Cognitive dysfunction underlies common mental health behavioral symptoms, and this study aimed to classify healthy vs. mild-to-moderate self-reported symptoms of each disorder using the cognitive neural markers measured. Mental health disorders are a leading cause of national and global disability, costing hundreds of billions of dollars every year. A challenge with optimizing treatment for neuropsychiatric disorders is the lack of simple, scalable measurement tools to identify objective biological markers for treatment. This study investigated whether an electroencephalography (EEG) can be used to identify healthy people who are self-reporting symptoms of common mental health disorders. Previous research has shown that alterations in cognitive brain functions are associated with common mental disorders. However, specific mental health symptoms have not been predicted from the standpoint of multiple neural markers underlying essential cognitive processes. This study used a scalable cognitive brain-mapping platform to analyze the data from 97 healthy adults performing five fundamental cognitive tasks: selective attention, response inhibition, working memory, interference processing, and emotion interference processing. All participants provided self-reported symptoms of anxiety, depression, inattention, and hyperactivity, and the analyses focused on how to optimally use the EEG data to classify the subjects based on their noted mental health symptoms. This design enables the investigation of the relationship between the brain patterns underlying the different cognitive processes within the same subject and the mental health of the subjects. Balasubramani et al. (2021) used EEG data from 97 healthy adult subjects to build a model of cognitive task-related EEG processing. The size of the dataset was Xi R68 3 5, i = 1,..., 97, or, in other words, each subject Xi had 68 ROIs 3 frequencies 5 tasks = 1020 dimensions. Each subject was binary-labeled for each of the four most common psychiatric symptoms. 102 adult human subjects (mean age 24.8 6.7 years, range 18–50 years, 57 females) participated in the BrainE neuro-cognitive assessment study, recruited using IRB-approved on-campus flyers at UC San Diego and via the online recruitment forum, ResearchMatch.org. All participants provided written informed consent for the study protocol approved by the UC San Diego Institutional Review Board (UCSD IRB). Participant selection criteria included healthy adult status, normal or corrected-to-normal vision and hearing, right-handedness, and at least a high school education level of 16 years. Five participants were excluded from the study as they had a current diagnosis for a psychiatric disorder and/or a current or recent history of psychotropic medications, so a total of 97 subjects were used for the analyses presented in this study. This study examined the experimental design of BrainE Neuro-Cognitive Assessments, developed and deployed by NEATLabs on the Unity game engine. The Lab Streaming Layer protocol was used to timestamp each stimulus and response event in each cognitive task. Participants underwent the following cognitive assessment modules within a 35-minute session: selective attention and response inhibition tasks; a working memory task with perceptually thresholded stimuli; a flanker interference processing task; and an emotion interference task. The Parametric Empirical Bayes (PEB) toolbox was used to train the machine learning model. Data preprocessing was conducted using the EEGLAB toolbox in MATLAB. EEG data were resampled at 250 Hz and filtered in the 1–45 Hz range to exclude ultraslow DC drifts and high-frequency noise produced by muscle movements and external electrical sources. Epoched data were cleaned using the autorej function in EEGLAB to remove noisy trials and excluded signals estimated to be originating from non-brain sources using the Sparse Bayesian Learning (SBL) algorithm. Block-sparse Bayesian learning (SBL) was used to map the underlying neural source activations for ERSPs using a two-step algorithm. Source space activity signals were estimated and partitioned into ROIs; artifact sources were removed to clean the EEG data; and the k-means method was used to cluster the IC scalp projections into brain, EOG, EMG, and unknown components. The SBL algorithm returns cleaned channel space EEG signals in addition to the derived cortical source signals as outputs. This study used the sparse Bayesian learning (SBL) algorithm to source localize 24 channels of electroencephalographic (EEG) signals. The SBL algorithm reduces the effective number of sources considered at any given time as a solution, reducing the ill-posed nature of the inverse mapping. It was benchmarked to produce evidence-optimized inverse source models at 0.95 AUC relative to the ground truth, verified using both data and simulations. The source signal envelopes were computed in MATLAB by a spline interpolation over local maxima separated by at least one time sample. Post-stimulus encoding occurs in the 100–300 msec range for the theta and alpha bands and the 400–600 msec spectral amplitude range for the beta band signals were chosen based on the peak global activity of the task-averaged signals in the respective frequency bands. This study used logistic regression to predict mental health symptoms using the logistic regression command in Jupyter notebook (Python). The task was to predict binary [0, 1] labels for all four symptoms for each subject. To achieve high accuracy, data augmentation, feature selection, and oversampling techniques were used to enhance the number of features and improve the number of samples. The L2 regularization term in the model is wTw=||w||22. Data augmentation is an operation to augment the original dataset to detect the hidden structure of the dataset. To do this, statistical measurements, products of two arbitrary features, and logarithms of each feature were used to compute the importance of the connectivity of two features in mental health prediction. Statistical measures were computed across three categories: 68 ROIs, 3 frequency bands (θ, α, β), and 5 cognitive tasks. The number of features increased considerably, and the number of features was enhanced to include base features (viz., 68 \* 3 frequency bands \* 5 tasks = 1020) + 105 + 1428 + 2380 = 4933. We computed the logarithm of the whole dataset, calibrated the value of each feature, and added it to the existing non-log features. We then selected a subset of the 24,339,422 augmented feature set and input these as predictors for the logistic regression model for predicting symptom scores. The chi-squared statistic is a generic measure of the co-occurrence between a feature and an objective nominal variable and is computed as follows for each feature: We used two methods to mitigate the drawbacks of a small sample size of 97 study participants: SMOTE (Synthetic Minority Oversampling Technique) and adding Gaussian noise. SMOTE creates new data points until the numbers of two labels for classification are equal. Adding Gaussian noise increased the number of samples by adding Gaussian white noise to the original dataset, resulting in a nine-times-larger set. We constructed a simple logistic regression model with augmented and selected features, rather than a complex model with the original features. We used basic statistical measures (max, min, …) and log or polynomial transforms to generate new features. A stratified cross-validation was conducted for each mental health symptom, and sensitivity and specificity were used to evaluate four binary classification problems. We determined a subject is true-positive for each symptom if their self-report score was no less than 5 or otherwise true-negative. This text discusses the concept of "hub-like" spectral activations that predict mental health symptom scores. It uses a simple arithmetic sum of chi-squared statistics to compute the "strength" of the connection of arbitrary pairs. The "strength" is the sum of the chi-square statistics of features, consisting of the product of all possible pairs of two features, with or without the logarithm transform. For each pair of vertices (ROIs) and each mental health symptom, the highly contributing feature (belonging to the frequency band and cognitive task) was computed. The most representative factor of pair (i1, i2) was computed as below by finding the max along one feature type and averaging across the other feature types. We computed current flow closeness and betweenness centralities to measure the extent to which a given ROI has the largest chi-square value, similar to a "hub" in a network, among all ROIs. We found that logistic regression performance was best when we used oversampling (OS), data augmentation (DA), and feature selection (FS) of the top 40,000 contributing features from our augmented data. We found >80% sensitivity and specificity, particularly for anxiety and inattention, when we employed data augmentation, feature selection, and oversampling techniques on the data. The 95th percentile confidence intervals based on the standard error of the proportion are 50 ± 9.95%. For more robust control, we performed 100 iterations of random shuffling of class labels and performed a stratified cross-validation procedure. The actual mean standard deviation accuracy for the DA, FS, and OS models were found to be above these random control thresholds, and they were 0.86 0.05, 0.78 0.10, and 0.73 0.09, respectively, suggesting the statistical significance of our results. We investigated the neural features that contribute to the prediction of mental health symptoms by mapping the chi-square statistic of pairs of ROIs to the graph. We found that the representative set of 40,000 selected features contains largely the products or logarithms of the products of a specific pair of features (as generated through data augmentation), whereas the chi-squared statistics, i.e., the rank of single features, tended to be much lower. Table 1 shows the top 10 features with the highest chi-square statistics for predicting anxiety. Table 2 shows the high-strength or highest-contributing vertices (ROIs) and their corresponding brain regions. Figure 5 shows which ROIs, frequency bands, and tasks contribute most significantly to symptom prediction.  (Kato et al., 2022)

### References

Kato, R., Balasubramani, P. P., Ramanathan, D., & Mishra, J. (2022, April 19). *Utility of Cognitive Neural Features for Predicting Mental Health Behaviors*. PubMed Central (PMC). https://doi.org/10.3390/s22093116