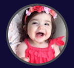
EEG-Based Epilepsy Seizure Classification Using Machine Learning

This deck explores using machine learning to classify EEG seizure types. It compares model performance before and after feature engineering and dataset balancing with SMOTE.



by Faiza Sifat

Made with GAMMA

Motivation and Dataset Overview

Motivation

Manual EEG analysis

Time-consuming, prone to human error, and inefficient for long-term monitoring

Automated EEG interpretation

Possibility with machine learning for better accuracy and efficiency

Transforming raw EEG data

Machine learning to derive meaningful clinical insights

Dataset Characteristics

The EEG dataset was collected from 6 patients diagnosed with focal epilepsy during presurgical evaluations at the American University of Beirut Medical Center from January 2014 to July 2015. Anti-seizure medications were halted to record habitual seizures.

- Patients: 6
- Total Seizures: 35
- Sampling Rate: 500 Hz
- Electrodes: 21 scalp electrodes (10-20 system)
- Filtered Range: 1.6Hz–70Hz (excluding 50Hz electrical interference)

Made with GAMMA

Data Preprocessing and Visualization

1

Preprocessing

1. Loaded '.npy' files for training and testing sets using NumPy.

2. Reshaped 3D EEG matrices into 2D arrays for model compatibility.

3. Balanced the dataset using SMOTE to address class imbalance, especially for Classes 2 and 3.

4. Scaled/normalized the data if required for specific model implementations.

2

Data Exploration

Analyzed the statistical properties and distributions of the EEG signals to gain insights into the dataset.

3

Class Distribution

Dominant Class 0, smallest Class 3 with only 6 samples

4

EEG Signal

Waveforms show brainwave activity varies across channels

5

Average Signal

Highlights key differences in signal strength and shape

6

Correlation Heatmap

Shows high correlation between adjacent EEG channels

7

Energy Plot

Class 1 has highest average energy, Class 0 lowest.

Made with GAMMA

Machine Learning Models

XGBoost

XGBoost (Extreme Gradient Boosting) is an optimized gradient boosting algorithm known for its speed and performance. It uses second-order gradient information to minimize the loss function and includes regularization to reduce overfitting.

Random Forest

Random Forest is an ensemble of decision trees built using bagging (bootstrap aggregation). It introduces randomness by selecting subsets of features during tree construction, thus improving model generalization. Its prediction is based on the majority vote from all trees.

LightGBM

LightGBM is a gradient boosting framework that grows trees leaf-wise (as opposed to levelwise in traditional GBDT). It is optimized for speed and memory efficiency and supports large-scale datasets with high dimensionality.

Made with GAMMA

Experiments and Evaluation

Without Feature Engineering

Raw reshaped EEG data was used for training: Table 1: Performance Without Feature Engineering

Model	Accuracy (%)	F1-Score (%)
XGBoost	91.66	91.57
Random Forest	89.09	88.78
LightGBM	91.01	90.72

With Feature Engineering + SMOTE

The data was balanced using SMOTE and reshaped to improve model learning: Table 2: Performance With Feature Engineering and Balancing

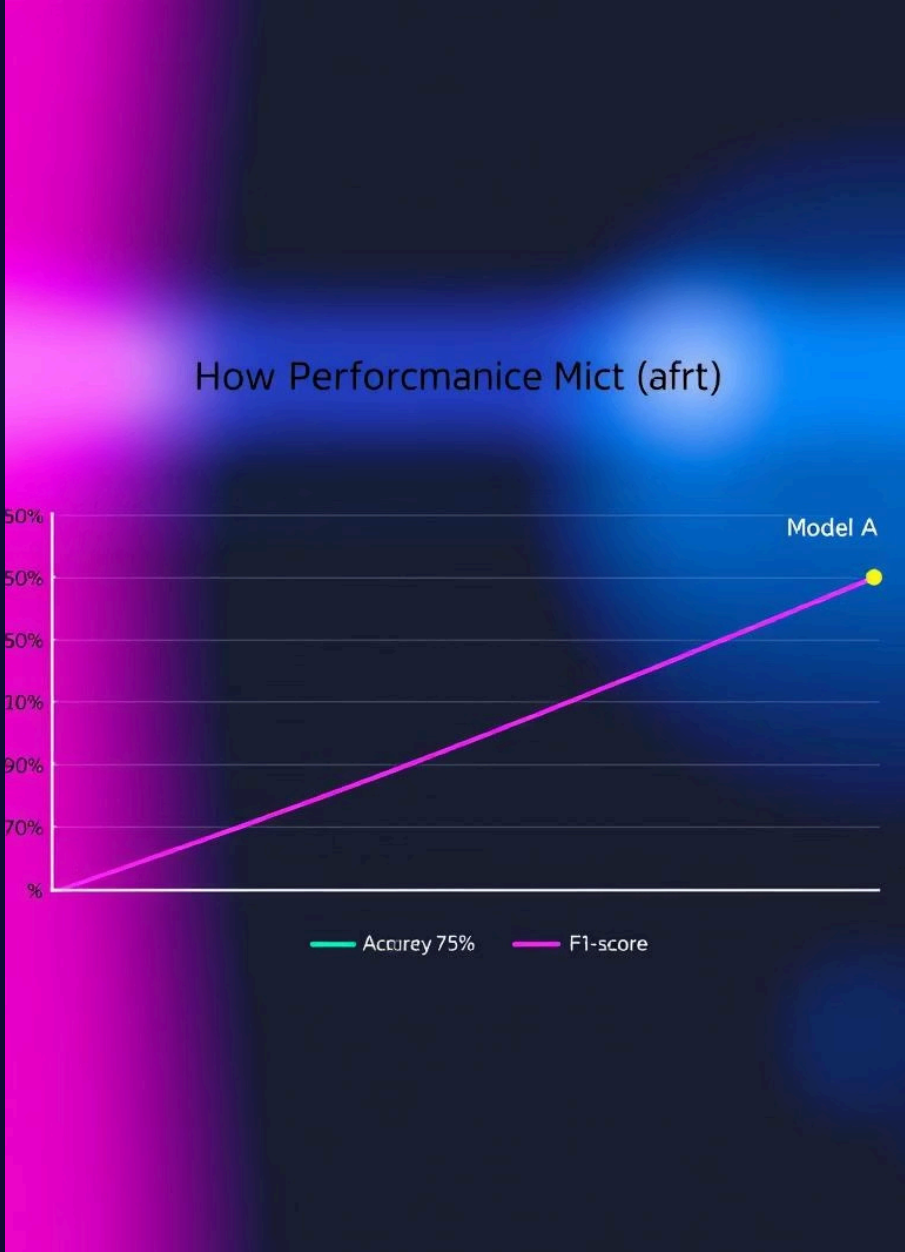
Model	Accuracy (%)	F1-Score (%)
XGBoost	91.66	91.72
Random Forest	88.70	88.83
LightGBM	91.40	91.46

Made with GAMMA

Visual Comparison of Results and Insights

Insights

- XGBoost has consistent high performance in both raw and engineered datasets. - LightGBM performs similarly, with slightly lower variance and better speed. - Random Forest lags slightly, especially in detecting rare classes. - SMOTE significantly improves recall for Classes 2 and 3.



Made with GAMMA

Conclusion and Future Work

Conclusion

This project demonstrated the effectiveness of machine learning models in detecting different types of epileptic seizures from EEG data. We showed that: - Feature engineering and balancing help improve minority class detection. - XGBoost and LightGBM offer robust solutions with high accuracy. - Random Forest is useful but slightly less accurate with imbalanced data.

Future Work

Future improvements could include deep learning methods like CNNs or RNNs to automatically extract spatial-temporal EEG features.

Made with GAMMA

References

- EEG Dataset: https://data.mendeley.com/datasets/5pc2j46cbc/1
- Chawla, N. V., et al. "SMOTE: Synthetic Minority Over-sampling Technique." Journal of Artificial Intelligence Research, 2002.
- Chen, T., and Guestrin, C. "XGBoost: A Scalable Tree Boosting System." 2016.
- Ke, G., et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." NIPS 2017.
- Scikit-learn: https://scikit-learn.org/
- XGBoost Documentation: https://xgboost.readthedocs.io/
- LightGBM Documentation: https://lightgbm.readthedocs.io/

Made with GAMMA