Project Report: Predicting ADHD with EEG Signals

Georgie Payne

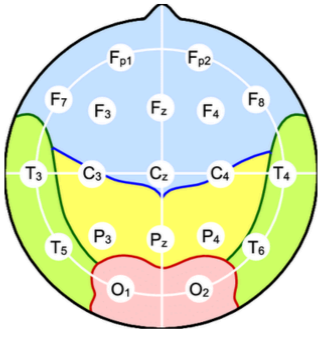
I. Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a very common behavioral disorder marked by patterns of inattention, hyperactivity, and impulsivity. Early diagnosis of ADHD is important for children because it allows them to avoid setbacks, especially in school, if they can come up with a way to manage symptoms through things like medication or ways to control their environment. Currently, diagnosing ADHD relies on subjective evaluations like self-reports or observations and an interview. While this works for the most part, having a nonsubjective, definitive test would be helpful for an accurate diagnosis. Because ADHD affects how the brain works, fMRI and EEG have been looked into as tools that may help identify it; however, neither has proven accurate enough to fit the criteria for a sole diagnostic tool.

Still, results using these devices have been improving over time as more research has been conducted and new machine-learning techniques have been developed, which means they have the potential to become a way to diagnose ADHD. EEG would be a more practical option than fMRI due to its cost and accessibility. Therefore, I used EEG recordings for my project to predict ADHD and research more about the disorder. My goals were to use a classification model to predict if a child has ADHD based on their EEG signals, identify features on an EEG that might be a marker for ADHD, and explore subgroups within the ADHD population that may represent distinct presentations of the condition.

II. Dataset

The data I used was from a public dataset on IEEE DataPort that contained the EEG recordings from 121 children, ages 7-12, where 61 of them had been diagnosed with ADHD. The other 60 were children with no history of psychiatric disorders, epilepsy, or high-risk behaviors. The recordings were conducted with electrodes on 19 channels according to the 10-20 standard at a 128 Hz sampling frequency. For the duration of the recordings, the children were provided with a visual attention task that asked them to count cartoon characters on a screen. The recording



Electrode placement for 10-20 system

was stopped when they provided an answer, which means the length of their recording was determined by their performance in the game.
https://ieee-dataport.org/open-access/eeg-data-adhd-control-children

III. Preprocessing

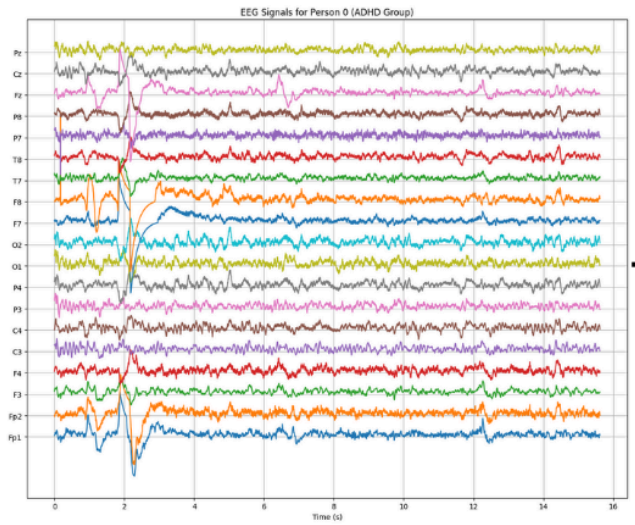
Preprocessing was an essential part of the process because EEG recordings do not exclusively capture brain activity; they also easily pick up noise from other movements in the body or external noise from the device or environment. Therefore, preprocessing is mainly concerned with separating brain signals of interest from other noise. The first step of the preprocessing is converting the signals for each person into MNE RawArray objects so that I could use functions from the MNE package to handle operations like artifact detection and apply filtering.

Once the RawArray objects were created, a high-pass filter was applied to the data with a low-frequency cutoff of 0.1 Hz and a high-frequency cutoff of 50 Hz. Applying this filter makes it easier for Independent Component Analysis (ICA), which I use next, to identify independent sources by removing extremely low drifts. After that, ICA was used to separate brain signals from artifacts, which was a function provided by the MNE preprocessing package. The ICA was run with n_components set to 19, matching the number of EEG channels, which ensures all the data is included and helps identify patterns specific to each channel. Next, the find_bads_eog function was used to detect components linked to eye movement artifacts, and these components were excluded before applying the ICA to clean the data. The frontal channels 'Fp1' and 'Fp2' were used to select these bad EOG components because they are located near the eyes and are most sensitive to eye movements.

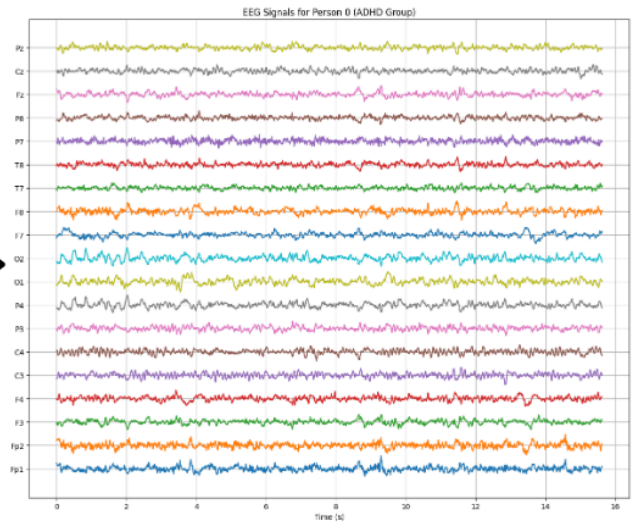
Following the ICA, a bandwidth filter was applied to remove all frequencies outside the 1 Hz to 40 Hz range. This range captures all the frequencies that should be related to brain activity during a visual attention task, including Theta, Alpha, and Beta waves. The bandpass filter also removed the powerline noise coming from the device at 50 Hz, which was present in the recordings of each participant. After filtering the data, I also re-referenced the EEG to the average across channels, which provides a baseline reference for all channels. This ensures that no electrodes dominate others and also cancels out any noise that is shared between all channels.

The final step was epoching the data into three-second intervals, which allows for the analysis of smaller, time-locked segments of the EEG. I found that three seconds gave me the best performance on my models because they were long enough to capture meaningful data but didn't capture too much variation. The auto-reject library was used on these epochs to detect and remove any epochs that might contain irregularities or noise. Once the data was cleaned, a new data frame was created that included the EEG signals for each person on every epoch. The labels at this point were the values for each channel, a person ID, an epoch number, and the person's group (0 for control, 1 for ADHD). The images below show the data before and after the preprocessing of the first subject.

Before

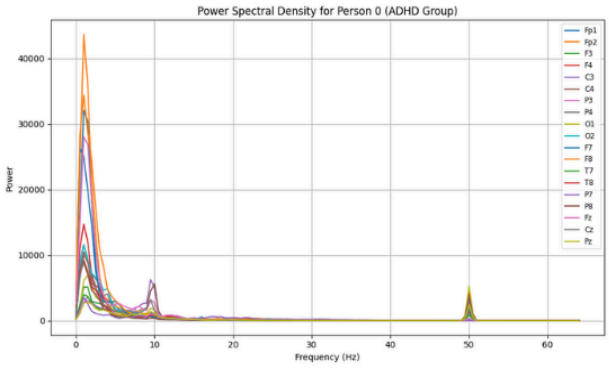


After

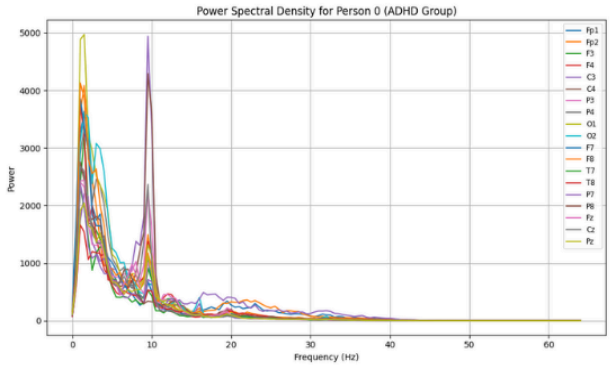


Activity across all channels before and after preprocessing

Before



After



Power Spectral Density plot of all channels before and after preprocessing

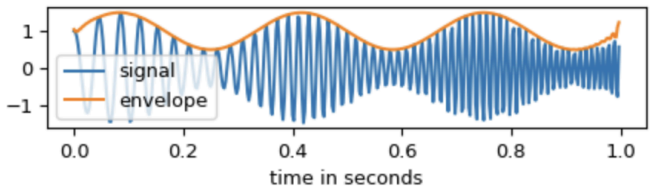
The first side-by-side illustrates the removal of noise that was resulting in messy signals. Particularly at around two seconds, there looks to be some kind of eye blink or other movement that the EEG recorded but was removed in the preprocessing. The PSD plots reveal a lot of low-frequency noise, which was filtered out by the bandpass filter. The powerline noise at 50 Hz is an example of external noise that was filtered out due to the filter. The channels in the second PSD plot also appear much more consistent with each other, reflecting less influence from channel-specific artifacts like eye blinks.

IV. Feature Engineering

Selecting features to train the models on also required a lot of careful thought because raw EEG data is extremely high-dimensional and not immediately interpretable, making it necessary to identify specific patterns or metrics in the data that may differentiate groups. Therefore, I extracted several features from the EEG signals, some of which are already considered biomarkers for people with ADHD, like Theta to Beta ratio. Each of the features I extracted was calculated for each person on every single epoch.

First, I calculated the power spectral density and got the frequency bands for Theta (4-8 Hz), Alpha (8-12 Hz), and Beta (12-30 Hz) frequencies on each individual channel. Then, using the Theta and Beta power values, I calculated the Theta-to-Beca ratio (TBR) for every channel. This feature is important because people with ADHD tend to have a higher TBR due to elevated Theta waves (associated with relaxation) and decreased Beta waves (associated with focused attention). These Frequency-Domain Features capture the distribution of power across different frequency bands.

The temporal features I used to capture brain activity over time were amplitude envelope and variation. The amplitude envelope, obtained through the Hilbert Transform, is a measure of the signal amplitude over time, and I took the mean and standard deviation of this envelope as features for each channel. The picture to the right is a visual representation of how the amplitude



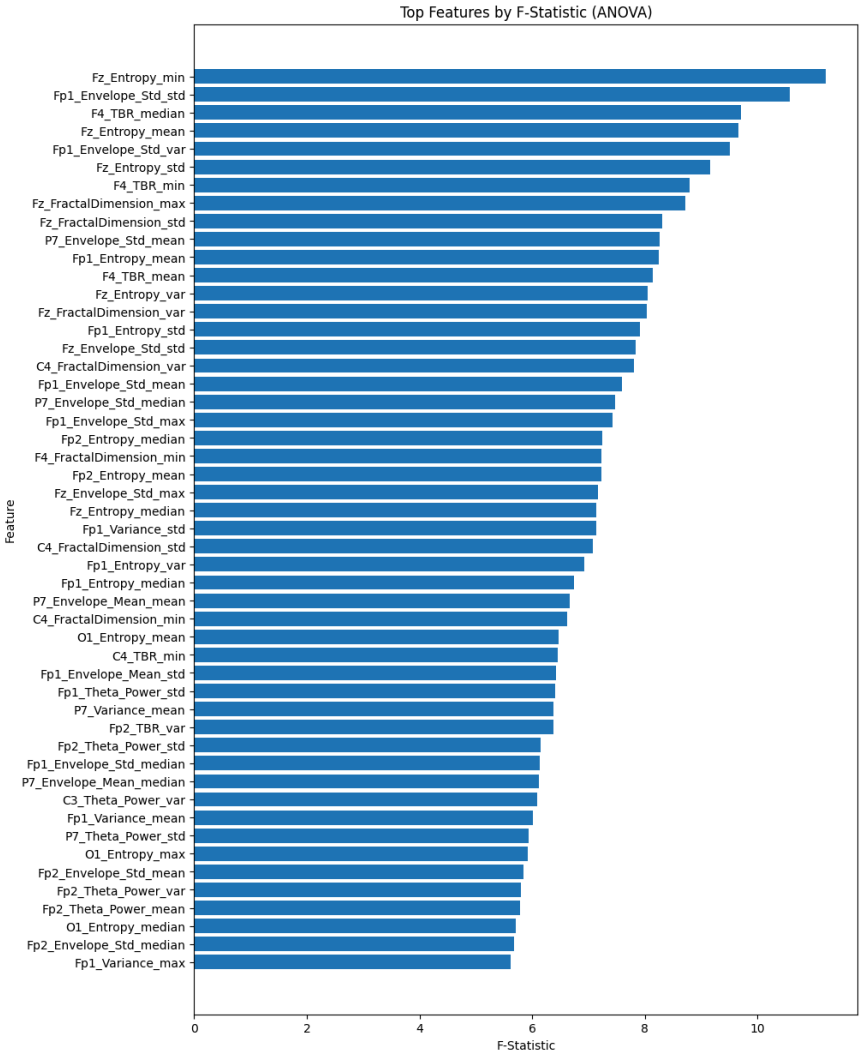
envelope works. I also took the variance of the raw signals on each channel to measure the signal fluctuations over time.

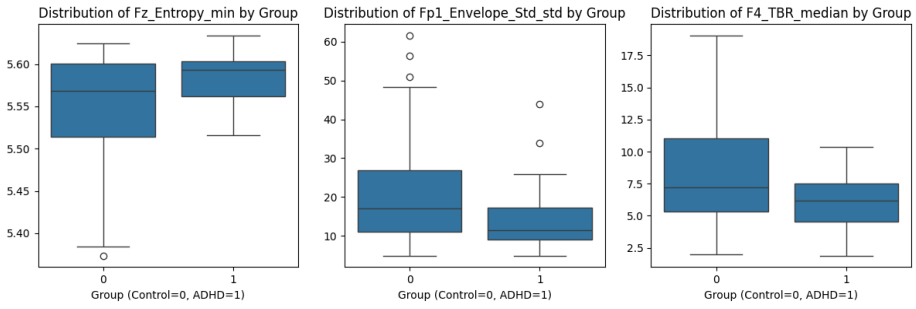
Fractal dimension and entropy were then calculated as a measure of epoch complexity on each channel. The fractal dimension, obtained from Detrended Fluctuation Analysis (DFA), is a measure of the complexity or self-similarity of the signal over time. The entropy focuses on randomness and unpredictability of signals. Finally, since TBR is very relevant to ADHD, specifically in the frontal lobe, for each epoch, I took the mean TBR across all frontal lobe channels as another feature.

This left me with a dataset of epochs containing all of the above features. Previously, for this project, I had split that dataset into a train/test set for model training, which meant my models were predicting whether each epoch belonged to a person with or without ADHD, but this didn't allow me to classify actual people, which was the goal. Therefore, what I have done is aggregate the epochs back together for each person by taking a few different measures of the epoch features over time. These were mean, median, variance, standard deviation, minimum, and maximum. After this, there was a single entry for each person in the data frame where each column was a measurement over time of one of the features described above, for example, Fp1_Theta_Power_var (variance in theta power over all epochs).

V. Feature Selection

After extracting all the features, I performed an analysis to explore the most important features and drop any that were unimportant. For this, I used both ANOVA and Lasso Regression. In a previous version of this project, I applied both methods to the entire dataset as opposed to a training subset. This mistake caused data leakage, where information from the test set influenced the feature selection process and resulted in models that might have generalized poorly to unseen data. I fixed this problem by creating a split first and then performing ANOVA and lasso regression on just the training set. I performed ANOVA first and dropped all features with a p-value less than 0.05 to eliminate statistically insignificant features, leaving me with 121 features for each person. Then, to get an idea of which features may be the most significant markers for ADHD, I made a graph of the top 50 features, ranked by F-statistic. I also made boxplots for the top 6 features with the highest F-statistic.


Top Features by F-Statistic (ANOVA)
Fz_Entropy_min
Fp1_Envelope_Std_std
F4_TBR_median
Fz_Entropy_mean
Fp1_Envelope_Std_var
Fz_Entropy_std
F4_TBR_min
Fz_FractalDimension_max
Fz_FractalDimension_std
P7_Envelope_Std_mean
Fp1_Entropy_mean
F4_TBR_mean
Fz_Entropy_var
Fz_FractalDimension_var
Fp1_Entropy_std
Fz_Envelope_Std_std
C4_FractalDimension_var
Fp1_Envelope_Std_mean
P7_Envelope_Std_median
Fp1_Envelope_Std_max
Fp2_Entropy_median
F4_FractalDimension_min
Fp2_Entropy_mean
Fz_Envelope_Std_max
Fz_Entropy_median
Fp1_Variance_std
C4_FractalDimension_std
Fp1_Entropy_var
Fp1_Entropy_median
P7_Envelope_Mean_mean
C4_FractalDimension_min
O1_Entropy_mean
C4_TBR_min
Fp1_Envelope_Mean_std
Fp1_Theta_Power_std
P7_Variance_mean
Fp2_TBR_var
Fp2_Theta_Power_std
Fp1_Envelope_Std_median
P7_Envelope_Mean_median
C3_Theta_Power_var
Fp1_Variance_mean
P7_Theta_Power_std
O1_Entropy_max
Fp2_Envelope_Std_mean
Fp2_Theta_Power_var
Fp2_Theta_Power_mean
O1_Entropy_median
Fp2_Envelope_Std_median
Fp1_Variance_max
Feature
0
2
4
6
8
10
F-Statistic


Distribution of Fz_Entropy_min by Group
Distribution of Fp1_Envelope_Std_std by Group
Distribution of F4_TBR_median by Group
Distribution of Fz_Entropy_mean by Group
Distribution of Fp1_Envelope_Std_var by Group
Distribution of Fz_Entropy_std by Group
Group (Control=0, ADHD=1)

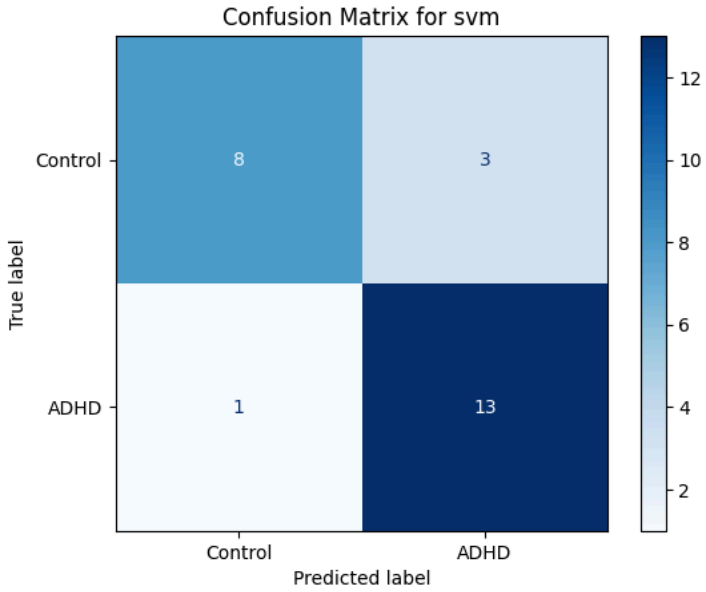
Then, to further reduce the number of features, I performed lasso regression on the remaining features, which left me with 32 features for each person.

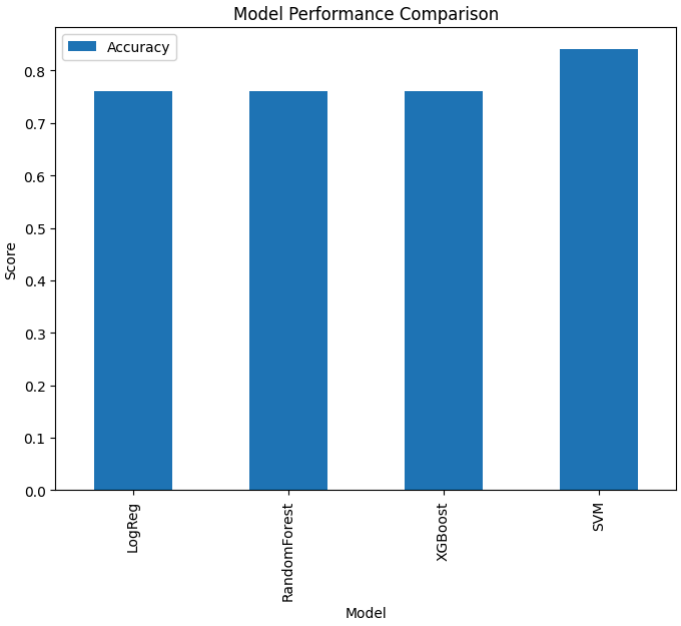
Selected Features:

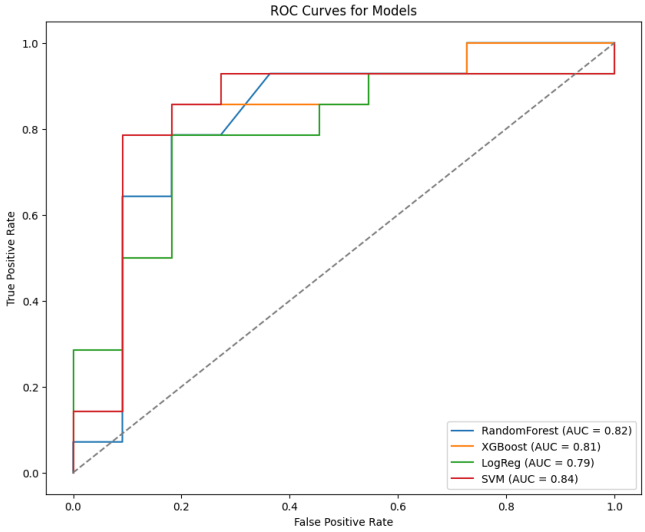
	Feature	Coefficient
61	C4_FractalDimension_var	0.212155
98	P7_Entropy_median	0.125975
50	F4_FractalDimension_median	0.108062
100	Fz_Envelope_Std_var	0.068962
73	O1_Entropy_max	0.068337
97	P7_Variance_median	0.057674
112	Fz_Entropy_max	0.055355
35	Fp2_Envelope_Std_std	0.051026
113	Fz_Entropy_median	0.049732
111	Fz_Entropy_min	0.048013
3	Fp1_Theta_Power_max	0.043190
71	O1_Envelope_Std_mean	0.040329
60	C4_TBR_median	0.038231
83	P7_Theta_Power_max	0.034728
25	Fp1_Entropy_median	0.025806
72	O1_Entropy_mean	0.022671
0	Fp1_Theta_Power_mean	0.011595
77	F7_Theta_Power_mean	0.007238
104	Fz_Variance_max	0.002818
45	F4_TBR_mean	0.001495
76	O2_Envelope_Std_median	-0.004723
36	Fp2_Envelope_Std_median	-0.009040
22	Fp1_Entropy_var	-0.010792
12	Fp1_Envelope_Std_std	-0.034019
55	C3_TBR_var	-0.051547
30	Fp2_TBR_var	-0.066014
49	F4_FractalDimension_min	-0.075758
67	P4_TBR_var	-0.086395
91	P7_Envelope_Std_min	-0.086669
46	F4_TBR_min	-0.095662
52	C3_Theta_Power_var	-0.118595
106	Fz_FractalDimension_std	-0.125142

VI. Models and Results

Next, I worked on fitting models that could predict whether a person had ADHD based on the features selected. Since there were only 32 features remaining for each person after ANOVA and Lasso, I did not further reduce the dimension with principal component analysis (PCA), as I previously did before correcting my mistake causing data leakage.. Using those 32 features, I trained four models to find the algorithm that performed best on my data. The ones I applied were logistic regression, random forest classifier, XGBoost, and support vector machines (SVM). SVM performed the best of these, with an accuracy of 84%, while the other three had an accuracy of 76%. The visual comparison of these models is shown below, along with the confusion matrix for SVM.



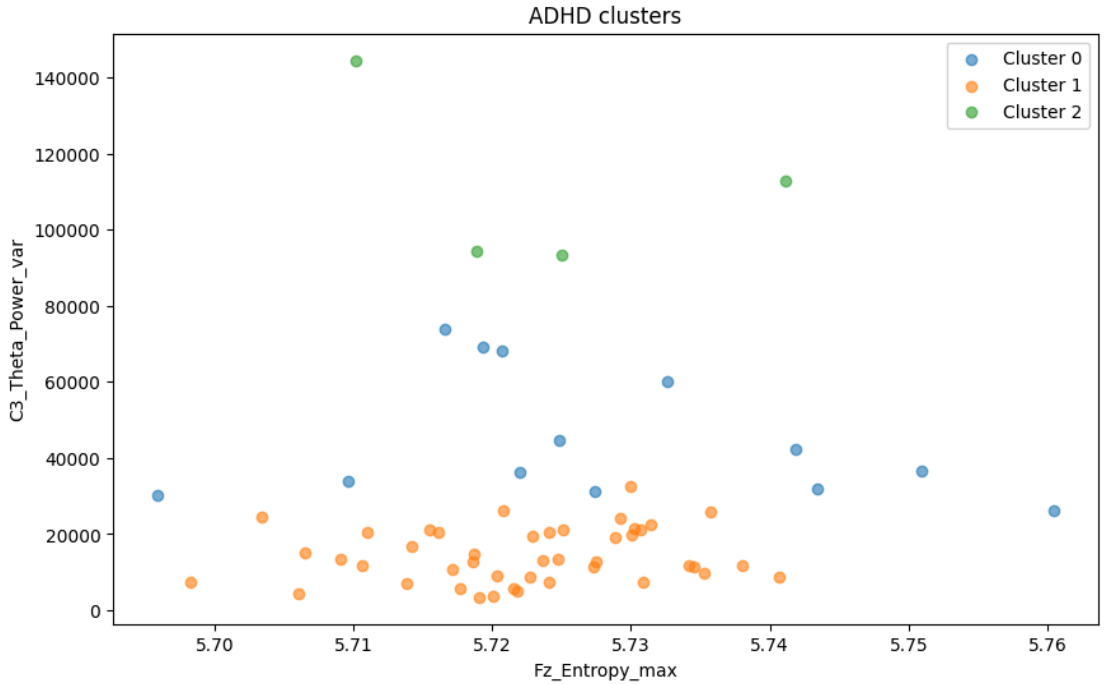




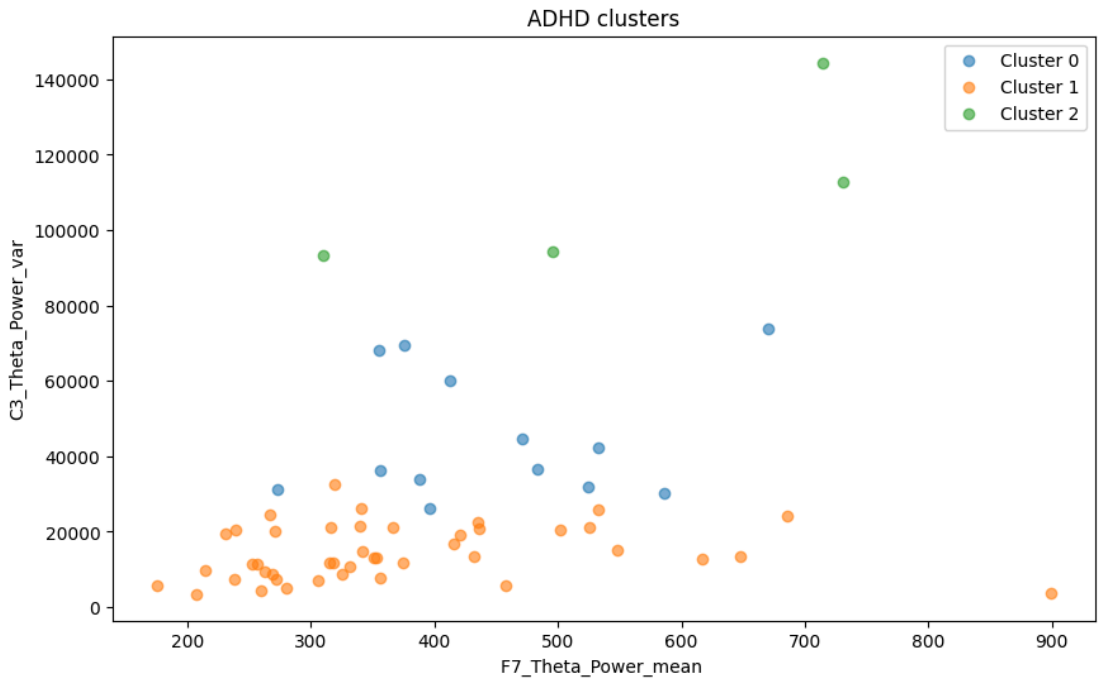
These results show a decrease in accuracy from my original approach, which can be explained by my mistake of performing a feature analysis that included the testing data, causing data leakage. Still, the results came relatively close to my previous accuracy of 88% which included features based on the testing data.

VII. Exploring ADHD Subgroups with Clustering

To get answers about potential subgroups of people with ADHD, I used K-means clustering to find clusters of people within the ADHD group. For this, I chose to split the data into three clusters because this created the most separated and logical clusters. In order to visualize the clusters, I looked at the silhouette score for each pair of features and plotted the clusters on the pair with the highest score, which were Fz_Entropy_max and C3_Theta_Power_var.



The results show a majority of the subjects in cluster 0 (yellow) with a low theta power variation on channel C3, and the other 2 have higher variation. Because groups had similar entropy values for the most part, I also plotted the clusters with F7_Theta_Power_mean on the x-axis to get more information on how theta power differs between the clusters.



I suspected the high variance on cluster 2 (green) might correlate to difficulties consistently concentrating because it could mean theta power was increasing at some points, marking a sudden state of relaxation where they shifted their attention away from the game. The second graph supported that idea because it turned out that the clusters with higher variance also had a higher mean theta power. A higher theta EEG power is generally associated with lower cognitive functioning because Theta frequencies are present during a state of relaxation. Therefore, these clusters may represent subgroups of children with ADHD that have different levels of executive functioning challenges or a harder time staying engaged with the task.

VIII. Conclusions

This project explored the potential of using EEG signals as a tool to predict ADHD diagnoses in children and identify potential subgroups within the ADHD population. By preprocessing the raw EEG data, extracting meaningful features, and experimenting with multiple machine learning algorithms, I was able to train models to classify ADHD with an accuracy of up to 84%. These results demonstrate the promise of EEG data and machine learning in aiding ADHD diagnosis. However, the accuracy is still not high enough to be used as a sole diagnostic tool. In the future, prediction accuracy could be improved by training models over a

larger dataset and by researching more features of EEG that may be markers of ADHD. Deep learning techniques could also be a good approach.

The features that I was able to extract and determine as relevant to ADHD were Theta power, TBR, amplitude envelopes, raw signal variation, fractal dimension, and entropy of the signals. The way these features change over time also proved to be an important aspect of how they relate to an ADHD diagnosis. The impact of TBR and Theta power of the results also supports a lot of research about differences in the ADHD brain.

My results from K-means clustering reveal three potential subgroups within ADHD that are highly dependent on the variation in theta power, specifically on the C3 channel. That channel is located in the cerebral cortex, which plays a key role in memory, thinking, learning, reasoning, problem-solving, emotions, consciousness, and functions related to the senses. The majority of subjects fell into a cluster with lower variance and a lower overall presence of theta frequency waves. The group with the highest clusters contained the fewest people, and this group had a very high theta power variation and mean, which most likely signifies lower cognitive functioning and trouble concentrating.