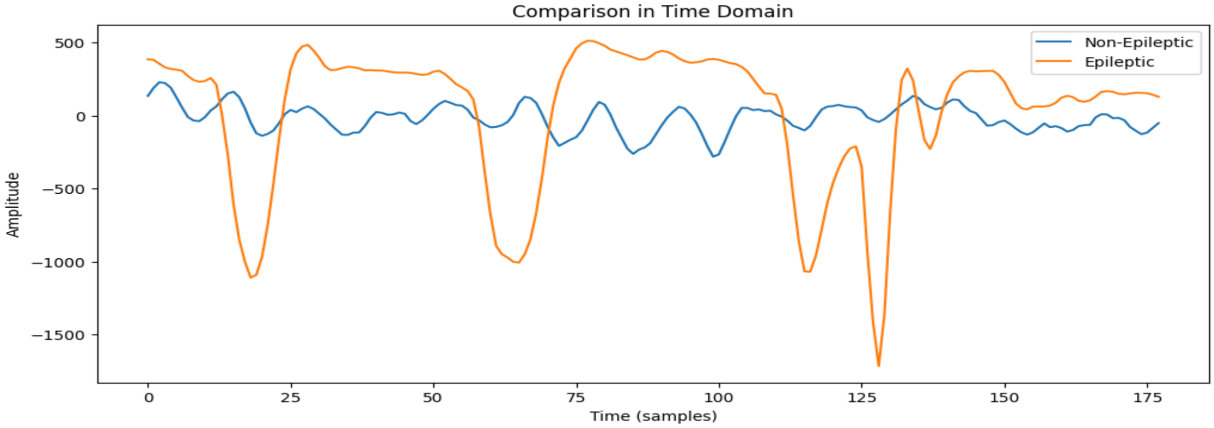


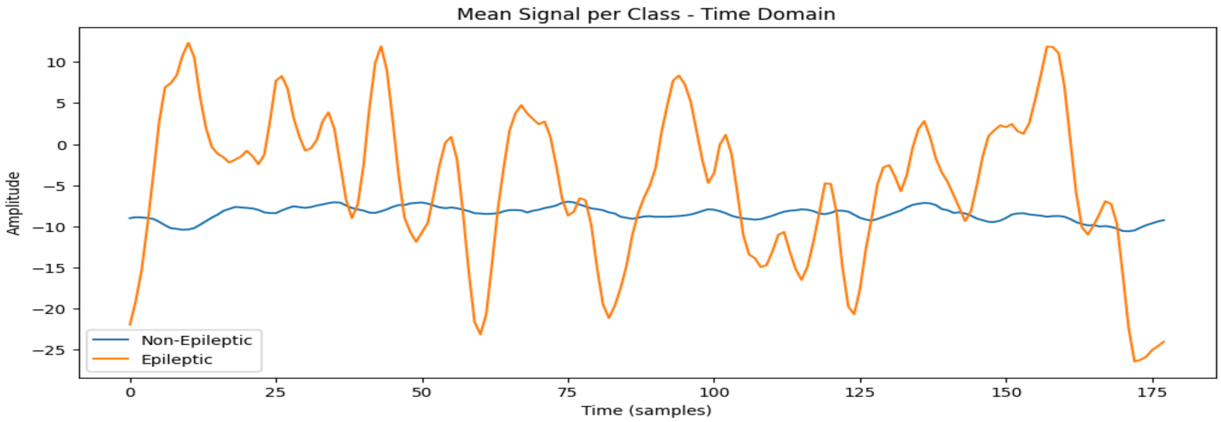
We can observe, records of Epilistic seizure are more smooth and look like they have tendency.



Time-domain visualizations & features data

Goal: visually compare signal shapes across labels and compute an envelope (mean ± std) for each label so that differences are obvious.



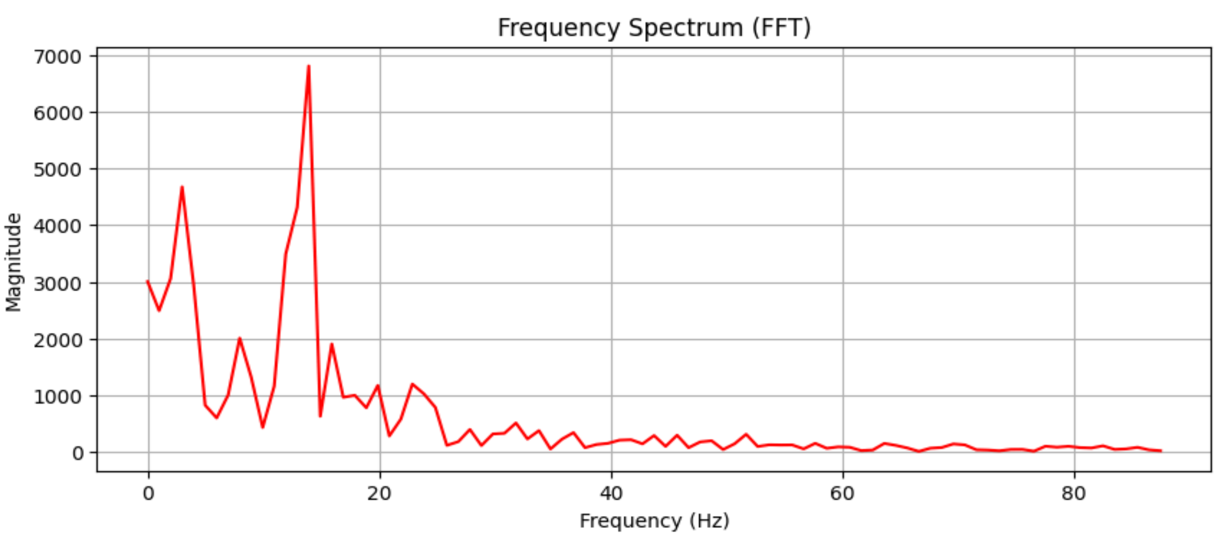


Epileptic signals are characterized by significantly higher amplitude variations and more irregular patterns compared to non-epileptic signals.

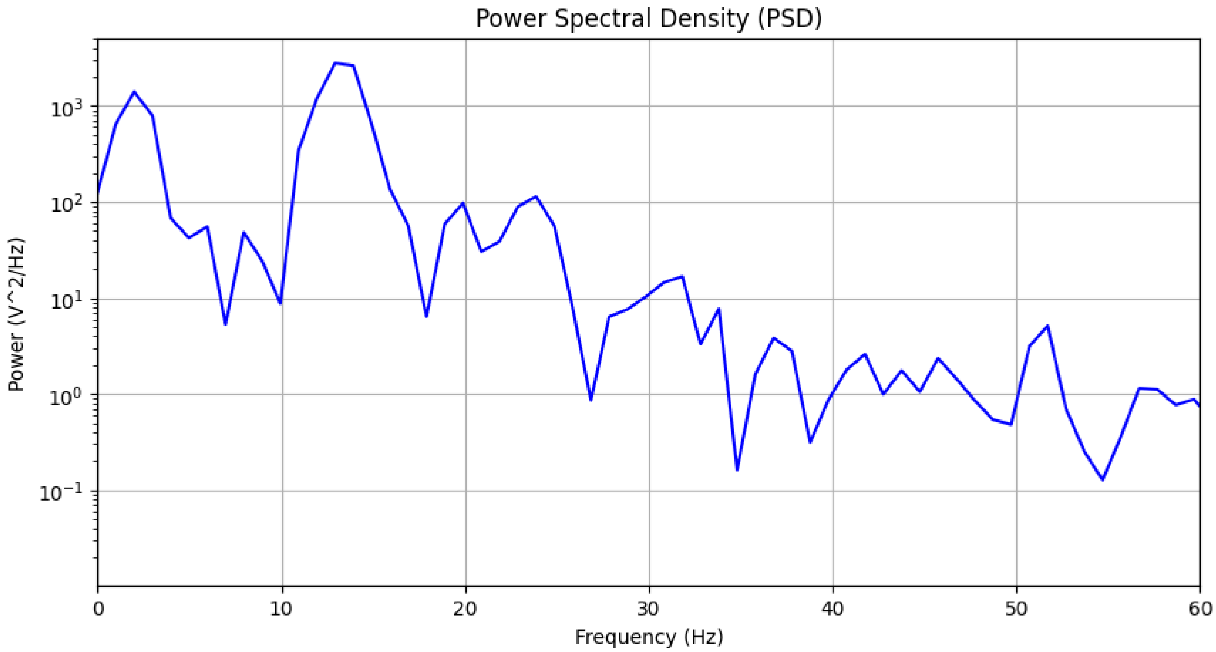
Frequency-domain visualizations & features data (PSD & bandpowers)

Goal: compute and plot average Power Spectral Density (PSD) per class and compare band-powers across classes.

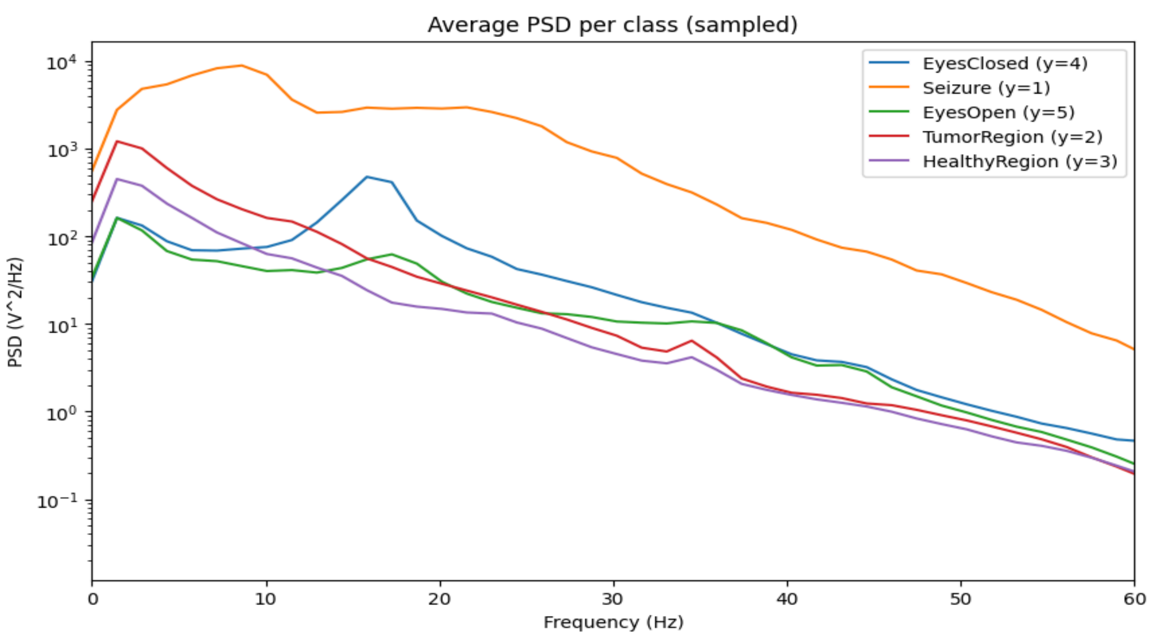
In the frequency domain, we can see what frequencies make up that signal and how strong they are. It also makes easier to remove noise and keep a specific part of the signal.



This plot clearly shows the frequencies that contain the most energy. The peak around **14-15 Hz** and another notable peak near **5 Hz** indicate that these frequencies dominate the signal's composition. The magnitude drops off significantly after about 20 Hz.



This plot is often preferred for analysis because it shows **power** instead of just amplitude. You see the same pattern as the FFT: high power near **0Hz** and a dominant peak near **14-15 HZ**. This is a clearer picture of where the signal's energy is concentrated.



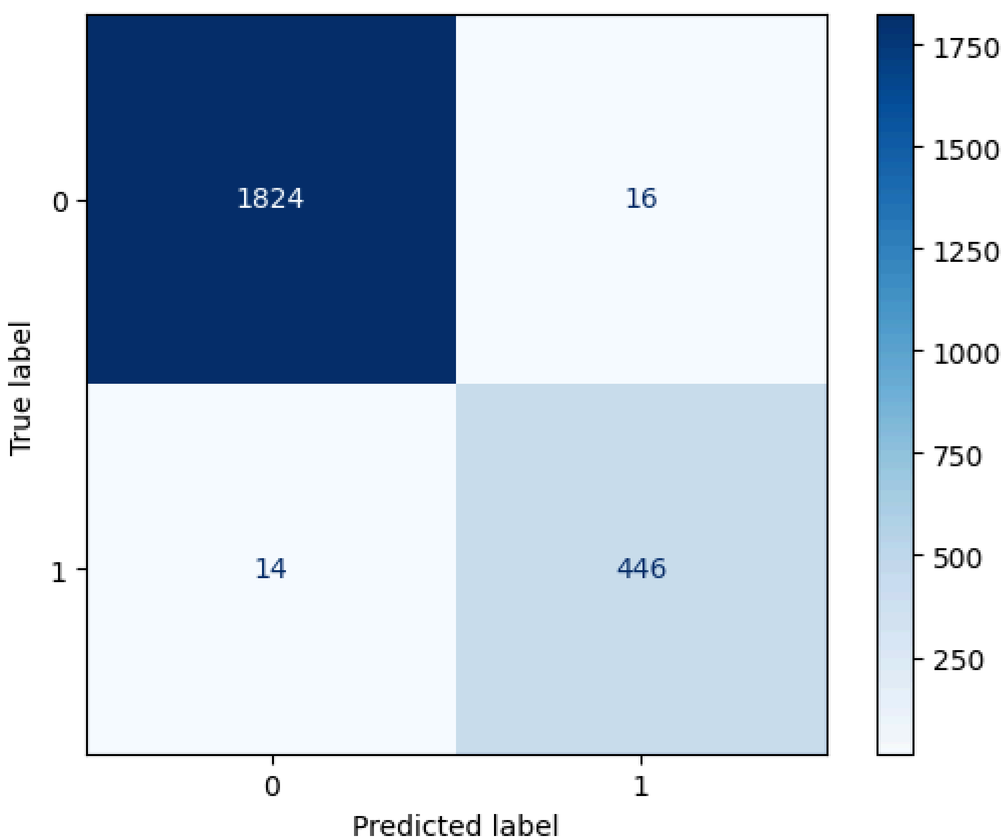
Building The Model

Multiple machine learning models (Logistic Regression, SVM, Random Forest, XGBoost, and MLP) are trained and evaluated. Features are standardized for models sensitive to scaling (Logistic Regression, SVM, MLP), while unscaled features are used for Random Forest and XGBoost.

XGBoost delivered the highest performance (balanced precision and recall), making it the most reliable model for seizure detection in this study.

The high recall for seizure class suggests the system is effective at minimizing missed seizure events, which is critical for real-world medical applications.

Overall, the results confirm that **EEG-based binary classification using machine learning is highly effective for epileptic seizure detection**, and ensemble models like XGBoost are particularly well-suited for this task. With further validation on real-time and patient-specific data, this approach has strong potential for clinical decision-support systems.



- The model **rarely misses seizures** (very low FN = 14), which is crucial for patient safety.
- **False alarms are minimal** (FP = 16), reducing unnecessary clinical interventions.
- Balanced high sensitivity and specificity indicate the model is **well-calibrated and reliable**.