EEG Data Analysis for Psychiatric Disorders

Ohad Gonen and Ora Friedman

Introduction

The aim of this project is to identify characteristic EEG brain activity patterns associated with different psychiatric disorders. Psychiatric disorders are mental illnesses that greatly disturb thinking, moods, and/or behavior. Therefore, we hypothesize that EEG activity will show significant differences both between psychiatric disorders and compared to healthy controls. We used an EEG dataset published by Su Mi Park on Kaggle, which includes detailed EEG recordings from 850 individuals diagnosed with various psychiatric disorders, along with 95 participants for the healthy control group.

We started with data cleaning and preprocessing, which included filling missing values, removing irrelevant columns, and standardizing categorical columns. Then, we wrote functions that calculated average electrode activity in each frequency band for specific disorders, visualized the brain activity, and performed statistical analysis to find significant differences in brain activity between disorders and healthy control. We continued to analyze the data and wrote functions to evaluate differences between the specific disorders and define correlations between far electrodes that show the different networks associated with each disorder. Finally, we tried to write a machine learning code that will enable us to predict a psychiatric diagnosis based on given EEG values.

Methods

Our project required a range of libraries and tools to clean, preprocess, analyze, and visualize the EEG data, ensuring accuracy and consistency.
For data cleaning, we relied on Pandas to handle data frames, fill missing values, and standardize columns, along with NumPy for numerical operations. During preprocessing, Scikit-learn was used for encoding categorical variables, splitting the data into training and testing sets, scaling features, and training machine-learning models. Data analysis involved statistical tests using SciPy.

For visualization, we used Matplotlib for basic plotting and Seaborn for advanced statistical visualizations. EEG-specific visualizations, including topographical maps, were created using MNE, while NetworkX enabled the analysis and visualization of long-range correlations as network graphs. Pytest was used to ensure the correctness of the code through rigorous testing.

We wrote several key functions to support our workflow. Data cleaning included the functions: fill_NaNs to handle missing values, standardize_categorical_columns to normalize text data, and reformat_electrode_columns to ensure consistency in electrode naming. The model training involved preprocess_data to encode variables and balance classes, split_data to create training and testing subsets, and train_model to train a Random Forest classifier.

For analysis, calculate_band_averages computed EEG band activity, while find_significant_differences identified key differences between psychiatric disorders and healthy controls. Additionally, find_strong_long_range_correlations uncovered connectivity patterns in brain activity by analyzing relationships between distant EEG electrodes.

Visualization played a critical role in interpreting the data. Functions such as visualize_brain_activity and visualize_all_disorders provided insights into EEG activity across disorders, while visualize_long_range_correlations highlighted connectivity strengths as network graphs.

Testing was conducted at multiple levels, including null, error, positive, negative tests for individual functions like test_fill_NaNs and test_calculate_band_averages, integration tests to verify workflows, and edge case tests to ensure robustness under unexpected inputs. These measures ensured the reliability and validity of the analysis pipeline.

Results and Discussion

Our visualizations results and statistical analysis were fascinating.

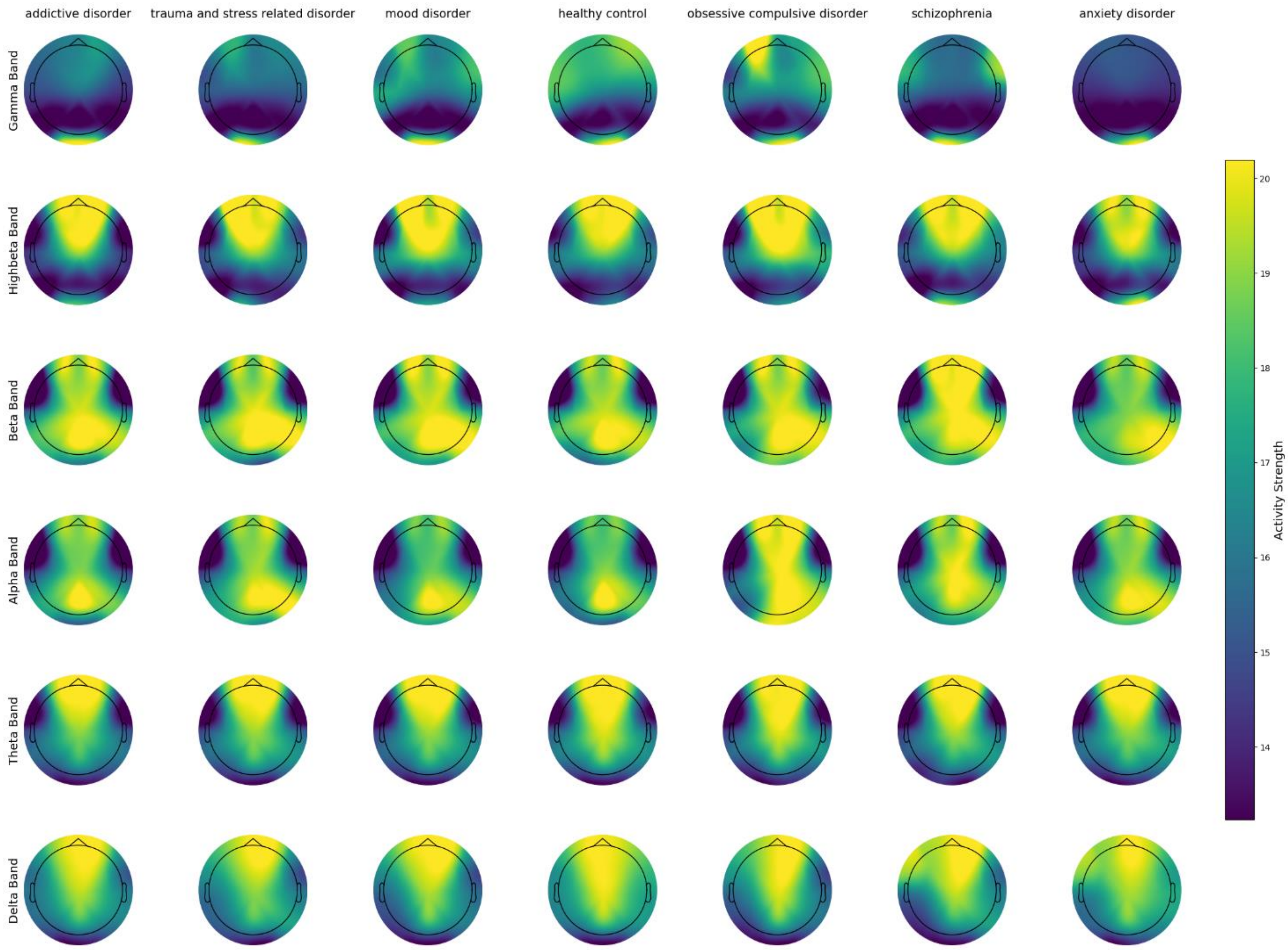

Figure 1. An overview of the topographic brain activity for all disorders according to the different frequency bands. Notice the middle column is the healthy control group.

We focused our first demonstration on schizophrenia and mood disorders, utilizing the knowledge we have gained throughout our degree to analyze and explain the results effectively.

The analysis reveals that individuals diagnosed with schizophrenia (Fig. 2), exhibit heightened delta activity in the occipital lobe, a region critical for visual processing, potentially linked to visual hallucinations and a dream-like neural state while awake. Elevated beta activity is contributing to thought disorders and executive dysfunction. In contrast to the highly significant differences observed at $P<0.001$, no notable changes were found in theta, alpha, gamma, or high-beta activity. This suggests that schizophrenia is characterized by specific neural disruptions rather than widespread neural excitation.

Individuals with mood disorders exhibit significantly increased beta and high-beta activity across multiple brain regions, which may be associated with heightened alertness, problem-solving, and excessive cognitive activity. This excessive beta activity is associated with anxiety, stress, and hyperarousal, common in conditions like major depressive and bipolar disorders (Fig. 3).

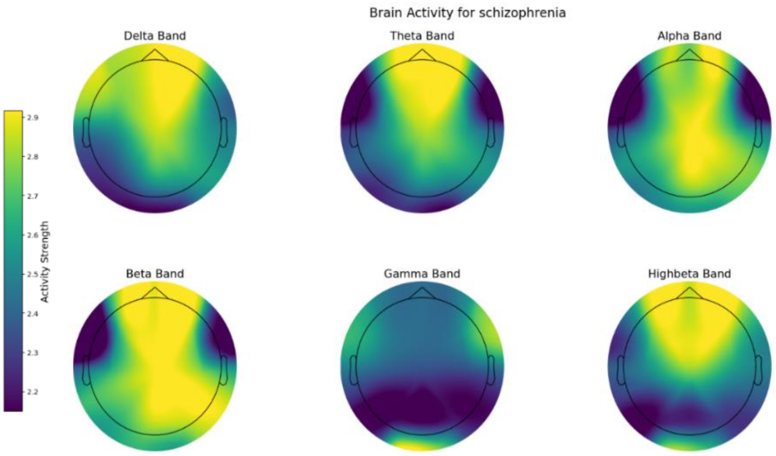

Figure 2. Visualization and significant differences between schizophrenia and healthy control (P<0.0001):
'delta': ['O1', 'O2'],
'beta': ['FP1', 'FP2', 'Fz', 'F4', 'F8', 'C4', 'O1']

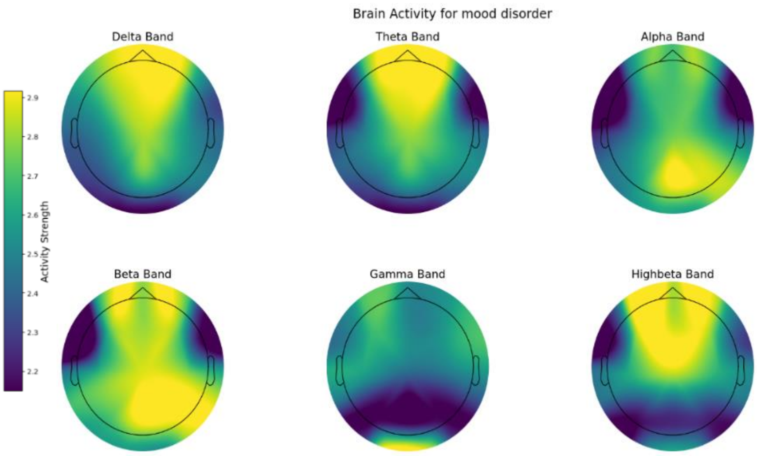

Figure 3. Visualization and significant differences between mood disorder and healthy control (P<0.01):
'delta': ['P4', 'O1', 'O2'], 'theta': ['O2'],
'beta': ['FP1', 'FP2', 'F7', 'F3', 'Fz', 'F4', 'F8', 'C3', 'Cz', 'C4', 'P3', 'Pz', 'P4', 'O1', 'O2'],
'gamma': ['P3'], 'highbeta': ['F3', 'C3', 'Cz', 'C4', 'P3', 'Pz', 'P4', 'O1'].

Statistical tests revealed no significant overall reduction in neural activity among psychiatric disorder patients compared to healthy controls. This may be explained by several factors. Impaired inhibitory control, such as GABA dysfunction, could lead to excessive neural firing and increased EEG power. The brain might also activate compensatory mechanisms, boosting activity in some regions to offset deficits elsewhere and maintain cognitive function. Additionally, psychiatric disorders may involve hyperconnectivity or dysregulated neural networks, leading to increased EEG power in specific frequency bands like theta and beta. Lastly, the effects of psychiatric medications, such as antipsychotics, SSRIs, and stimulants, could elevate EEG activity in certain bands, contributing to the observed enhancements.

Variations in Brain Activity Within Specific Disorders

The differences between specific disorders of each main disorder, are very interesting and can emphasize and explain their different symptoms.

Between the trauma and stress related disorders, PTSD shows the highest gamma activity, indicating heightened cognitive engagement or stress responses compared to other disorders. Adjustment disorder exhibits the highest alpha activity, suggesting relatively better relaxation or focus among the disorders. Acute stress disorder displays consistently high delta and theta activity, reflecting its association with stress and baseline brain processing. Highbeta activity shows minimal variation across disorders, indicating that it may not be as sensitive to differences within this category (Fig. 4).

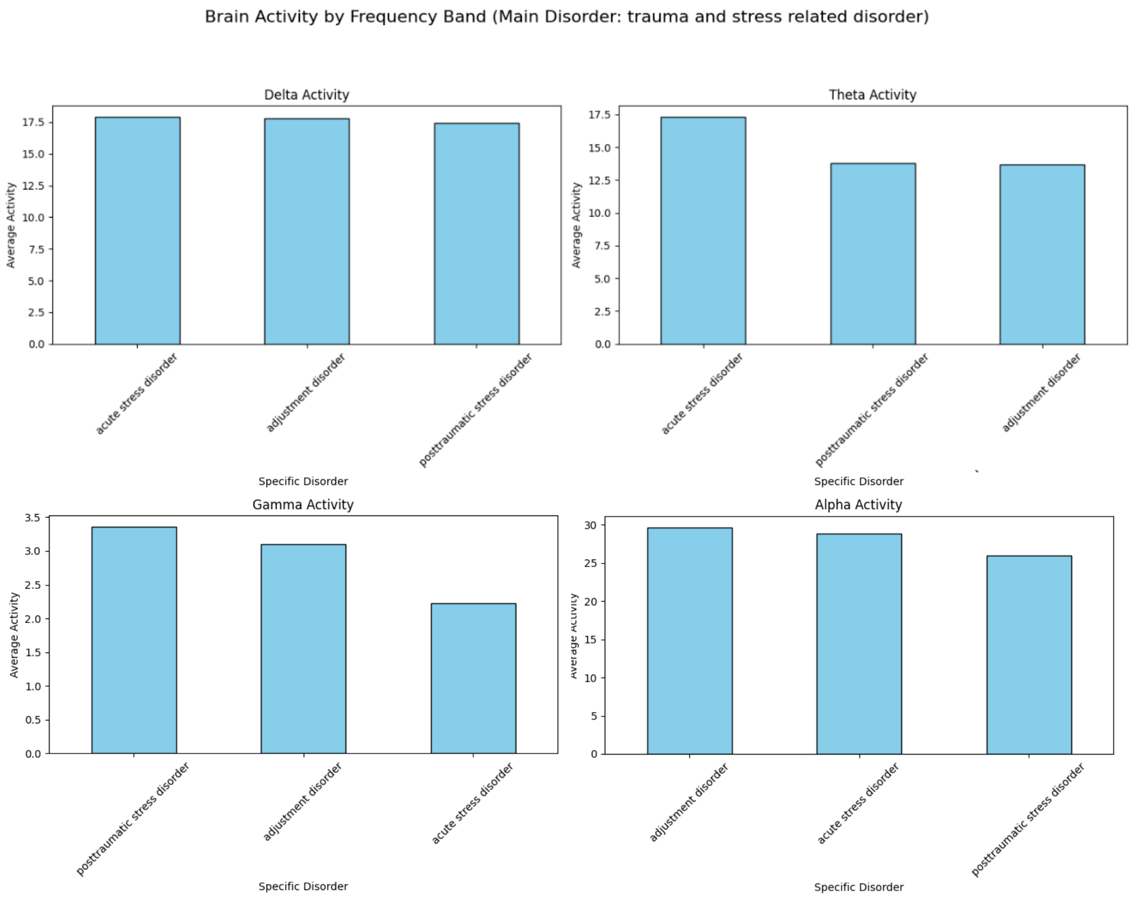

Figure 4.

EEG Activity Variations Among Trauma- and Stress-Related Disorders.

Another example of brain activity variations can be seen within the category of addictive disorders (Fig. 5). Alcohol use disorder is characterized by higher alpha activity, suggesting greater relaxation, while behavioral addiction disorder shows increased gamma activity, indicating heightened cognitive engagement or stress. In contrast, delta and theta activity remain similar in both, reflecting comparable baseline brain processing.

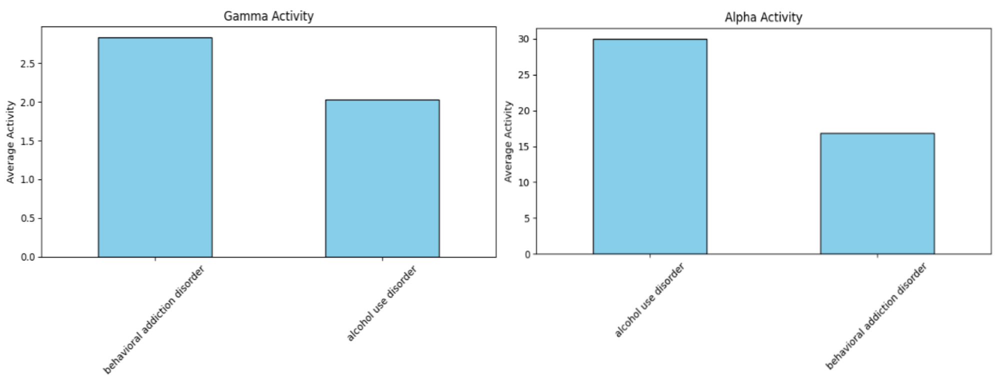

Figure 5. EEG brain activity variations among addictive disorders.

correlations between distant electrodes

Distant electrodes are defined as pairs located in anatomically and functionally separate brain regions. This approach focuses on analyzing true long-range correlations, excluding nearby or highly connected regions.

During rest, the brain's spontaneous activity forms networks like the DMN, visual, and sensorimotor networks, with distant correlations reflecting intrinsic connectivity (Fig. 6). The EEG correlation network reveals strong long-range and functional connections, with T4 as a central hub linking frontal, parietal, temporal, and occipital regions. The T4-Pz-P4-T6-O2 pathway suggests multimodal sensory integration involving auditory and visual processing. Strong frontal-temporal links as Fz-T3-T4-F3 highlight interactions between executive functions and memory, while frontal-occipital connections as FP1, FP2, F4-O2 indicate top-down sensory modulation.

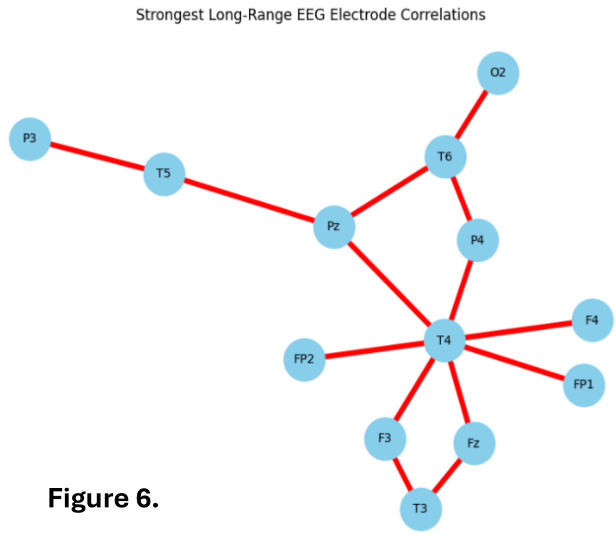

Figure 6.

In contrast (Fig. 7), OCD networks show increased local connectivity in posterior-temporal regions (Pz, T5, T6, O1, O2) and weaker frontal-central integration, suggesting hyperactive sensory processing and impaired executive control. This hyperconnectivity may underlie compulsive behaviors and rigid thinking associated with OCD.

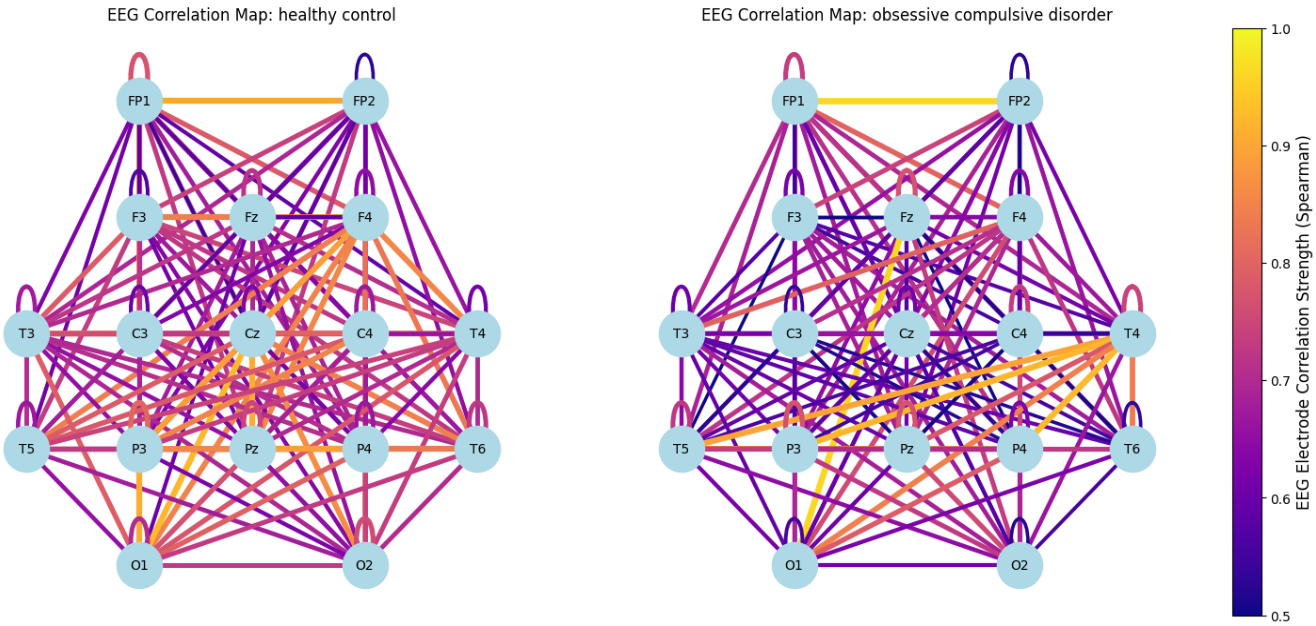

Figure 7. EEG correlation map of healthy control in contrast to OCD.

Can we predict the main psychiatric disorder based on the EEG data?

To explore whether EEG brain activity can predict psychiatric disorders, we developed a machine learning pipeline that preprocesses the data, splits it into training and testing sets, and trains a Random Forest model. The model's performance is evaluated using accuracy metrics and a classification report.

Our machine learning model achieved an accuracy of 34%, indicating challenges in distinguishing between psychiatric disorders based on EEG features. Mood disorders had high recall (77%) but low precision (33%), suggesting frequent misclassification of other conditions as mood disorders. Healthy controls had high precision (75%) but low recall (16%), meaning the model correctly identified them when predicted but missed many actual cases. Anxiety disorder showed high precision (67%) but poor recall (13%), while OCD and schizophrenia had extremely low recall (0% and 5%, respectively), making them the hardest to classify. These results suggest that certain disorders may have overlapping EEG patterns or insufficiently distinctive features for the model to learn effectively.

```
Model Accuracy: 0.34

Classification Report:
                                   precision    recall  f1-score     support
addictive disorder                  0.382353  0.382353  0.382353   34.000000
anxiety disorder                    0.666667  0.125000  0.210526   16.000000
healthy control                     0.750000  0.157895  0.260870   19.000000
mood disorder                       0.325758  0.767857  0.457447   56.000000
obsessive compulsive disorder       0.000000  0.000000  0.000000    9.000000
schizophrenia                       0.125000  0.050000  0.071429   20.000000
trauma and stress related disorder  0.428571  0.085714  0.142857   35.000000
accuracy                            0.343915  0.343915  0.343915    0.343915
macro avg                           0.382621  0.224117  0.217926  189.000000
weighted avg                        0.389731  0.343915  0.282384  189.000000
```

Conclusion

Through this project, we gained valuable insights into EEG patterns associated with psychiatric disorders and deepened our understanding of brain activity differences between conditions. We found the topic fascinating and believe EEG analysis has great potential for psychiatric research.

For further information we recommend reading the code itself.

We hope you enjoyed and found our project interesting!

Acknowledgements

Park, S. M. (2021, August 16). EEG machine learning. Retrieved from osf.io/8bsvr

https://www.kaggle.com/datasets/shashwatwork/eeg-psychiatric-disorders-dataset?select=EEG.machinelearing_data_BRMH.csv