Automated Seizure Detection from EEG Signals Using Spectral and Wavelet-Based Machine Learning Techniques

Bisera Nikoloska
bn1872@student.uni-lj.si
Faculty of Computer and
Information Science,
University of Ljubljana
Večna pot 113
SI-1000 Ljubljana, Slovenia

ABSTRACT

This project explores automated seizure detection from EEG data through the integration of signal processing and machine learning techniques. By extracting key features from both power spectrum and wavelet-transformed EEG signals, we aim to improve the accuracy of seizure detection. Three machine learning classifiers—Support Vector Classifier, Logistic Regression, and Random Forest—were trained and optimized using grid search with cross-validation. Our findings indicate that the combination of spectral and wavelet features enhances detection performance. This study underscores the potential of machine learning to advance clinical seizure diagnosis.

KEYWORDS

wavelet transform, eeg, seizure detection, machine learning, spectral features

1 INTRODUCTION

Electroencephalography (EEG) is an essential tool in the study of brain activity, providing detailed insights into electrical signals generated by the brain. Accurate seizure detection using EEG signals is crucial for diagnosing and managing epilepsy. However, the complexity of EEG data, characterized by its high dimensionality and susceptibility to noise, poses significant challenges for reliable seizure detection.

This project aims to develop a robust automated system for detecting epileptic seizures from EEG recordings. The approach involves preprocessing the EEG data to filter out noise and enhance signal quality. Key spectral features such as dominant frequency, spectral centroid, spectral bandwidth, and spectral entropy are extracted from the preprocessed signals to capture essential characteristics of the brain's electrical activity. Additionally, wavelet transformations are applied to the EEG data to obtain time-frequency domain features, including energy and entropy, which further enrich the feature set used for classification.

The dataset consists of annotated EEG recordings,used from CHB-MIT Scalp EEG Database with clearly marked seizure events, enabling supervised training and evaluation of the classifiers.

2 METHODS

2.1 Data Acquistion

The EEG data used in this analysis were obtained from the CHB-MIT Scalp EEG Database. This publicly available dataset consists of EEG recordings from pediatric patients with epilepsy, captured at the Children's Hospital Boston. The dataset includes long-term EEG recordings with annotated seizure events, making it an invaluable resource for developing and evaluating automated seizure detection algorithms.

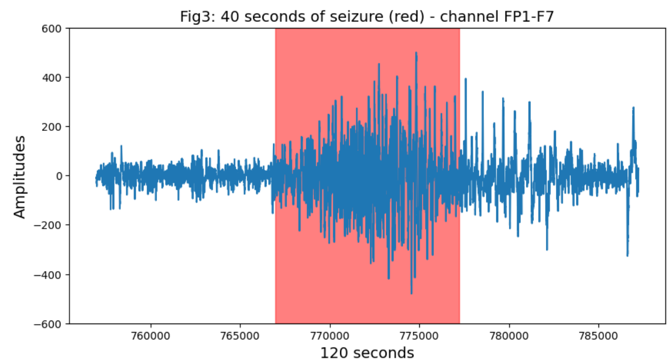

Figure 1: Seizure episode on the FP1-F7 channel

2.2 Preprocessing

To enhance the quality of the EEG data, we performed preprocessing steps, including filtering and artifact removal. A bandpass filter to the raw EEG signals was applied in order to attenuate frequencies outside the range of interest (1-60 Hz). This frequency range was chosen to focus on neural activity while minimizing noise from external sources such as powerline interference and muscle artifacts. The filter design was based on a finite impulse response (FIR) filter [1], which offers linear phase characteristics and stability.

2.3 Power Spectrum Analysis

To characterize the frequency content of the EEG signals, we computed the power spectral density (PSD) using the Welch method. [3] The PSD provides information about the power distribution across different frequency bands, including delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (>30 Hz). The PSD was calculated using the following formula:

$$P_{xx}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-j2\pi fn}\right|^2 \tag{1}$$

where $P_{xx}(f)$ is the power spectral density, N is the number of samples, $x(n)$ is the EEG signal, and f is the frequency.

From the power spectrum, we extracted features such as the dominant frequency, spectral centroid, spectral bandwidth, and spectral entropy. By extracting these features from the power spectrum, we aimed to capture key characteristics of EEG signals related to neural activity and dynamics. These features serve as informative descriptors for subsequent analysis and classification tasks, facilitating the detection and characterization of seizures and other relevant brain events.

2.4 Wavelet Transformation

Wavelet transformation was employed to capture both time and frequency information in the EEG signals. The continuous wavelet transform (CWT) was applied using Morlet wavelets, which are complex sinusoidal waveforms modulated by a Gaussian envelope [2]. The wavelet coefficients were computed as:

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^{*}\left(\frac{t-b}{a}\right) dt \tag{2}$$

where $W(a,b)$ is the wavelet coefficient, $x(t)$ is the EEG signal, ψ^{*} is the complex conjugate of the wavelet function, a is the scale parameter, and b is the translation parameter.

From the wavelet coefficients, we derived features such as energy and entropy to characterize the temporal and spectral properties of the EEG signals.

2.5 Machine Learning Models

We applied various machine learning models, including support vector machines (SVM), logistic regression, and random forests, to classify EEG recordings as normal or epileptic. These models were trained on the extracted features and evaluated using cross-validation.

3 RESULTS

3.1 Wavelet Transform Feature Extraction

- **Cross-Validation Scores with Random forest:**
  - Mean Accuracy: 0.867
  - Standard Deviation of Accuracy: 0.163

3.2 Results for Support Vector Machine (SVM)

Table 1: SVM results for wavelet transform feature extraction

Metric	Value
Accuracy	0.75
Precision	0.5
Recall	1.0
F1-score	0.667
AUC-ROC	0.833

3.3 Power Spectrum Feature Extraction

Table 2: Results for power spectrum feature extraction

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
Logistic Regression	0.624	0.784	0.209	0.330	0.582
Random Forest	0.878	0.920	0.793	0.852	0.869
SVM	0.878	0.920	0.793	0.852	0.869

These results demonstrate the performance of different machine learning models on the EEG data, with wavelet transformation and power spectrum feature extraction methods.

3.4 Discussion

The performance of machine learning models varied depending on the feature extraction method applied to EEG data. For the power spectrum features, the Random Forest and Support Vector Machine (SVM) models achieved similar accuracies of approximately 87.8%. These models also demonstrated high precision, recall, and F1-score values, indicating robust performance in classifying seizure activity.

Compared to the power spectrum-based features, the wavelet transform offered a competitive performance. While the power spectrum-based models achieved an accuracy of 87.8% with the Random Forest classifier, the wavelet-based models achieved a slightly lower accuracy of 86.7%. However, it's essential to note that the wavelet features provide complementary information, particularly in capturing transient changes and localized features in the EEG signals

Future work could explore ensemble methods or deep learning architectures to further improve classification accuracy and robustness, especially in scenarios with complex EEG patterns or a larger number of classes. Additionally, investigating alternative feature extraction techniques or combining multiple modalities, such as EEG and additional physiological signals, could enhance the discrimination power of the models for seizure detection tasks.

REFERENCES

[1] U Rajendra Acharya, Subbaram Naidu Al Sree, and Mookiah Krishnan. 2008. Analysis of EEG signals. *Medical & biological engineering & computing* 46, 9 (2008), 879–887.

[2] Stéphane Mallat. 1989. Wavelet transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 7 (1989), 674–693.

[3] Gert Pfurtscheller and Fernando Lopes Da Silva. 1986. Power spectrum analysis of EEG signals. *Biological cybernetics* 53, 6 (1986), 357–366.