Barbara Pobiedzińska

Agata Semerjak

Maciej Wieloch

Identification of Epilepsy from EEG Data:

A Machine Learning Approach

Introduction

Epilepsy, a neurological disorder characterized by recurrent seizures, poses significant challenges in diagnosis and management. Detecting epileptic activity from EEG (Electroencephalogram) data plays a crucial role in understanding and treating this condition. Accurate classification of EEG recordings can aid in timely intervention and personalized treatment plans. In this study, we apply machine learning techniques to EEG data for the recognition and analysis of epileptic seizures.

Methods

Data Source and Composition

The dataset, obtained from Kaggle, comprises EEG recordings from 500 individuals, each with 4097 data points representing 23.6 seconds of brain activity. Subjects are categorized into five classes, with classes 2-5 representing non-seizure conditions and class 1 indicating epileptic seizures.

Data Processing and Format

The dataset is organized into 5 folders, each containing 100 files corresponding to individual subjects. EEG recordings are divided into 23 chunks, each containing 178 data points, resulting in a total of 11,500 records. The response variable ('y') denotes different brain activity states.

Models

We employ three classification models: K-Nearest Neighbors (KNN), Decision Tree, and Random Forest Classifier. These models are trained to distinguish between epileptic seizure (class 1) and non-seizure conditions (classes 2-5).

Results

Analysis of classification results in the "EEG Epilepsy Recognition" project shows that all three methods used - K-Nearest Neighbors (KNN), Decision Tree and Random Forest Classifier - achieve high accuracy. KNN and Decision Tree showed significant improvements in accuracy after using extracted features, with the greatest difference in accuracy for KNN. Random Forest Classifier, although it achieved the highest accuracy without features, did not show significant improvement after using them.

	K-Nearest Neighbors	Decision Tree	Random Forest
Accuracy (no features)	0.9243	0.9397	0.9780
Accuracy (after using features)	0.9530	0.9699	0.9774
Precision	0.9303	0.9571	0.9691
Recall	0.9218	0.9481	0.9598
Accuracy difference	0.0287	0.0301	-0.0006

Summarized findings

Discussion

The success of machine learning models in accurately classifying EEG data for seizure recognition highlights their potential in aiding medical professionals. KNN and Decision Tree, in particular, demonstrate enhanced performance with feature extraction. The Random Forest Classifier, despite its initial high accuracy, shows limited sensitivity to additional features.

Conclusion

This study emphasizes the effectiveness of machine learning in EEG-based epilepsy recognition. The models, particularly KNN and Decision Tree, showcase promising results in distinguishing seizure and non-seizure states. Further research could explore deep learning models, data augmentation techniques, and the integration of additional data sources to enhance overall accuracy and applicability in clinical settings.

In conclusion, the project's objective is to compare the effectiveness of different classification algorithms in distinguishing between normal brain activity and epilepsy-related activity using EEG data.