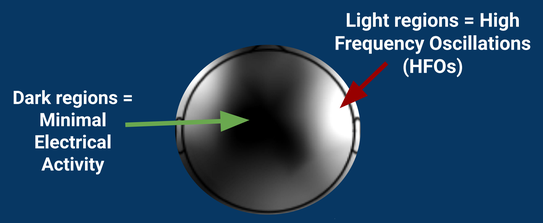


Sample topographical image and how to analyze it (reference for upcoming sections)

Using Convolutional Neural Networks to Determine the Impact of High-Frequency Oscillations in Epileptic Pediatric Patients

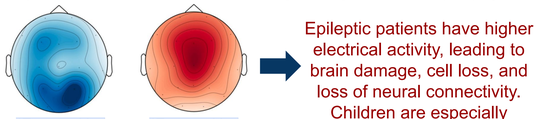
Sneha Sharma

Acknowledgements
Mentor: David DiStefano, Department of Psychology, Tufts University

Credits
All images, unless otherwise noted, have been created by Sneha Sharma. Full list of references for this project can be found in the project binder.

INTRODUCTION

Epilepsy is a **neurological disorder** that causes patients to get anywhere from less than one seizure per year to several daily. **75%** of epilepsy cases begin in childhood, but almost **14%** of people with epilepsy go undiagnosed in their lifetime. This means a substantial part of the global population may face seizures throughout their entire life, but never get diagnosed for epilepsy.

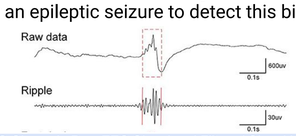


Topographical Image of Person Without Epilepsy | *Topographical Image of Person With Epilepsy*

Source: Zijlmans et al.

Epileptic patients have higher electrical activity, leading to brain damage, cell loss, and loss of neural connectivity. Children are especially susceptible, as they have underdeveloped brains.

High-frequency oscillations (HFOs) are a biomarker for epilepsy. These are abnormal electrical signals in the brain within the frequencies of **80 to 250 Hertz** and are typically observed in electroencephalography (EEG) scans. My research focused on HFOs, as they are one of the few biomarkers for epilepsy that occur during *and* in-between seizures, meaning that researchers don't have to wait for an epileptic seizure to detect this biomarker.



HFO Shown in Raw EEG Signal and Filtered Signal in the Ripple Band (From 80 to 200 Hertz). Source: Zuo et al.

PROBLEMS + RESEARCH QUESTION

Clinical Problem

Few efficient diagnosis methods for pediatric epilepsy.

- Current methods are expensive, time-consuming, and not accurate (an EEG scan in the USA costs ~$1000)

Scientific Problem

Little research on age's influence on epilepsy diagnostics.

- Researchers don't know how the spectral frequency and location of HFOs change with age

Tech Problem

No formal way to analyze EEG data.

- EEG data is very high-dimensional, so conventional methods (DBSCAN, K-nearest-neighbors) do not work well

Research Question

How can a machine learning model analyze EEG data to determine the impact of age on the spectral frequencies and locations of HFOs in pediatric epileptic patients?

DATASET DESCRIPTION

The dataset used in this project was sourced from an publicly-accessible dataset platform. The data was collected by **Dorottya Cserpán et al.** in 2021. In this study, researchers took scalp EEG readings at 3-hour intervals* from **30 pediatric epileptic patients** undergoing non-rapid eye movement (NREM) sleep. By taking measurements during NREM sleep, the doctors reduced the number of artefacts in the data, meaning that I did not have to account for artefacts during pre-processing steps.

**Patient 12 only had a 2-hour recorded interval*

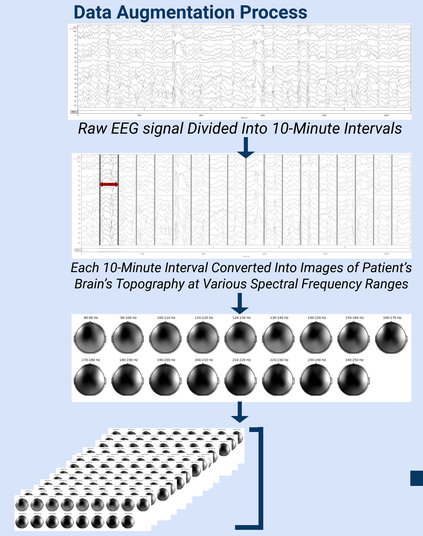
METHODOLOGY

Step 1: Data Collection + Processing

Original Dataset Description: Scalp EEG readings of 30 pediatric epileptic patients (aged from 0.7 to 17.4 years old). Because the dataset size was so small, I had to use data augmentation to increase the number of samples.

Augmented Dataset Description: 535 data points, each with 17*125*125 = **265,625 features, which is a lot**. I needed to reduce the dimensionality of my dataset to find patterns.

Data Augmentation Process



Raw EEG signal Divided Into 10-Minute Intervals

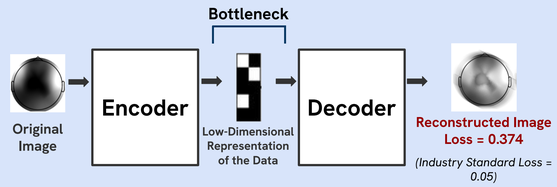
Each 10-Minute Interval Converted Into Images of Patient's Brain's Topography at Various Spectral Frequency Ranges

$$[17_{\#\text{of images at frequency bands}} \times 125_{\#\text{of vertical pixels}} \times 125_{\#\text{of horizontal pixels}}]$$

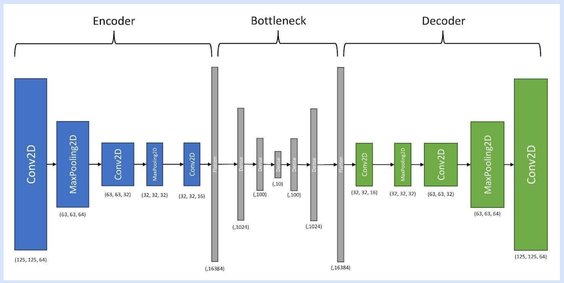
Step 2: ML Model for Dimensionality Reduction

I needed to reduce the dimensionality of each data point, so I decided to use an **autoencoder**. An autoencoder is the only architecture that would work for this data, as the data is unlabeled and is too high-dimensional to find meaningful patterns as is.

Typical Autoencoder Architecture: uses **convolutional** and **fully-connected layers** and **mean-squared-error loss** to compress and recreate an image, as seen below. A typical autoencoder, however, wasn't able to handle the amount of noise present in this small dataset; it got lost in the irrelevant details.

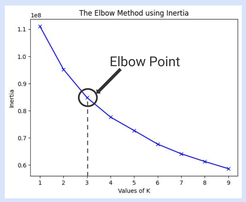


Novel Autoencoder Architecture: added MaxPooling layers, used Batch Normalization and Leaky ReLu. Also, a standard loss function can't ignore the noise in the images' white backgrounds, so I custom loss function focused on decreasing loss within the relevant parts of the image. These changes yielded an autoencoder that took the 265,625 features in each data point and represent them with only **10 features at a loss of 0.0029**.



Architecture of my autoencoder. The encoder compresses the 265K features. The bottleneck further compresses it into 10 features. The decoder ensures that the 10 features are representative of the data point.

Step 3: Pattern Detection with K-Means



I wanted to find patterns within the smaller representation of the data. I decided to use the KMeans clustering algorithm due to its robustness and success in similar research. In order to determine the optimal number of clusters in the data, I used an Elbow Plot, which that that I should use **k = 3 clusters**.

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} |x - m_i|^2 \qquad WCSS(K) = \sum_{i=1}^{K} \sum_{x \in C_i} ||x - \mu_i||^2$$

KMeans Clustering Algorithm Formula (right) and Elbow Plot Algorithm Formula (left)

Step 4: Analysis

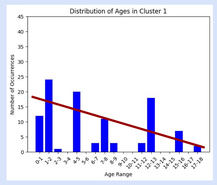
I used two analysis methods to find patterns: **(1)** to find age-related variations, I graphed the number of data points per age range per cluster; **(2)** to find spectral-frequency-related variations, I created plots of the topographies of each cluster's central data point.

RESULTS

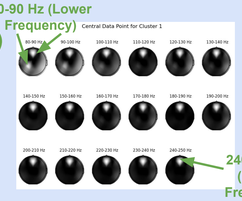
Below are the results of my visualizations for Clusters 1 and 2*

Cluster 1:

(1)



(2)

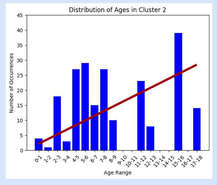


(1) Number of patients in cluster 1 decreased with age. **(2)** More high-frequency oscillations (HFOs) at lower frequencies compared to higher frequencies. Also, I determined that 89% of young data points that had HFOs in high frequencies also had them in low frequencies, but the reverse wasn't true.

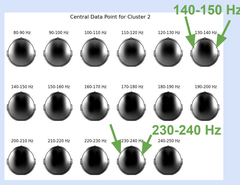
*Cluster 3 visualizations are shown in research paper and project binder

Cluster 2:

(1)



(2)

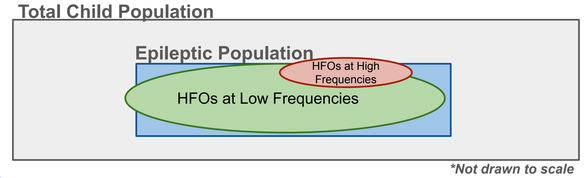


(1) Number of patients in cluster 2 increased with age. **(2)** HFOs are consistent across all spectral frequencies. All of the topography images, from 80 to 250 Hertz, had HFOs in the same regions of the brain, meaning that all older children had consistent HFOs.

RESULTS CONT.

The visualizations of the clusters show that HFOs at lower spectral frequencies are better predictors of epilepsy in children. The visualization below summarizes my results:

1. More epileptic children **experience HFOs at lower frequencies** in comparison with HFOs at higher frequencies
2. HFOs at lower frequencies encompass most of the HFOs at higher frequencies



**Not drawn to scale*

CONCLUSION

Focus on Lower Frequencies for Children's HFOs:

- Looking for HFOs from **80 to 140 Hz** will lead to **lower false-negative** and **false-positive** rates across all ages
- Don't have to search through such a broad range of frequencies anymore → **can build cheaper EEG machines** to detect epilepsy

Autoencoders are Powerful to Remove Noise in High-Dimensional Data

- Allowed me to focus on the data's most important features

Current Limitations

- I only had 30 patients, so I **couldn't get statistically significant results**

FUTURE DIRECTIONS

Step 1: Generalize Findings to All Children

Currently speaking to researchers in the Seattle Children's Research Hospital to get access to their past data and **expand my dataset**.

Step 2: Introduce in Clinical Settings

After my results have been generalized to the full pediatric population, they will help doctors search for HFOs at smaller frequencies, which means that **less powerful EEG machines** (like the EMOTIV Insight) **will be needed** to detect HFOs.

Step 3: Build Custom EEG Headset to Detect HFOs

Ultimate goal is to work with companies to **build an EEG headset that's optimized to detect HFOs**. Because the headset would only need to look for HFOs at a lower frequency, it would have the **potential to be significantly cheaper than current tools**.

KEY REFERENCES

Cserpán, Dorottya, Ece Boran, Santo Pietro Lo Biundo, Richard Rosch, Johannes Sarnthein, Georgia Ramantani. "Scalp High-Frequency Oscillation Rates Are Higher in Younger Children." Brain Communications 2 no. 3.

Gliske, Stephen V., Zachary T. Irwin, Cynthia Chestek, Garnett L. Hegeman, Benjamin Brinkmann, Oren Sagher, Hugh J. L. Garton, Greg A. Worrell, and William C. Stacey. 2018. "Variability in the Location of High Frequency Oscillations during Prolonged Intracranial EEG Recordings." Nature Communications 9 (1): 2155.

Gramfort, Alexandre, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. 2013. "MEG and EEG Data Analysis with MNE-Python." Frontiers in Neuroscience 7 (267): 1–13.

Jacobs, Julia, and Maeike Zijlmans. 2020. "HFO to Measure Seizure Propensity and Improve Prognostication in Patients With Epilepsy." Epilepsy Currents 20 (6): 338–47.

Zijlmans, Maeike, Premysl Jiruska, Rina Zelmann, Frans S.S. Leijten, John G.R. Jefferys, and Jean Gotman. 2012. "High-Frequency Oscillations as a New Biomarker in Epilepsy." Annals of Neurology 71 (2): 169–78. https://doi.org/10.1002/ana.22548.

Paper bibliography contains full list.