EEG Classification Model

Aarya Gadekar

Andrew Roth

Pooja Patil

Table of Contents

1. Introduction

1.1 Background

Electroencephalography (EEG) is an vital method in scientific diagnostics, specifically for understanding epileptic seizures. Good interpretation of EEG recordings requires medical knowledge and a significant time investment. Because of this, deep-learning models are increasingly more important to detect seizures as it could help reduce any human error. This project is designed to evaluate classical and modern-day models on the EEG datasets from CHB-MIT and Bonn.

The primary goals are is to preprocess raw EEG signals by applying filtering, noise removal, segmentation, and normalization. To do so, we will extract key features in order to train and evaluate traditional and deep learning models. Then, we will assess model performance on various datasets, visualize EEG data and model predictions, and analyze the results.

As mentioned above, two datasets are used, one from CHB-MIT and the other from Bonn. The CHB-MIT Scalp EEG Database includes multiple channels (23 EEG channels) and has long recordings from real clinical settings, with notes indicating when seizures occurred. The Bonn EEG Dataset has short recordings from a single channel, where the EEG segments are already divided into parts that show normal, between-seizure, and seizure-related activity.

2. Data Preprocessing

2.1 CHB-MIT Dataset

According to the report (pages 3–5) , each EDF file contains 23 channels with a sampling rage of 256 Hertz (Hz). The duration of the sample is around 3600 seconds (1 hour) where seizure intervals are parsed from summary text into a structured data frame.

A preprocessing function applies band-pass filtering, notch filtering and resampling. The band-pass filter keep only the frequencies from 0.5 to 40 Hz and is used to keep the frequencies where most brain activity occurs. The notch filter removes frequencies in a small range around 60 Hz, because power line operate at this frequency, and it can disprupt actual brainwave activity. Lastly, the data is resampled to reduce the data volume and processing time, while highling the frequencies recorded below 64 Hz.

Signals are then segmented into 10 second windows with a step size of 5 seconds, meaning there is a 5 second gap between windows and each window overlaps by 5 seconds with the previous one. A window is labelled as ‘seizure’ if it overlaps with a seizure annotation by at least one second. Lastly, each window undergoes per-channel z-score normalization to standardize amplitude differences.

2.2 Bonn EEG Dataset

The Bonn EEG dataset contains 500 single-channel signals with 4097 samples per signal, or one reconding every 23.6 seconds. The sampling rate is 173.6 Hz over the sets Z, O, N, and F, which contain signals without seizures, and set S, which contain seizures. A Butterworth band-pass filter is applied, keeping only signals between 0.5 and 40 Hz and z-score normalization is applied.

3. Feature Extraction

3.1 Time-Domain Features

Extracted metrics include simple features such as mean and standard deviation, as well as advanced features including skewness, kurtosis, peak-to-peak amplitude, and energy. We then used Welch's Power Spectral Density (PSD) to evaluate each signal at different frequencies.

3.2 Combined Feature Matrix

The final feature matrices are shows as:

Dataset	**Shape**	**Description**
CHB-MIT	(4771 × 13)	Windows × features
Bonn	(500 × 13)	Segments × features

4. Dataset Splitting

4.1 CHB-MIT Split

Set	**Samples**	**Seizure Windows**
Train	3339	70

Validation	716	15
Test	716	15

4.2 Bonn Dataset Split

Set	**Samples**	**Seizure Count**
Train	349	70
Validation	75	15
Test	76	15

5. Model Development

5.1 Random Forest

Our random forest is made up of 200 decision trees, weighted by balance, meaning that the rarer the seizures, the higher the weight of the seizure, where it is important to not misclassify a window that should be a 'seizure'. Each window is fitted with 13-dimensional feature vectors.

From CHB-MIT, the random forest had a high overall accuracy but just around a 7% seizure recall. On the other hand, the Bonn dataset random forest had a 97% accuracy and an 87% seizure recall.

5.2 Deep Learning Models

Now, we will use deep learning models to analyze the datasets. The CNN takes full raw EEG windows, including Three Conv1D and MaxPooling block, a dense layer with 128 units, a dropout and a sigmoid output. The LSTM has a single layer with 64 units followed by dense layers. It overfits quickly and had poor effects on seizure detection.

To prevent overfitting, we used early stopping, which tells the model to stop training when the model stops getting better. We also used an Adam optimizer with Learning Rate of

0.001 to train the data faster and more stably. Lastly, we used a batch size of 32 to train 32 samples at a time before changing model weights.

6. Model Evaluation and Visualization

A comprehensive evaluation of all models was performed using accuracy, precision, recall, F1-score, confusion matrices, and visual diagnostic tools. These evaluations clearly highlight the performance differences between classical machine-learning models and deep-learning architectures when applied to EEG-based seizure detection.

The CNN model demonstrates exceptional classification capabilities. Its confusion matrix shows near-perfect discrimination of normal EEG windows, with zero false positives and only three false negatives across the 15 seizure windows in the test set. This indicates that the model captures subtle temporal and spatial signatures of seizure activity that are difficult to encode through handcrafted features alone. The CNN's training and validation curves further support the fact that the loss decreases smoothly during training, and the validation loss stabilizes early, reflecting a well-generalized model that does not overfit. These behaviors align with the strengths of convolutional architectures in analyzing multichannel biomedical signals, where local temporal structures carry meaningful diagnostic information.

In contrast, the LSTM model, though theoretically suited for sequential data, exhibits less stable learning behavior. The validation loss fluctuates noticeably across epochs, suggesting difficulty in capturing long-term dependencies in EEG sequences without additional convolutional preprocessing. Additionally, its confusion matrix reveals a significantly lower seizure recall, meaning the LSTM frequently misclassifies seizure windows as normal. This underperformance is attributed to the model's sensitivity to class imbalance and its tendency to overfit to the dominant non-seizure class, despite the use of regularization and early stopping.

The Random Forest classifier, trained on handcrafted features, provides useful interpretability but limited predictive capability for the CHB-MIT dataset. Feature importance analysis indicates that alpha and beta band powers are the most influential predictors. These findings are consistent with known neurophysiological patterns, where seizure activity disrupts normal rhythmic oscillations within these frequency ranges. However, despite informative features, the Random Forest model fails to capture the complex nonlinear structure of real seizure dynamics, as reflected in its extremely low seizure recall for CHB-MIT recordings. Conversely, its performance on the cleaner Bonn dataset is much stronger, highlighting that classical models are effective only when signals are noise-free and well-structured.

Finally, visual inspection of EEG windows reinforces these quantitative results. Normal EEG windows exhibit smoother, lower-amplitude oscillations with consistent rhythmic qualities. During seizures, the EEG transforms dramatically: amplitude increases, oscillations become irregular, and rapid bursts and fluctuation patterns emerge. The CNN model effectively learns these distinctions directly from the raw multichannel input, whereas traditional models relying solely on summary features lose access to the fine-grained temporal structure necessary for reliable seizure detection.

7. Discussion

The CHB-MIT dataset is highly challenging for classical machine learning due to noise, multichannel interactions, and extreme class imbalance. Random Forest models fail to identify seizure patterns effectively. In contrast, CNNs excel by learning spatiotemporal filters capable of capturing non-linear EEG structures. LSTMs fail because of long sequence length and imbalance. The Bonn dataset, because it is clean and single-channeled, is well-suited for simpler models.

8. Conclusion and Future Work

CNNs significantly outperform classical models for multichannel EEG seizure classification. Future enhancements include adding RQA and recurrence-network features and using spectrogram-based 2D CNNs. We could also explore transformer EEG, using them independently or alongside a CNN model. Additionally, we could highly focal loss to examine rare examples or examples that have cloudy classifications. Lastly, we could also try synthetic EEG augmentation to further reduce overfitting, especially with a small seizure sample size.

9. Tables of Observations

Table 1: CHB-MIT Window Statistics

Metric	**Count**
Total windows	4771

Seizure windows	100
Non-seizure windows	4671
Channels per window	23
Samples per window	1280

Table 2: Bonn Dataset Observations

Subset	**Description**	**Class**	**Samples**
Z	Healthy (eyes open)	Non-seizure	100
O	Healthy (eyes closed)	Non-seizure	100
N	Interictal	Non-seizure	100
F	Interictal	Non-seizure	100
S	Seizure	Seizure	100

Table 3: Model Performance Summary

Model	**Dataset**	**Accuracy**	**Seizure Recall**	**Notes**
Random Forest	CHB-MIT	98%	**~7%**	Fails at seizure detection
CNN	CHB-MIT	99.6%	**80–93%**	Best-performing model
LSTM	CHB-MIT	98%	**~27%**	Overfits, weak recall

Random Forest	Bonn	97%	**~87%**	Performs well

Table 4: Feature Importance (Random Forest)

Feature	**Importance Rank**
Alpha power	1
Beta power	2
Gamma power	3
Delta power	4
Theta power	5
Energy	6
Kurtosis	7
Skewness	8
Peak-to-peak	9
Max	10
Min	11
Mean	12