Classifying Seizures in Electroencephalogram Data using Spectral and Fractal features

Nadia Lach-Hab, Yuanzhu Li, Svetlana Pack, Alexander Novak, Yarong Xiao

University of Rochester
StemForAll 2025

August 8, 2025

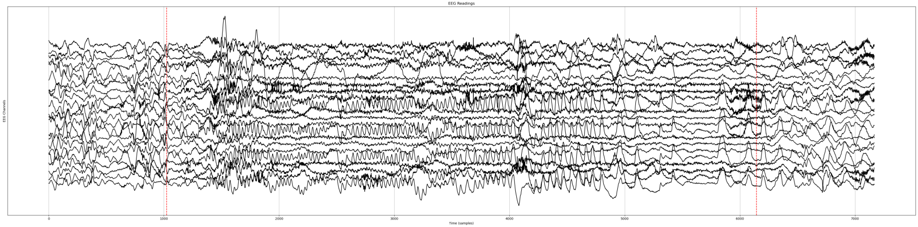


FIGURE: Raw data from a seizure period in the CHB-MIT Database

Seizures

- Sustained abnormal electrical activity ($\geqslant$ 7 seconds)
- Pediatric epilepsy affects 1 in 100 children.
- Diagnosed by visual inspection of brain waves by a clinician
- Automating seizure classification: area of ongoing research

Difficulties of detecting seizures in EEG data

- Seizures are rare (limited training data)
- Size of raw EEG data
  - EEG sampling rate: $f_s = \mathbf{256}$ samples per second
  - EEG electrode count: $d = \mathbf{23}$ electrodes
  - Result: 1 minute of data is in $\mathbb{R}^{d \times n}$, $n = 60 * f_s = \mathbf{15360}$ samples.

Difficulties of detecting seizures in EEG data

- Seizures are rare (limited training data)
- Size of raw EEG data
  - EEG sampling rate: $f_s = 256$ samples per second
  - EEG electrode count: $d = 23$ electrodes
  - Result: 1 minute of data is in $\mathbb{R}^{d \times n}$, $n = 60 * f_s = 15360$ samples.

Key question: How can we exploit the distinct mathematical structure of seizure brainwaves to perform dimension reduction on EEG data?

Past approaches to Seizure Classification in the Literature

Gotman et. al. (1982)[1] and Shoeb et. al. (2010)[2]

- Classifying seizures using limited bandwidth of **0** - **30** Hz Fourier components of EEG data
  - Earliest seizure detecion algorithms
  - Focus on only biologically relevant frequencies

Past Work on Seizure Detection

Past approaches to Seizure Classification in the Literature

Gotman et. al. (1982)[1] and Shoeb et. al. (2010)[2]

- Classifying seizures using limited bandwidth of 0 - 30 Hz Fourier components of EEG data
  - Earliest seizure detection algorithms

Namazi et. al. (2015)[3]

- Classifying seizures using the Hurst exponent
  - EEG data might have **higher** fractal dimension during seizures

Definition

The **discrete s-energy** *of a finite dataset* $\{x_k\}_{k=1}^n \subset \mathbb{R}^d$ *for parameter* $s > 0$ *is defined by*

$$I_s\left(\{x_k\}_{k=1}^n\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}|x_i - x_j|^{-s}$$

where $|.|$ *denotes the Euclidean norm on* $\mathbb{R}^d$*.*

StemForAll 2024

- Discrete energy → estimate EEG fractal dimension
- Feedforward neural network trained on discrete energy differences: **86.79%** test accuracy

Compare three approaches to classifying seizures from EEG data in pediatric epilepsy patients using models trained on:

- Full raw EEG recording ($\mathbb{R}^{23 \times n}$)
- Biologically relevant (0-30Hz) frequency component time series
  - Compress further using Principal Components Analysis
- Discrete energy time series ($\mathbb{R}^{\frac{n}{256}}$)

Children's Hospital Boston MIT Scalp Electroencephalogram Data (CHB-MIT) Database

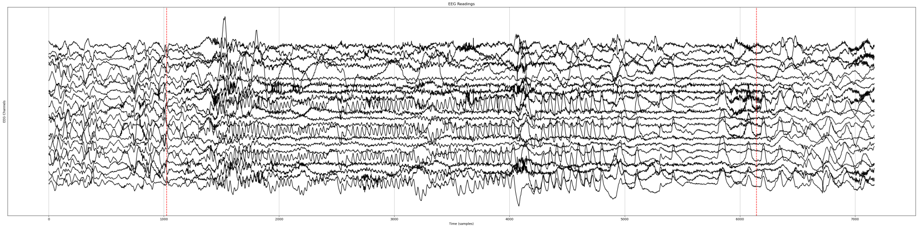
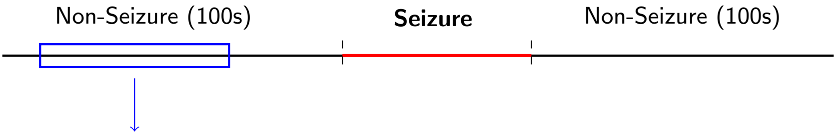


FIGURE: Raw data from a seizure period in the CHB-MIT Database

Patient Demographics

- 23 pediatric patients
- Ages 1.5 - 22 years
  - 4 patients had age $\leqslant 3$
  - Neonatal seizures were excluded

- Model: Long Short-Term Memory (LSTM)
- Dataset: 1 EEG file (23 channels) from each patient
- Excluding: patients under 4 years old, and seizures shorter than 7 seconds



Sliding windows for collecting training dataset

- Test accuracy: **87%**
- Limitation: only 1 EEG file from each patient is used for training due to limitation of computers, so it might not be generalizable to other EEG files.

- Applied Fourier Transform to convert raw EEG data from the time domain to the **frequency** domain.
- Observation: For both seizure and non-seizure plots, most of the signal power is concentrated in the **0–30 Hz** range.
- Conclusion: Focus on low frequencies.

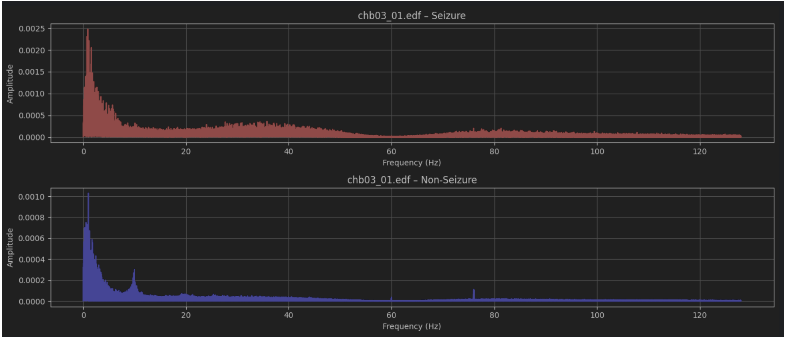


FIGURE: FFT magnitude spectrum for seizure (red) and non-seizure (blue) EEG segments

- Applied Fourier Transform to EEGs using a **1-second sliding window**, aligned with seizure onset time.
- Observation: In some patients, there is a clear **rise** of normalized Fourier transform data during the seizure.

Figure: Normalized low-frequency Fourier transform over time.

- **Goal**: Use Fourier Transform features (0–30 Hz) to classify seizure vs. non-seizure EEG segments.
- **Features**: Low-frequency Fourier data across common channels, reduced by PCA ($n = 7$).
- **Classifier**: Random Forest with balanced class weights.
- **Results**:
  - Accuracy: **86.8%**
  - Seizure class F1-score: **0.69**
  - Non-seizure class F1-score: **0.92**

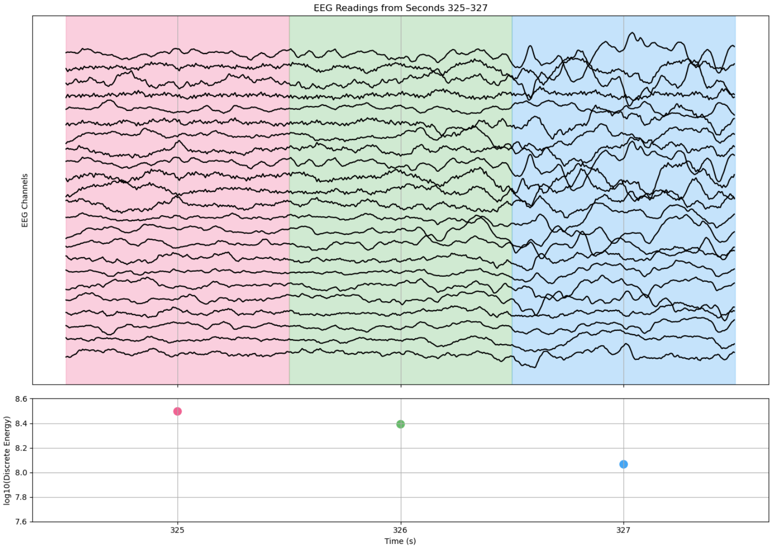


Figure: Finding the discrete energy for each second of EEG readings. This process reduces the dimension of the data by a great degree, since each data point produced will be the result of 256 time steps, each with 23 features

Reducing Dimensionality with Discrete Energy

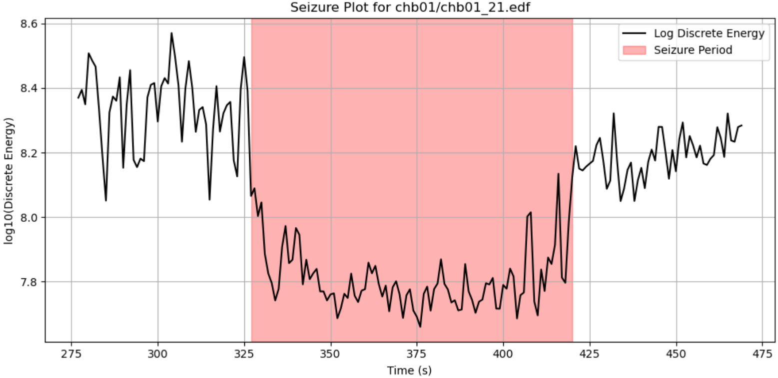
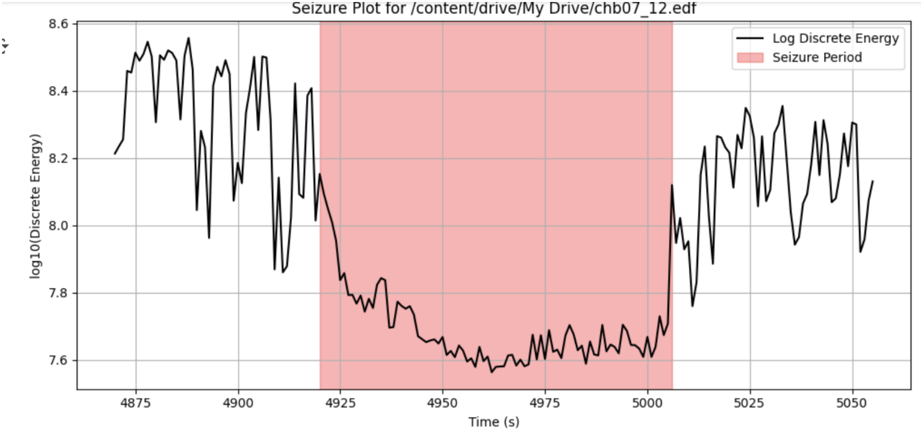


Figure: Discrete energy often drops during seizure periods (highlighted in red)

- **Training dataset**:
  - Datapoints in the **seizure group**: the discrete energy over time (calculated for each second) during each full seizure period as well as a random buffer of timesteps before and after the seizure
  - Datapoints in the **non-seizure group**: the discrete energy over time (calculated for each second) for random periods of time that were not during seizures
- **Goal**: to classify whether a given discrete energy time series contains a seizure period or not

- Model performance was tested with 5-fold cross validation
- **Classification performance**:
  - Overall accuracy of **87%**
  - Detected **90%** of discrete energy time series with seizures and misclassified **16%** of non-seizure time series
- **Conclusion**: despite the input data being reduced to a 1-dimensional time series with 256 times fewer timesteps, enough information was retained that a model was still able to predict the presence of seizures with relatively good accuracy

Computing integrals of the Seizure Discrete Energy Plots




```
Estimated seizure integral (area under curve): 661.3600926832951
Estimated preictal integral (area under curve): 713.6737973519548
Estimated postictal integral (area under curve): 694.7157979924451
```

FIGURE: The discrete energy integral from patient 7 when s=1. Note apparent differences in pre-ictal, post-ictal, and seizure integrals

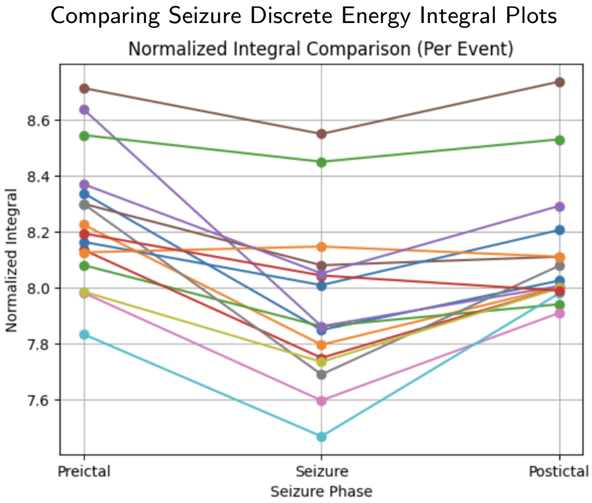


Figure: Discrete energy integrals of 16 patients during pre-ictal, post-ictal, and seizure periods. Note apparent decline in discrete energy during seizure periods.

Comparing the Discrete Energy of Non-Epileptic and Epileptic Patients

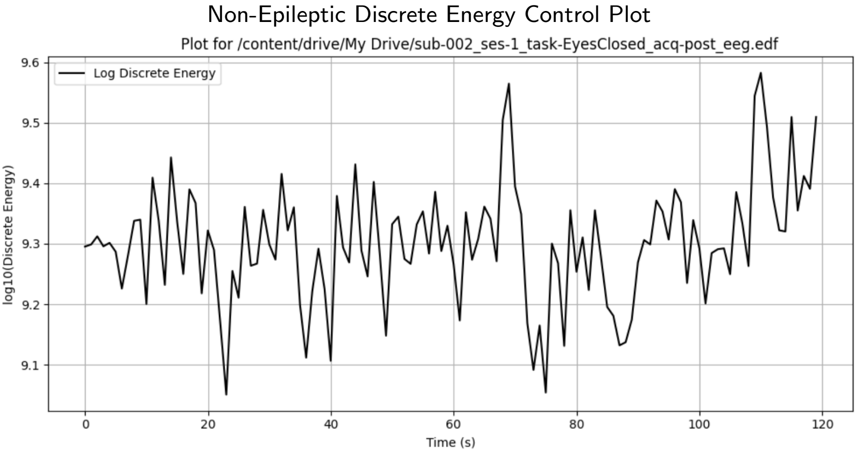


Figure: Discrete energy plot of healthy, non-epileptic patient. Note apparent difference in discrete energy range for healthy and epileptic patients.

This feedforward neural network (FNN) uses 9 ratio-based features derived from discrete energy integrals. The model was trained on standardized ratio features capturing relative energy shifts between seizure states, enabling it to learn discriminative patterns. It uses a simple 2-layer architecture trained with cross-entropy loss.

```
Total Test Accuracy: 0.8636
Class 0 Accuracy: 1.0000
Class 1 Accuracy: 0.6000
Class 2 Accuracy: 0.8333
Class 3 Accuracy: 1.0000

Predicted labels: [1, 0, 0, 3, 3, 2, 0, 0, 1, 1]
True labels:      [1, 0, 0, 3, 3, 2, 0, 1, 1, 1]
```

Figure: FNN accuracy and predicted classification. Class 0 is the seizure ratios, class 1 is the pre-ictal ratios, class 2 is the post-ictal ratios, and class 3 is the no seizure ratios.

Future Directions

- Identifying start time of seizure (sliding window): the first interval being classified as seizure is the start time
- Investigating why some patients' seizures do not show a decline in discrete energy when setting the s parameter to 1, but do for higher values of s.
- Creating a classification neural network that combines both discrete integrals and differences to detect seizures.
- Researching outliers such as neonatal seizure data further to see if we can draw similar patterns/comparisons as we did for the pediatric data.

Ackknowledgements

We would like to acknowledge Alex Iosevich and the other StemForAll instructors whose volunteering makes this project possible.

We would also like to acknowledge the Children's Hospital Boston and MIT for collecting the data the CHB-MIT EEG Scalp database, and making it publicly available under a Open Data Commons Attribution License v1.0

Thank you!

Questions?

Gotman, J. 1982. "Automatic Recognition of Epileptic Seizures in the EEG." Electroencephalography and Clinical Neurophysiology 54 (5): 530–40. https://doi.org/10.1016/0013-4694(82)90038-4.

Shoeb, Ali H., and John V. Guttag. 2010. "Application of Machine Learning to Epileptic Seizure Detection." International Conference on Machine Learning, June, 975–82. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.170.923&rep=rep1&type=pdf.

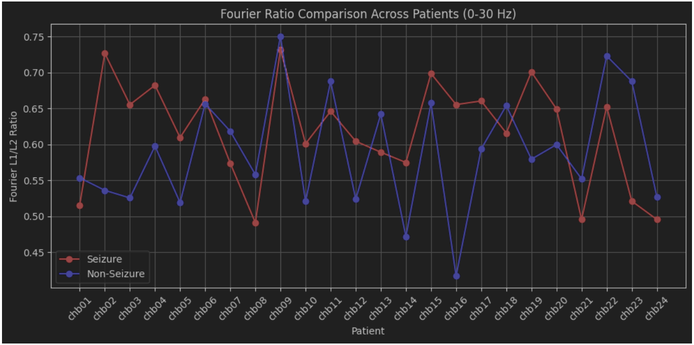
Namazi, Hamidreza, Vladimir V. Kulish, Jamal Hussaini, Jalal Hussaini, Ali Delaviz, Fatemeh Delaviz, Shaghayegh Habibi, and Sara Ramezanpoor. 2015. "A Signal Processing Based Analysis and Prediction of Seizure Onset in Patients With Epilepsy." Oncotarget 7 (1): 342–50. https://doi.org/10.18632/oncotarget.6341.

Guttag, John. "CHB-MIT Scalp EEG Database" (version 1.0.0). PhysioNet (2010), https://doi.org/10.13026/C2K01R.

Appendix

- Computed the ratio of the L1 norm to the L2 norm of the fourier data (0–30 Hz).
- Observation: Although seizure and non-seizure ratios do not always show a large difference, the ratios range from **0.45 to 0.75** across patients.
- Conclusion: EEG time series are highly **nonforecastable**. Further there is no clear correlation between the relative degree of forecastability between seizure and non-seizure time series

$$\text{Fourier Ratio} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{f}_i\right|}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{f}_i\right|^2}}$$



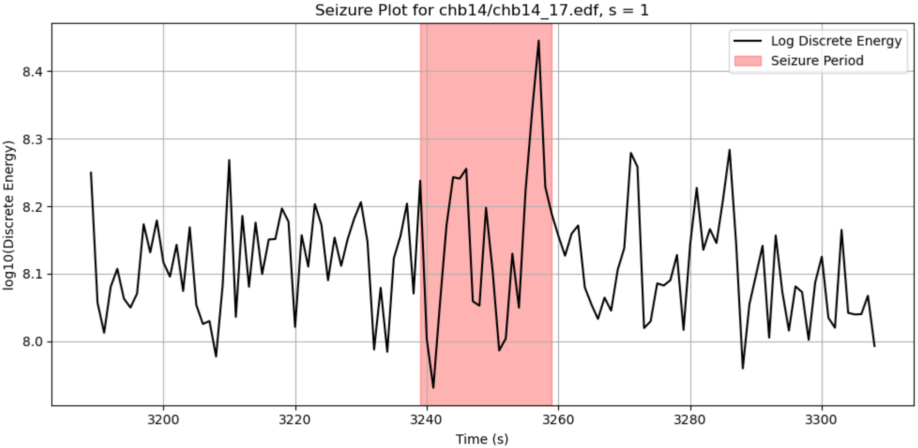


Figure: The discrete s-energy of EEG 17 from patient 14 when s=1. Note apparent lack of change during the seizure duration (highlighted in red)

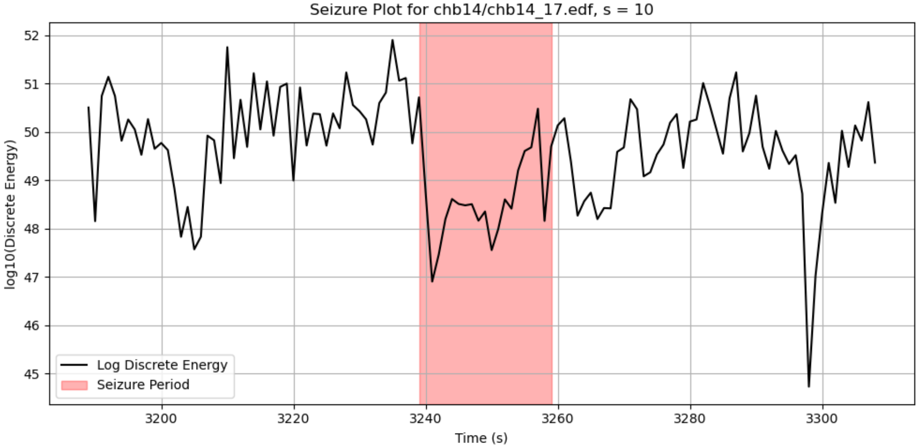


Figure: The discrete s-energy of EEG 17 from patient 14 when s=10. The sustained decrease in energy during the seizure duration (highlighted in red) is now visible