Assessing the utility of brain and gut cognitive electrophysiology for early prediction of treatment outcome in major depressive disorder

Amal Jude Ashwin, Alok Bajpai, Hari Prakash Tiwari, Nandini Priyanka B, Pragathi Priyadharsini Balasubramani

Translational Neuroscience and Technology Lab, IIT Kanpur, Kanpur 208016
***corresponding pbalasub@iitk.ac.in**

ABSTRACT

According to World Health Organization, about 5% of adults globally suffer from clinical depression, and in India it is about 4.5% of people. Oral medication is a common treatment against depression. However, more than half of those treated do not respond to the pharmacological treatment strategy in the first trial and may require switching or augmentation with other medications. There is a strong need for precise models for arriving at a personalized treatment strategy in a sooner timeline. Some earlier models using clinical information along with electroencephalogram (EEG) data showed good performance for early predicting treatment outcome in depression. However, the critical features identified by those studies including the presence of differential frontal theta power and frontal alpha asymmetry in depression patients has been challenged in the recent times due to contradictions in interpretability and robustness: when the theta and alpha frequency signals were teased apart from their aperiodic component, the resulting periodic components were not robust for prediction. On the other hand, gut abnormality in depression has been reported by many earlier studies but have not been used for predictive or prognosis purposes in depression. Our study aims are twofold: first to identify the features that can early predict treatment outcome, and interpret them for different patient subgroups, and second to understand the utility of longitudinal data collection and gut-brain interactions to predicting treatment outcome. About 161 participants (naïve patients = 99) registered for our longitudinal study spanning three visits, and our aim was to investigate whether visits 1 (baseline) and visit 2 (in 7-10 days) could predict the antidepressant treatment outcome in visit 3 (after 30 days). After attrition, electroencephalography and electrogastrography data from 89 participants were collected in visit 2 (patients = 42), and 61 in visit 3 (patients = 21). We used electrophysiological features in the brain and the gut along with clinical data to train simple predictive models, and it was able to reliably predict non-response to depression medications with specificity 78% and sensitivity 84%. The significant features explaining the treatment outcome were ranked, altogether offering a scalable, whole body cognition tool for clinicians for guiding their medication strategy.

Keywords. Electrogastrography, Electroencephalography, gut-brain coupling, longitudinal assessments, early prediction, treatment outcome, Depression

1. INTRODUCTION

According to WHO, Major Depressive Disorder (MDD), commonly referred as depression, is a mental disorder that is associated with consistent lack of pleasure/motivation to perform activities and negative moods. Other major symptoms of depression include irregular sleep and appetite, poor concentration, feeling of hopelessness, low energy and thoughts of suicide. Once clinically diagnosed with depression, the clinician recommends certain generic lifestyle changes and antidepressants in cases of mild depression and combination of therapies for severe cases of depression. The state of art method of treating depression using medications involving combinations of selective serotonin and norepinephrine reuptake inhibitors typically involve a trial-and-error process: that is, in this traditional approach, patients are prescribed medications, but it can take 4 to 6 weeks before healthcare providers can accurately assess whether the prescribed treatment is effective for the individual. Non-remission or non-response to treatments is as high as ~50-60% for depression treatments just after the first selected medication treatment. Altogether many patients may experience side effects or find that the medication does not alleviate their symptoms, leading to a need for adjustments or changes in treatment. (Bauer et al., 2007; Fochtmann & Gelenberg, 2005; Pigott et al., 2010; Quitkin et al., 2003).

We observe that there lacks a neurophysiology tool that accounts for precise neuroplastic effects to accurately predict treatment outcomes within a short timeframe of less than two weeks. Earlier attempts utilize a combination of clinical health records, behavioral data, and genomic information to achieve a commendable efficiency rate of around 72% within a 4-week timeframe (Athreya et al., 2021; Bares et al., 2015; Jaworska et al., 2019; John Rush et al., 2006; Keitner et al., 2008; Leuchter et al., 2009; Trivedi et al., 2006; Warden et al., 2007). Relatedly, there are a few studies that look into neurophysiology at any timepoint to predict depression severity (Cao et al., 2019; Hasanzadeh et al., 2019; Sadat Shahabi et al., 2021), and claim to achieve greater than 90% accuracy. However, they may not be sensitive to treatment related changes through time and prediction of treatment responses.

Some notable studies nearly a decade ago, that utilized tracking of changes in neural activation due to medicine to predict the treatment outcome suggests *slowing* of brain activity that is an increased frontal theta power, and also reduced frontal alpha asymmetry to be significantly explaining the response to treatment even for medicine resistant patients with greater than 91% accuracy, only when combined with features from depression rating scale (Bares et al., 2015). Moreover, frontal beta power were also identified to be significantly relating to depression (Knott et al., 2001; Lieber & Prichep, 1988). In contrast, studies in the past 5 years suggested opposing views, with some showing that measures like frontal alpha asymmetry is not sensitive to treatment response at all (Kołodziej et al., 2021; Lee et al., 2020; van der Vinne et al., 2019); and some other suggests it is not the increase in low frequency power but instead the aperiodic activity that can explain the response in patients (Smith et al., 2023). Altogether the contradicting views question the reliability and robustness of the earlier identified neurophysiological markers.

Interestingly of late, we find increasing evidence to suggest it is not just the brain physiology but the whole-body physiology that can represent the severity of symptoms in depression. Gut dysfunctions such as irritable bowel syndrome, gastroparesis often are confound depression symptoms (Lydiard, 2001; Masand et al., 1995; Overs et al., 2024; Zamani et al., 2019). In that light, studies point at gut-brain interaction being more integral to understanding the depression symptoms (Drossman, 2016; Goyal et al., 2021). Yet, these gut measures are not considered for investigating the markers of relevance to depression, and nor they contribute to the standard guiding tool of treatment strategy in depression.

Our study is novel in many folds: One that it considers the whole-body cognition especially the gut-brain coupling in addition to the electroencephalography markers for predicting treatment response. And next, the whole-body cognition is studied using various cognitive task probes to understand the task based neural circuit recruitment for the prediction. More specifically, we hypothesize that whole body physiology measures from the brain and gut during various cognitive processes, acquired using methods such as electroencephalography and electrogastrography, along with demographics and clinical scores, could be utilized to early predict the treatment outcome in depression. We present a model for early predicting treatment outcome for depression, as early as by two weeks, for the outcome that is by state of the art measured after 4-6 weeks, that is built and tested using hundreds of patients data collected in India. An early indicator tool that can help clinical experts for administering the treatments.

2. METHODOLOGY

The methodology section can be broadly divided into 4 sub-sections. 1) Data collection: This sub-section covers the experimental setup and tools used to collect longitudinal electrophysiological data and the self-reported questionnaires. 2) Feature extraction: Using the collected electrophysiological data, the quantitative features extracted are discussed elaborately in this sub-section. 3) Feature selection: The approaches used to select the relevant features for model development and analysis are defined in this sub-section. 4) Model development: Finally, the Machine Learning models developed for classifying clinical participants into responders/non-responders are outlined in the last sub-section of the methodology section.

2.1. Ethical clearance

Ethical clearance for the study was obtained by the Institutional Ethics Committee of Indian Institute of Technology Kanpur in February 2022. The informed consent from the participants were obtained orally and in a written format. The participants were awarded coupons for each visit and were allowed to voluntarily quit the study at any time without any deduction in compensation for the visit. The data confidentiality was maintained and a subject ID was assigned to the participants to mask their what and whereabouts.

2.2. Inclusion criteria

Healthy young adults from the university campus and community settings such as schools and colleges were included as a control group. Naïve depression patients were recruited with the help of the institute psychiatrist at Indian Institute of Technology, Kanpur who sees patients via. his psychiatry clinic. Participants were included from all Handedness and Genders. Depressed participants received unimodal medication treatment during the time of study.

2.3. Exclusion criteria

Patients who are in their Pregnancy or Postpartum period, with any neurological or medical comorbidity that could influence research investigations were excluded from the study. Additionally for depressed patients, participants with suicidal intent or any psychiatric emergency, bipolar depression, psychotic symptoms and substance use disorder were excluded from the study. Participants with multimodal treatment strategy were excluded from the study.

2.4. Participants consort

Out of 161 people who participated in the study, electrophysiological data of brain (electroencephalogram) and gut (electrogastrogram) were collected from 138 participants including both control and patient population at the baseline visit. However, only 62 participants turned up for all three visits and EEG and EGG were collected from them, where visit 1 (baseline visit) was at the day 0 of participant visit, visit 2 (intermediate follow-up visit) between 7-14 days from then, and the next visit 3 (final follow-up visit) after 30-40 days (refer to **Figure 1A**). Control population corresponds to the population that is not seeking any clinical help and are not diagnosed with any mental disorders. Patient population corresponds to the participants who are clinically diagnosed to have depression, naïve in medicine intake, and are intaking clinical assistance specifically through oral medication (refer to **Figure 1C**). The medication was chosen according to the judgement of the attending physician. All the participants had normal eye-sight or corrected to vision eye-sight. No other treatments were prescribed to the patient population or control population. A small token of gratitude was provided to the control group for their time.

2.5. Cognitive tasks setup

The participants performed simple tasks that have been reported to evoke various cognitive processes such as interoception, attention, memory, arithmetic calculation and language. Participants were asked to perform certain tasks on all 3 visits, while the other tasks were performed only on the second visit. Resting state eyes open and eyes closed tasks were performed for approximately 5 minutes each while resting on the chair in a quite setup. An interoceptive breathing task was then performed where the participants were asked to focus on their breathing, hyperventilate and deep breath by holding it for about 5 secs for approximately ten times spanning a total of 2 to 3 minutes. Varying frequency photic task was administered, where the photic stimulus that is present at the right top location of the participant flickers at the rate of 3,5,7,9,11,13,15,17,19,21 Hz for 10 seconds each frequency, inter-frequency-interval of 1 sec, and a total of 110 secs. The above mentioned 4 tasks were performed on all the 3 visits. During visit 2 and 3, in addition to the 4 tasks, the participants performed picture description task, Counting backwards in multiples of 7 from 100 to 40 in an arithmetic task. Mini mental state examination was performed by the participant and the results of the same were stored to assess different cognitive abilities of the participants.

2.6. Data acquisition

We administered self-report questionnaires such as PHQ9 (Kroenke et al., 2001), GAD7 (Spitzer et al., 2006), MMSE (Arevalo-Rodriguez et al., 2015), expert administered HDRS (Hamilton, 1960), during visits 1 and 3, wellbeing (Tennant et al., 2007), along with collection of electrophysiological measures. A daily activity log was administered between visits 1 and 2, whose data was not used in the current study. Electrophysiology signals were obtained while the participant performed all the simple cognitive tasks resting comfortably in a chair. Electroencephalogram (EEG) was recorded using the 24-channel montage with 22 recording electrodes, 1 reference, 1 ground electrode following 10-20 system from manufactured by Clarity medicals in the name of BrainTech machine. The sampling rate of the data collected was 256 Hz. The Electrogastrogram (EGG) was recorded using an OpenBCI setup with 1 ground, 1 reference, 2 recording electrodes, placed as mentioned in the figure (refer to **Figure 1B**). The data was acquired at a sampling rate of 250Hz. The same setup was used for all three visits. For about 23 subjects who were recruited during the final stages of the study, a 7 channel EGG setup was used, while for the remaining subjects, a 5-channel setup was used.

2.7. Electrophysiology data preprocessing

2.7.1 EEG

EEG data cleaning was performed using EEGLAB v2022.1, MATLAB R2022b.

Bandpass filtering: The EEG data was bandpass filtered using a Hamming windowed FIR filter (pop_eegfiltnew function in EEGLAB) with 'locutoff', 'highcutoff' and 'filtorder' parameters set to 0.5,35 and 3300 respectively. Slow frequency drifts and high frequency channel noise are removed during bandpass filtering.

Labeling bad channels: Bad channels were labeled post filtering based on whether the standard deviation of any channel data exceeds the 75th percentile of standard deviation of the rest of the channel and is greater than 100 or less than 1 microVolts.

Source localization-based artifact removal: Independent Component Analysis (ICA) was executed for the data without the bad channels using pop_runica function with 'icatype' parameter set as runica. For *n* number of channels, ICA returns at most *n* number of components. The labels for the components were extracted using iclabel function in EEGLAB which classifies the components into **'brain', 'eye', 'muscle', 'heart', 'line noise', 'channel noise', 'others'** and returns the predictive probabilities of the components for each class. Bad components, the ones which have sum of predictive probability of the brain and others to be less than 0.1, were removed from the channel signals using pop_subcomp function in EEGLAB.

Interpolation of bad channels: The labeled bad channels were interpolated using pop_interp function with the 'method' parameter set to spherical.

Common average referencing: Finally, common average re-referencing was performed using pop_reref function. The common noise that is recorded by all the channels is reduced due to the common average referencing method.

Z-score normalization: The resulting EEG signals for a task and a subject was Z-score normalised across all channels.

2.7.2. EGG

Bandpass filtering: To obtain normogastric signal, the EGG signal was bandpass filtered between 0.03 (1.8cpm) and 0.07 (4.2cpm) using the fir2 function in MATLAB with a transition width of 0.01 and filter order of 3. For tachygastric signals, the lower and upper frequencies were set to 0.07 (4.2cpm) and 0.15 (9cpm) respectively (Wolpert et al., 2020). The power spectral densities of the filtered signals are used to validate the process of filtering (refer to **Figure 1D**).

2.8. Quantitative electrophysiology feature extraction

The data was fragmented into 1minute fragments and quantitative features were extracted using preprocessed EEG and EGG. The EEG features were grouped region-wise into 6 regions, frontal right (Fp2, F4), frontal left (Fp1, F3), central right (C4, P4), central left (C3, P3), occipital right (O2, T6) and occipital left (O1, T5).

2.8.1. Absolute band power of brain signals

The power spectral density was computed by considering the Z-score normalized channel signal and performing fast fourier transform with window size as 5 seconds and 50% overlap. The average band power (theta (4-7 Hz), alpha (8-12 Hz) and beta (13-30 Hz)) were calculated within the range as the absolute band power.

2.8.2. Relative band power of brain signals

The relative band power was computed by dividing the absolute band power and total power between 1 to 35Hz.

2.8.3. Separating the aperiodic and Periodic band power of brain signals

The Power spectral density (PSD) was obtained using pwelch function in MATLAB considering 5 second windows and 50% overlap between windows and were normalized using eqn. (1). Brain signals that are recorded using EEG has a 1/f component i.e., there is more power at lower frequency and less power at higher frequency. Therefore, fitting oscillations one over frequency (FOOOF) method was used to separate the aperiodic and the periodic components in the power spectral density (Donoghue et al., 2020). FOOOF fits 2 aperiodic parameters, exponent and offset, to the log transformed power spectral density of the broadband EEG signal and generates an aperiodic component. Removing the aperiodic component, the periodic component is obtained that is comparable across frequencies. Band powers were computed using only the periodic component. The aperiodic parameters, offset and exponent, were also stored after fitting oscillations one-over frequency.

$$Power_{f_{norm}} = \frac{Power_f}{\sum_{i=0.5}^{35} Power_i} \quad (1)$$

2.8.4. Approximate entropy

Approximate entropy is a measure of randomness in the signal. It was calculated using the approximateEntropy function in MATLAB and the lag was set to be 1 second. A higher value of approximate entropy means higher randomness and a lower value means higher predictability and less randomness.

2.8.5. Theta cordance

Historically, theta cordance has been used as a measure of energy consumption in a given region. It is calculated by taking the average of absolute theta power and relative theta power as computed in sections 2.8.1 and 2.8.2 respectively.

2.8.6. Band power asymmetry

Band power asymmetry between right and left region is the difference of their relative band powers. A negative value indicates that the band power is greater in left region and a positive value indicates that the band power is higher in the right part of the region.

2.8.7. Magnitude squared coherence - a measure of functional connectivity

Magnitude squared coherence computes the cross power spectral density between 2 signals and provides insights about the similarity between the PSD of 2 signals. If a particular frequency

is present in signal A and signal B, the coherence at that particular frequency is closer to 1 and similarly if it is absent in both, it is closer to 1. In any other case, the coherence value is low and tends towards 0. In our study, coherence was computed using the mscohere function in MATLAB and using the clean broadband EEG signals of the electrodes. The coherence at specific bands were computed by taking the average coherence at the frequencies of interest.

$$Coh_{AB}(f) = \frac{|P_{AB}(f)|^2}{P_{AA}(f) * P_{BB}(f)} \quad (2)$$

where the numerator term is the cross power spectral density between signal A and B at frequency 'f' and the denominator is the normalizing term used to limit the coherence value from 0 to 1.

2.8.8. Weighted - phase lag index

Weighted PLI between 2 signals is a measure of presence of consistence phase difference between after accounting for the effect of noise and volume conduction to an extent (Vinck et al., 2011). If there is a consistent phase difference between 2 signals, commonly referred to as phase synchronization, then the value of wPLI tends to 1, otherwise, it will tend towards 0.

$$wPLI = \frac{|E(Im(e^{i\Delta\theta}).|Im(e^{i\Delta\theta})|)|}{E(|Im(e^{i\Delta\theta})|)} \quad (3)$$

where $\Delta\theta$ is the instantaneous phase difference computed by using the hilbert transform of the signals and $Im(e^{i\Delta\theta})$ is the imaginary part of the complex phase difference.

2.8.9. Band power of gut signals

The peak band power for the filtered gut signal was computed using the pwelch function in MATLAB with 1 minute window and 50% overlap. The peak frequency corresponds to the frequency at which the peak power is observed.

2.8.10. Phase amplitude coupling between EEG and EGG - a measure of gut-brain coupling

The collected EEG and EGG signals were phase locked at the level of seconds and the analytical signal was constructed using the instantaneous amplitude of the higher frequency EEG signal and the instantaneous phase of the lower frequency EGG signal. The instantaneous amplitude was computed using the absolute of the hilbert transform of the EEG signal and the instantaneous phase was computed using the angle of the hilbert transform of the Z-score normalized, filtered EGG signal. Phase amplitude coupling (PAC) for the analytical signal was computed using the following formula,

$$PAC_{EEG,EGG} = \frac{\sum_{t=1}^{T} |A_t^{EEG} * e^{i\phi_t^{EGG}}|}{\sqrt{T} * \sqrt{\sum_{t=1}^{T} (A_t^{EEG})^2}} \quad (4)$$

The PAC measure was computed using the timepoints belonging to the top 5 percentile of instantaneous amplitude to represent gut-brain coupling.

2.9. Feature scrutinization for Prediction model - criteria based

The extracted features were tested for inter-feature correlation. Highly correlated features (>0.8) were merged by taking their average and the final set of features were passed for criteria-based scrutinization as mentioned in the following section.

Criteria (i): Does the feature show differences between healthy and depressed individuals? The extracted electrophysiology features from the baseline visit recordings were tested for statistical difference between healthy and depressed populations based on mental health scores (PHQ9, healthy: baseline PHQ9 <=5, depressed: baseline PHQ9 > 5). If the p-value was less than 0.05, then the feature passes the criteria.

Criteria (iia): Do the change in feature value over time reflect mental health severity? The change of the electrophysiological feature values at the intermediate follow-up visit from baseline (visit 2 - visit 1) were correlated with the change in PHQ9 scores at final follow-up visit from baseline (visit 1 - visit 3). If the correlation coefficient was significant (p-value<0.05), then the feature qualifies the criteria.

Criteria (iib): In the same lines as before, the change in any feature value at the final follow-up visit from baseline (visit 3 - visit 1) were correlated with the change in PHQ9 score at final follow-up visit from baseline (visit 1 - visit 3). If the correlation coefficient was significant (p-value<0.05), then the feature is said to qualify the criteria.

The features that satisfy one of the above criteria were filtered for the predictive modeling. For baseline features, if criteria (i) is satisfied, the baseline electrophysiological feature was used for model development. On feature plasticity, if criteria (iia) or criteria (iib) is satisfied, then the change in feature value at intermediate follow-up visit from baseline visit (visit 2 - visit 1) is considered.

The final set of features used for predictive model development are presented in **Supplementary table A**.

2.10.1. Feature generation for the prediction models

The criteria-based scrutinized quantitative electrophysiology features, along with the demographics of the participants and the individual answers to the questionnaires such as PHQ9, GAD7 and MMSE recorded at visit 1 and visit 2 were used as input to develop the predictive models. For the electrophysiology data fragmented into 1-minute fragments, corresponding demographics and scores of each subject were repeated for all the fragments. The fragments of various tasks were concatenated such that the 1st minute of a particular task was concatenated with the 1st minute of another task.

Feature Imputation: For feature categories such as baseline EEG, baseline EGG, change in EEG and change in EGG with missing data for less than 2 tasks and shorter tasks, the missing values were imputed using other features of the same category using Multiple Imputation using Chained Equations algorithm. IterativeImputer function from sci-kit library in PYTHON software, where the values were iteratively imputed in a round-robin fashion for a maximum of 10 iterations until the values reach an asymptotic stability. Features were generated for 5 fragments in any task, the average inter-fragment correlation across features and groups was 0.77±0.19.

2.10.2. Stepwise testing of feature categories for predicting treatment response

In order to illustrate the significant incremental variance explained by the longitudinal and gut-brain coupling features, in addition to the classical baseline EEG, we sequentially added selected longitudinal EEG and gut-brain coupling (PAC) feature groups and predicted the change in PHQ9 score at final follow-up visit from baseline (visit3-visit1) using Ridge Regression **(Figure 2B).**

2.11. Model outcome operationalization

Five models were developed for comparing the performance and explainability of treatment outcome prediction. The model outcomes were Patients with significant change (SC) or no change (NC) or healthy controls: That is, if there was >=30% reduction in final follow-up PHQ-9 score as compared to the baseline visit or if the final follow-up (visit3) PHQ9 was <=5, then there is positive change in mental health and the subject was considered as a "Significant Change" (SC). If not, then the output is treated as no change in mental health and the subject was considered as a "No Change" (NC). Healthy controls are non-clinical participants with baseline PHQ9 <=5.

Models 1 and 2 are Multi-Layer Perceptron that uses all the scrutinized electrophysiological features along with demographics and questionnaires, and outputs whether the subject is a no change in mental health/significant change in mental health (Model 1) and NC/SC/HC (Model 2) based on baseline PHQ9 score and change in PHQ9 scores at the baseline visit and final follow-up visit. The MLP was built using 2 hidden nodes with tanh activation function with constant learning rate of alpha=0.001. The parameters were finalized using Grid search algorithm.

Site of Recruitment	**PHQ9<=5 (visit 1)**	**PHQ9>5 (visit 3), >=30% reduction (visit 1-3)**	**PHQ9>5 (visit 3), <30% reduction (visit 1-3)**
Non-clinic (control)	Psychotypical healthy controls (HC)	Non-clinical responders	Symptomatic controls (NC)
Clinic (patient)	Sub-threshold Depression	Significant change in Mental Health (SC)	No change in Mental Health (NC)

Table 1: Terminologies used to define the participants based on their symptomological profile. The groups are categorized as HC/SC/NC and are used for model development and analysis. The rows indicate distinct sites of participant recruitment, while the first column indicates the status of initial visit 1 PHQ9, and the second and third columns indicate the final visit 3 PHQ9 scores.

Model 3 is a logistic regression model that outputs whether a depressed subject is a SC/NC. For the development of this model, only the data points of subjects whose PHQ9 score at baseline is >5 was used. Therefore, all the healthy controls were excluded and only 2 classes were used in this model.

Models 4 and 5 are Recurrent Neural Networks that learns the time dynamics signature differentiating various subject groups, with input as the 1 minute fragments through time and had a design of 2 hidden nodes to the output layer with either 2 nodes representing NC, SC or 3 nodes representing NC, SC, HC, respectively.

Oversampling and Synthetic Minority Oversampling Technique (SMOTE) was used by all the models to address the issue of class imbalance and increase the datapoints in the training dataset

respectively. The imblearn library in PYTHON was used with the default parameters of sampling_strategy as "auto" and n_neighbours as 5. Only the minority classes were oversampled to match the number of samples in the majority class.

2.12. Model closeness measure for prediction

Sample similarity measures were computed, and only the closely matched sample related model learning weights were used for predicting outcomes of any sample during testing. We designed a nested k-fold setup for model training and testing, each instance of the model is trained with a subset of dataset (following k-fold partition, and stratified in sample distribution) for performing prediction. For a test sample, we identify those instances of models with high closeness of the test sample to the training data and prioritize the prediction of those selected weights. More precisely, the euclidean distance between the test subject and the centroids of each class data is computed. In order to ensure that the test subject is closer to only one class, the least distance value was used as a closeness measure,

$$closeness = d_{minimum} \quad (5)$$

The C% of close models were used to predict the output of the test subject **(Figure 4B).** The total number of instances run in the nested loop was 150. The performance of the model was evaluated using accuracy of predictions in a nested k-fold cross validation setting for reliability. In the nested k-fold, the dataset was separated into n folds (n=5) where 1 fold was held out and other n-1 folds were used for training the model. The training dataset was further divided into m folds (m=5) and different instances of the model were trained and used for prediction depending on the closeness measure.

2.13. Feature importance

SHAPley value is a measure of marginal contribution of the feature towards the prediction of a class for a given datapoint. Depending on the feature value, the SHAPley value increases or decreases, and this relationship is captured by the slope of the linear model fit between the feature value of the all the datapoints and their corresponding SHAPley values. The feature importance score is calculated using the following formula,

$$\text{Feature Importance } Score = |Slope| * |SHAP_{SC} - SHAP_{NC}| \quad (6)$$

In the above formula, the score is maximum when the magnitude of slope is maximum and marginal contribution of the feature increases the prediction of one class (SC or NC) while bringing down the prediction of the other.

2.14. Baseline psychotyping

The distribution of baseline questionnaires (PHQ9, GAD7, MMSE) was presented in a reduced dimensional form using Kmeans clustering algorithm with number of clusters set as 3 because of the presence of 3 groups (healthy, responders and non-responders).

2.15. Statistics

For all correlation analysis, the test for normality was performed using Anderson-Darling test and if the data was normal, the correlation was calculated using Pearson correlation method.

If the feature failed to pass Anderson-Darling test, then Spearman rank-based correlation method was used. As demographics have categorical variables, chi-square test was performed to compute statistical significance.

For feature comparisons based on existing theories, based on normality, t-test or rank-sum test was performed to calculate the p-value. As the validation of existing theory was not a part of our hypothesis testing, no corrections were performed to the p-values.

3. RESULTS

Our primary study objective is to assess whether any behavioral or whole person neurophysiology markers can predict the treatment outcome as soon as 7-10 days after intake of medicine in depression patients? To answer, we setup our experiment in a longitudinal design and recruited treatment naïve patients, control subjects; the patients start their medicine intake from the index date on visit 1 to the clinic, follows up in 7-10 days time for visit 2 experimental procedures, and again after 30-40 days for follow up visit 3. These patients were prescribed to take Selective Serotonin Reuptake Inhibitors (SSRI), Benzodiazepines or atypical antidepressants. The age range-matched controls were recruited from the community outside of the clinic and were assessed in our research laboratory setup. **Figure 1** presents the schematic of the experimental timeline, the setup and the consort diagram. We collected initial patient information, demographics, medicine information, history of trauma, during the index visit 1 date. During visits 1 and 3, we collected electrophysiological data including electroencephalogram (EEG), electrogastrogram (EGG) for various cognitive tasks, along with administration of PHQ9, GAD7, MMSE, and conducted HDRS clinical interview. During visit 2, we collected their EEG, EGG for cognitive tasks, and collected wellbeing scores. The cognitive paradigms include being in simple eyes open resting state, eyes closed resting, performing hyperventilating interoception, and photic stimulation, that were administered to all participants in all 3 visits.

The control population had a mean age of 34.3 yrs (±12.17) with 51 male participants and 10 female participants. The patient population has a mean of 35.4 yrs (±15) with 73 male participants and 26 female participants. The signal characteristics **(Figure 1D)** shows that broadly across visits, there are no statistically different spectral EEG presentation between control and depressed population when averaged across regions, however, theta power in central regions were significantly greater for the depressed population when compared to the control population (p=0.014 (left), 0.045 (right). The control population had a higher normogastric EGG power **(**p=0.048, **Table 1).**

Figure 2A presents the initial patient characteristics from various clinical data collected from visit 1. Our initial analysis showed there were significant differences between control and depressed groups in visit 1 for medicine and trauma history, PHQ9, GAD7, MMSE **(Table 1)**. The clincal participants who showed significant change in mental health (N=28) had a slightly greater mean PHQ9 and GAD7 scores (question-wise) when compared to those who showed no significant change in baseline visit 1 (N=22, **refer to Supplementary figure 1).** Patients who presented significant change in mental health response showed improvement across all questions in PHQ9 and GAD7. **Figure 2B** presents the results of the psychotyping and we observe that the blue cluster predominantly contained healthy subjects (PHQ9<5, 73%) and the green cluster represent patients showing significant changes in response (67%) and no change in mental health (33%). The orange cluster had almost equal proportion of responders and non-responders.

Feature	Control	Depressed	p-value
Age	34.3 yrs (±12.17)	35.4 yrs (±15)	0.496
Handedness*	50 right-handed, 10 left-handed	94 right-handed, 5 left-handed	0.049
Gender	51 male, 10 female	73 male, 26 female	0.300
History of Trauma*	57 No	84 No, 10 Yes	0.005
History of medications*	1 Yes, 59 No	38 Yes, 61 No	<0.001
PHQ9*	4.85 (±4.26)	12.48 (±5.2)	<0.001
GAD7*	4.06 (±3.67)	10.98 (±4.46)	<0.001
MMSE*	26.75 (±3.36)	25.24 (±2.62)	0.001
EGG Normogastric power*	0.0061 (±0.0011)	0.0057 (±0.0012)	0.048 (t-test)
EGG Tachygastric power	0.0031 (±0.0003)	0.0031 (±0.0002)	0.881 (t-test)

Table 2: Summary of demographics and baseline electrophysiology in our population. The statistics indicate the p-value of chi-square tests for Handedness, History of Trauma, History of medications, PhQ9, GAD7, MMSE, and t-test of the normogastric EGG power are significantly different between non-clinical (control) and clinical (patient)groups.

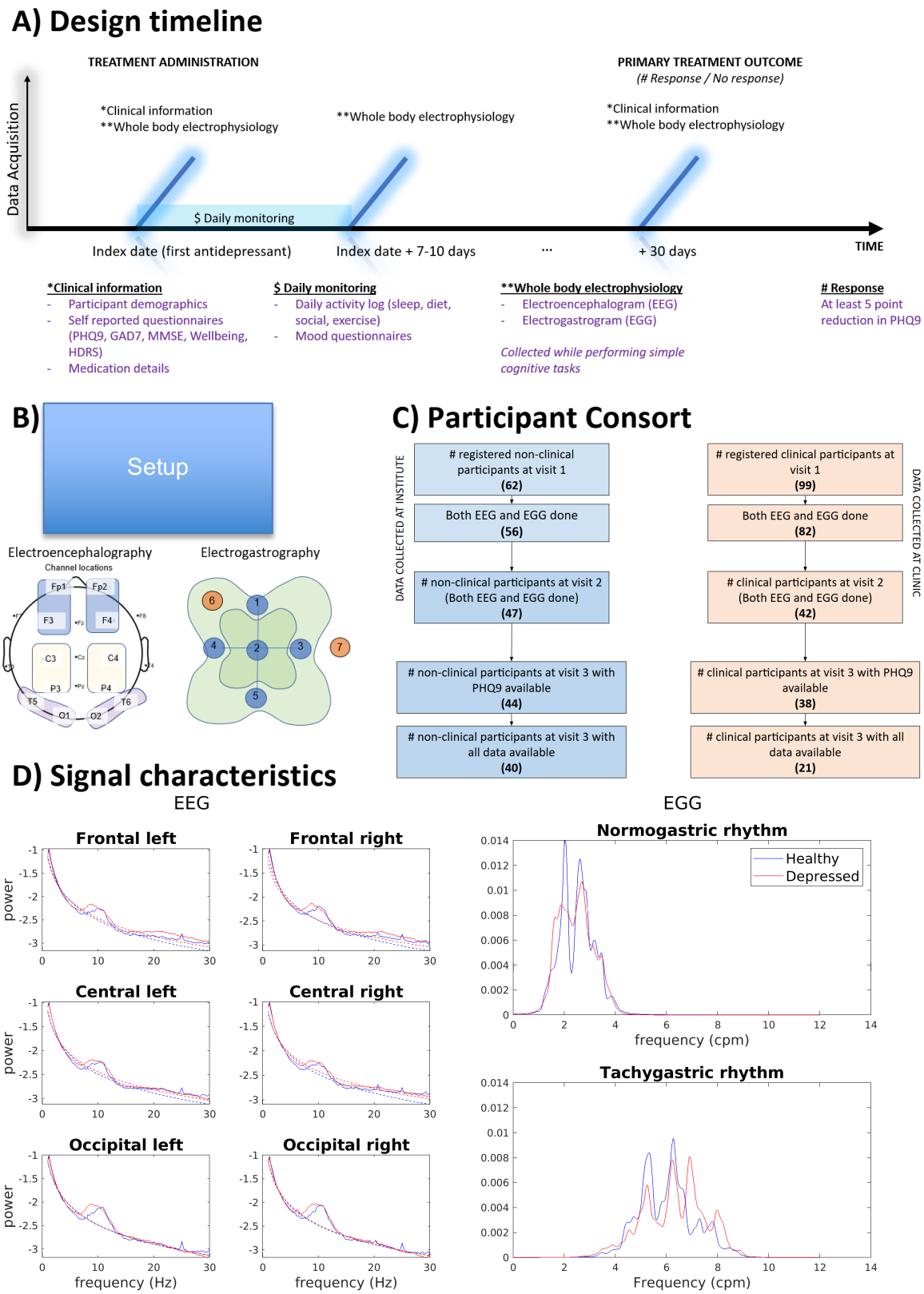


Figure 1: Study design and baseline electrophysiology characteristics. A) Study timeline: In visit 1, clinical information such as demographics, questionnaires and baseline whole body electrophysiology

(electroencephalography-EEG, electrogastrography-EGG) data are collected. During visit 2, only whole-body electrophysiology while performing simple cognitive tasks is collected. In the final visit, self-reported questionnaires and whole-body physiology are collected. **B)** Experimental setup: The grouping of EEG electrodes in 6 regions and the location of electrodes for electrogastrogram set-up are depicted. **C)** Participant consort: The number of participants for each visit with electrophysiology and PHQ9 questionnaire data are highlighted. **D)** Averaged power spectral density curves for clean EEG and EGG across tasks: Log transformed PSD and aperiodic fit (dotted lines) for the filtered band range of healthy and depressed population are plotted for comparison, central theta EEG were significantly greater in the depressed population. In EGG, we present the normogastric (0.03-0.07Hz) and tachygastric (0.07-0.15Hz) power for the two groups, and the normogastric signals were higher in control.

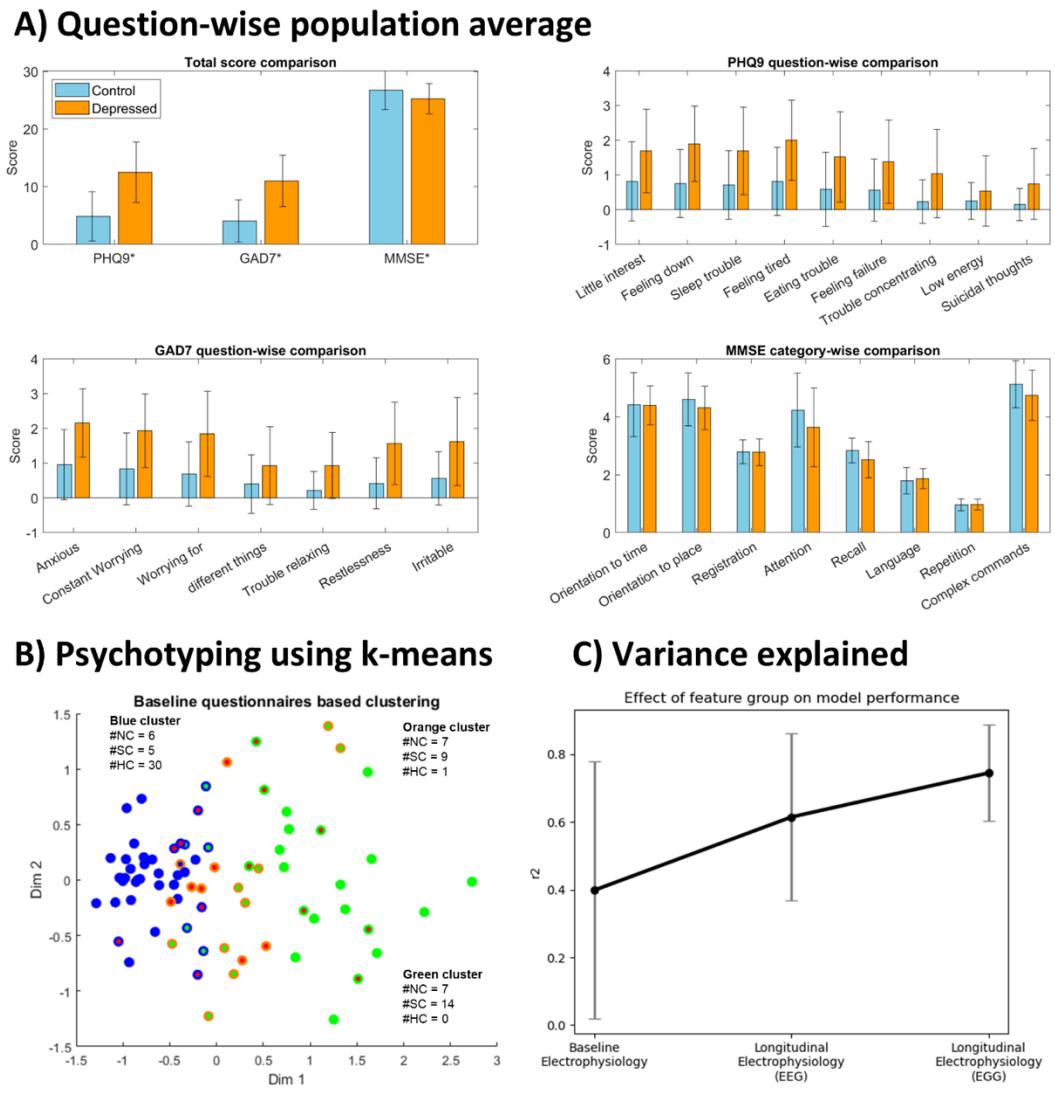


Figure 2: Participants clinical characteristics. A) Question-wise comparison of mean scores for the control and depressed populations. **B)** Baseline questionnaire clustering using k-means algorithm: Psychotyping of information obtained in visit 1 into 3 groups based on their eventual response characteristics in visit 3 (Healthy controls HC, Patients showing significant change in mental health SC, no change in mental health NC), performed using baseline PHQ9, GAD7 and MMSE responses. **C)** Importance of longitudinal design and brain-gut coupling for characterizing treatment strategy of patients: Ridge regression model performance after including various feature groups in steps.

Initial analysis provides evidence for a longitudinal design and including of brain-gut coupling to predict treatment outcome.

How do different forms of data, such as clinical reports, EEG, EGG, collected at different time points of treatment broadly contribute to response prediction? We start investigating this question using a ridge regression model in this section and later in the manuscript through various machine learning models. We setup the simple regression with just the baseline EEG features collected during visit 1. The cross-validated r2 score of the model was 0.4, indicating it explained about 40% variance in treatment outcome data. Interestingly, further adding the longitudinal brain information improved the explanation of variance to approximately 60%. And having information about longitudinal changes of brain and gut through time till visit 2 explained >70% of data **(Figure 2C)**.

Major theories on frontal activations and brain-gut interactions were sensitive to treatment response

Earlier studies hint at least 4 main theories for predicting treatment response in depression, such as increased theta cordance and increased magnitude of frontal alpha asymmetry in depression. There has been a decrease of excitation inhibition ratio suggesting a plausible role of aperiodic exponent, and an increased gut symptom presentation suggesting a plausible role of gut-brain coupling in patients. We asked whether these four features: gut-brain coupling, aperiodic exponent, theta cordance and frontal alpha asymmetry (right – left) can be a reliable marker for our study as well?

At baseline, we did not observe significant differences in theta cordance between control and depressed individuals. However, the feature was predictive of early response fairly in visit 2 during eyes closed cognitive state ($d = 0.389$, $p = 0.007$) and were reliable even in visit 3 ($d = 0.360$, $p = 0.043$), where the non-responders showed greater reduction in frontal theta cordance when compared to responders.

We also observed increased alpha activations in the right compared to left during baseline in controls than in depressed individuals in resting state eyes open task ($d = 0.280$, $p = 0.003$). Interestingly, in eyes closed task ($d = 0.344$, $p = 0.011$) the feature was lower in controls, indicating a balancing right to left activations **(**refer **Figure 3)** and they were early predictive of treatment response especially during interoceptive cognitive state ($d = 0.593$, $p = 0.004$).

The grand average aperiodic exponent across frontal brain regions was significantly more for depressed individuals when compared to healthy participants in breathing task ($d = 0.381$, $p = 0.042$). Over time, we noticed the feature decreased more in non-responders especially during eyes closed state ($d = 0.328$, $p = 0.048$) in visit 2, but shows the opposite trend in eyes opened state ($d = 0.391$, $p = 0.027$) in visit 3.

Although there is no significant difference for gut-brain coupling across tasks at baseline, medication increases the coupling value for responders for eyes open task in visit 2 ($d= 0.287$, $p = 0.029$) and breathing task in visit 3 ($d = 0.749$, $p = 0.033$).

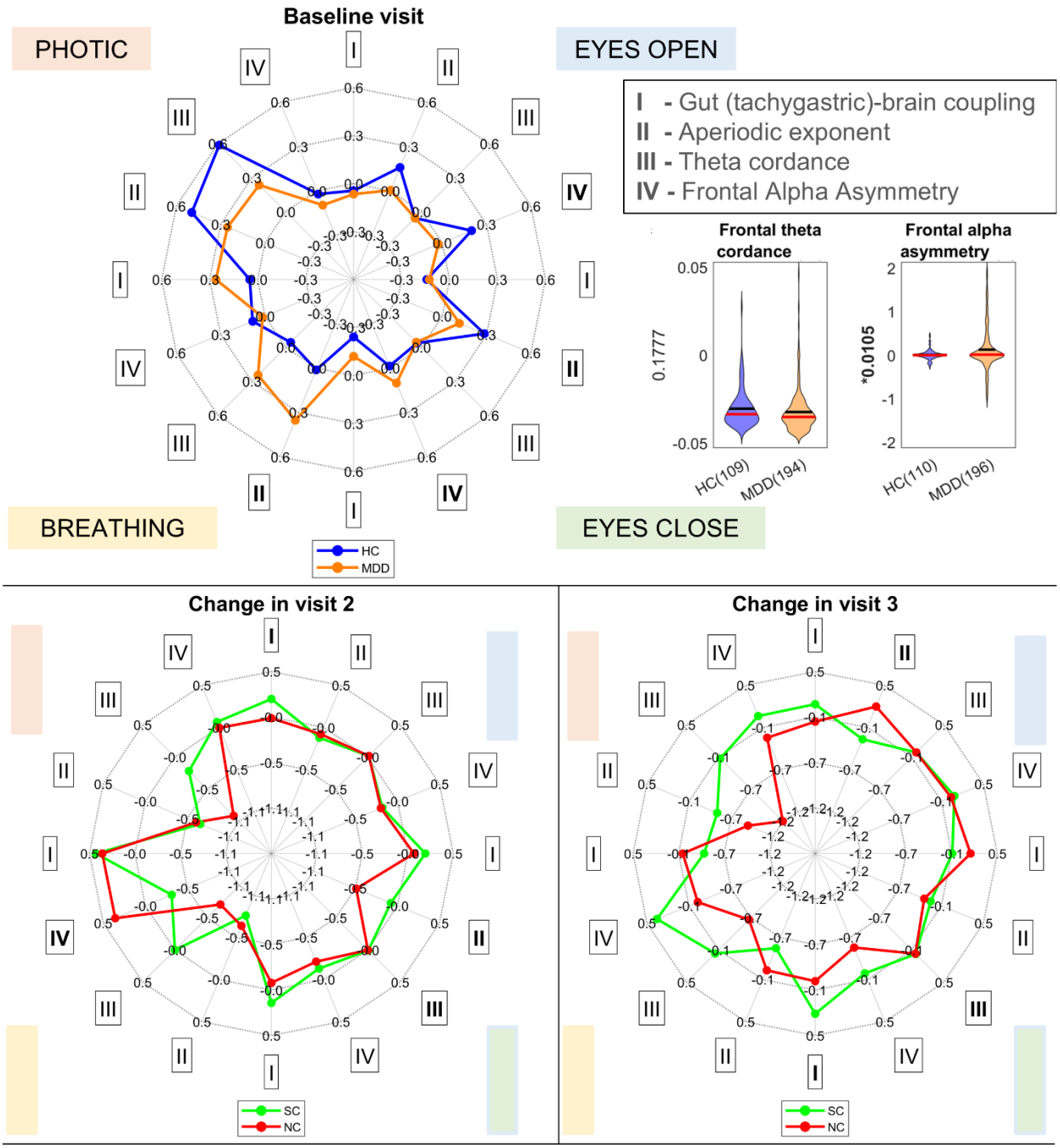


Figure 3. Summary features based on theories of depression. We explore four different theories of depression treatment prediction- increased gut-brain coupling (I), lower aperiodic exponent (II) and lower theta cordance (III), higher frontal alpha asymmetry (IV), when compared to baseline predicts significant response to treatment. Many of these features were sensitive to the cognitive state context to significantly predict the response outcome in patients in baseline visit 1, visit 2 or visit 3. The boldened Greek numerals represent significant differences in the healthy versus patient samples (baseline) or the responders versus non-responders (visits 2, 3) in two sample hypothesis testing, not controlled for multiple comparisons.

A perceptron model can efficiently early predict patients showing significant response to treatment.

Finally, we asked whether we can early predict the treatment response right at visit 2? Forty eight participants had complete EEG and EGG data for all the tasks of all 3 visits and were included in developing the predictive model **(Figure 1C)**. We used multi-layer perceptron (MLP) and simple recurrent neural network (RNN) models to classify the data for those patients that will exhibit significant reduction of mental health severity (SC), and no change in mental health (NC), controls (HC). The feature selection process resulted in electrophysiological features: 12 (baseline) + 11 (longitudinal) from resting state eyes open task, 10+22 from state eyes closed task, 6+14 in breathing task and the remaining 7+12 while performing photic administration task (refer **Supplementary table 1–** for the feature list), demographics (5) and questionnaires (24). First, we deployed 3-class models, they were validated using k-fold cross validation (k=5) run for 10 instances. The overall cross-validation mean accuracy for MLP was >74% (~37/48) and for RNN was >70% (34/48) which was well above the chance level of 33% (**Table 3A**).

We also tested two class models specifically trained on patients data, to early predict whether a patient will show significant change/no change in mental health by end of the treatment course. For this question, we compared the performances of the earlier described MLP and RNN architectures but for two class output, along with another Logistic Regression. All three models were trained and cross-validated using 28 subjects and performed well above the chance level of 50%. The MLP model outperformed the other 2 with an accuracy of ~81% (23/28), while the accuracy of the Logistic Regression of ~79% (22/28) and RNN was ~75% (21/28). The sensitivity of the 2-class MLP model to predict a participant with no change in mental health was ~84%. The relative risk of our model, computed as the proportion of subjects falsely predicted as non-responders is 2 subjects (16%). The Matthew's coefficient accounting for all the depressed subjects is 0.62. The specificity of the model for predicting significant change in mental health is 78% (**Table 3B**). Out of all the 5 individual models and also their nesting architecture, the MLP architecture had a least relative risk, and the model was selected to further explain the importance of each feature towards the prediction using SHAPley values.

A

Model	**#classes** **3: (HC, NC, SC)** **2: (SC, NC)**	**Accuracy%**	**Sensitivity% (NC)**	**Specificity% (SC)**
Multilayer Perceptron	3	78.3	73.1	76.7
	2	**81**	**84.1**	77.9
Recurrent Neural Networks	3	72.5	61.4	72.5
	2	75	75.4	74.7
Logistic Regression	2	80	80.9	**79.2**

B

Multilayer Perceptron		Predicted	
		NC	**SC**
Actual	**NC**	10.6 (±1.4)	3.4 (±1.4)
	SC	2 (±1)	12 (±1)

Table 3: Model selection for predicting treatment outcome. A) Based on model comparisons, we observe that 2class MLP is best suited for predicting participants with no change in mental health and Logistic Regression is more sensitive for predicting significant change in mental health. **B)** Confusion matrix from the 2class MLP model analysis predictions on 10 simulations.

Feature marginal contribution is sensitive to symptom presentation in patients.

We computed Feature importance as the marginal contribution of top 20 feature groups towards response prediction (**Figure 4**) in the 2-class multi layer perceptron. The results suggest that overall, changes in left fronto-central absolute theta power, global coherence across bands during photic task and coherence between regions other than central left during eyes closed task are higher for participants showing significant change in mental health. On the other hand, changes in coherence during breathing task, central left coherence during eyes closed task, beta asymmetry between hemispheres (frontal), periodic theta power in frontal left region, aperiodic exponent, and tachygastric gut rhythm-brain coupling decreases.

We also observed differences in the features based on the specificity of symptom severity. Longitudinally, in interoceptive breathing task, global beta coherence and tachygastric gut rhythm coupling with central brain regions reduced for responders independent of the symptom manifestation. Aperiodic exponent decreased selectively for responders with high sleep symptom manifestation. Similarly, left fronto-central theta absolute power increased selectively for responders with social symptoms. Longitudinally, decrease in beta frontal asymmetry and left-fronto-central theta periodic power during breathing task were significantly sensitive to treatment response for all symptom groups except anxiety. Overall, we observed different electrophysiological biomarkers to be predictive of treatment outcome depending on their psychotype and physiological symptom presentation (**Table 4**).

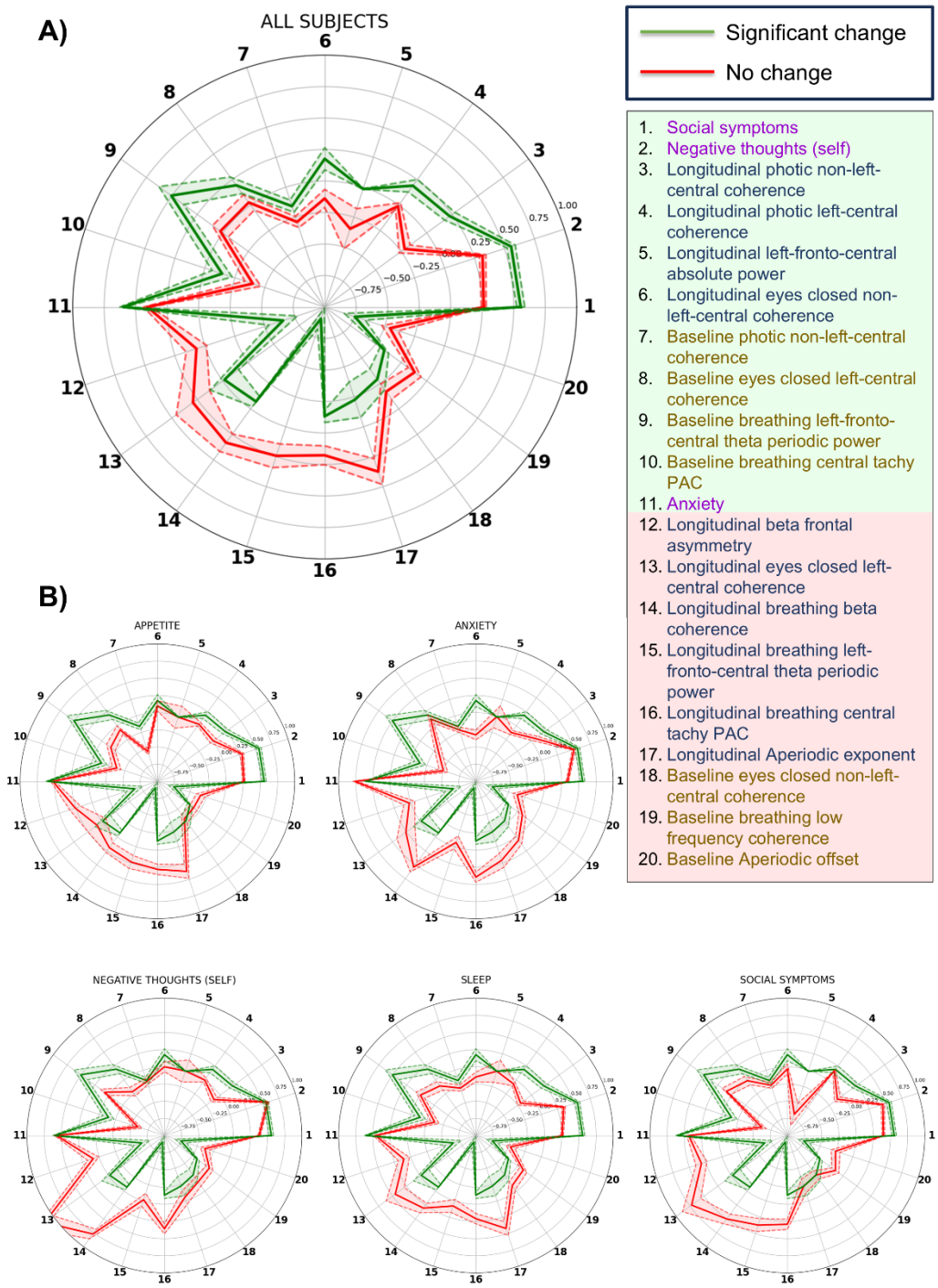


Figure 4: Interpretable and Predictive Modeling of Treatment Outcome. A) Zscored normalized comparison between responders (SC-green) and non-responders (NC-red) of top 20 features significantly contributing the multi layer perceptron 2-class model performance. Shaded region represents s.e.m for N=28, the green legend shading reflects feature increase for response while the red indicates the contrary. **B)** The schematic feature profiles for selected participants with high severity in appetite, anxiety, negative thoughts about self, sleep and social symptoms.

	Anxiety	Appetite	Negative thoughts (self)	Sleep	Social symptoms
Social symptoms	0.333±0.22, 0.622±0.28	0.259±0.22, 0.593±0.26	0.378±0.17, 0.633±0.27	0.256±0.18, 0.574±0.28	0.397±0.1, 0.581±0.26
Negative thoughts (self)	0.5±0.12, 0.667±0.26	0.296±0.25, 0.58±0.33	0.578±0.08, 0.689±0.23	0.344±0.26, 0.62±0.26	0.46±0.18, 0.573±0.29
Longitudinal photic non-left-central coherence	0.064±0.45, 0.181±0.43	0.004±0.45, 0.284±0.5	-0.109±0.53, 0.131±0.42	-0.196±0.46, 0.222±0.45	-0.101±0.51, 0.174±0.4
Longitudinal photic left-central coherence	-0.123±0.37, 0.254±0.52	0.029±0.26, 0.331±0.5	-0.001±0.31, 0.217±0.55	-0.07±0.3, 0.248±0.52	0.161±0.17, 0.147±0.51
Longitudinal left-fronto-central absolute power	-0.012±0.01, -0.018±0.02	-0.018±0.01, -0.016±0.02	-0.012±0.01, -0.016±0.02	-0.015±0.01, -0.016±0.02	-0.675±1.93, -0.015±0.02
Longitudinal eyes closed non-left-central coherence	-0.323±0.33, 0.377±0.78	0.098±0.47, 0.237±0.84	0.001±0.56, 0.293±0.78	-0.147±0.66, 0.273±0.75	-0.03±0.49, 0.216±0.75
Baseline photic non-left-central coherence	-0.213±0.24, -0.214±0.42	-0.542±0.29, -0.104±0.37	-0.171±0.38, -0.216±0.42	-0.274±0.39, -0.215±0.41	-0.231±0.36, -0.137±0.42
Baseline eyes closed left-central coherence	0.119±0.61, 0.259±0.67	-0.074±0.67, 0.345±0.77	-0.211±0.39, 0.263±0.65	-0.108±0.56, 0.241±0.71	-0.012±0.64, 0.264±0.65
Baseline breathing left-fronto-central theta periodic power	-0.34±0.41, 0.207±0.89	-0.16±0.43, 0.7±1.15	0.072±0.98, 0.367±1.1	0±0.71, 0.398±1.02	0.095±0.86, 0.521±1.05
Baseline breathing central tachy PAC	-0.501±0.46, -0.237±0.72	-0.376±0.62, -0.106±0.82	-0.593±0.45, -0.186±0.7	-0.368±0.59, -0.189±0.74	-0.609±0.45, -0.188±0.66
Anxiety	0.762±0.11, 0.7±0.13	0.524±0.32, 0.593±0.22	0.571±0.24, 0.652±0.18	0.462±0.27, 0.631±0.19	0.435±0.26, 0.626±0.18
Longitudinal beta frontal asymmetry	0.021±1.17, -0.331±0.81	0.259±0.71, -0.612±1.12	0.09±1.06, -0.398±0.83	0.105±0.83, -0.588±1.04	0.315±0.51, -0.48±0.79
Longitudinal eyes closed left-central coherence	0.261±0.29, 0.063±1.24	0.087±0.39, 0.249±1.33	1.154±1.95, 0.032±1.26	0.453±1.59, 0.11±1.2	0.723±1.76, -0.106±1.15
Longitudinal breathing beta coherence	0.543±0.78, -0.026±0.47	0.192±0.12, -0.278±0.33	0.772±0.79, -0.054±0.47	0.305±0.79, -0.08±0.45	0.501±0.56, -0.039±0.41
Longitudinal breathing left-fronto-central theta periodic power	-0.066±0.93, -0.586±0.43	0.235±0.71, -0.969±0.82	-0.018±0.84, -0.623±0.47	0.078±0.75, -0.825±0.75	0.37±0.91, -0.753±0.5
Longitudinal breathing central tachy PAC	0.394±0.74, -0.113±0.4	0.282±0.6, -0.176±0.38	0.356±0.61, -0.143±0.39	0.21±0.67, -0.111±0.42	0.291±0.56, -0.13±0.39
Longitudinal Aperiodic exponent	0.209±1.09, -0.294±1.34	0.377±1.03, -0.299±1.41	-0.049±1.09, -0.214±1.36	0.422±1.02, -0.322±1.3	-0.228±0.68, -0.167±1.24
Baseline eyes closed non-left-central coherence	0.058±0.38, -0.342±0.76	-0.33±0.57, -0.054±0.92	-0.167±0.53, -0.318±0.76	-0.1±0.66, -0.374±0.72	-0.29±0.44, -0.251±0.86
Baseline breathing low frequency coherence	-0.164±0.5, -0.347±0.37	-0.348±0.54, -0.416±0.37	-0.193±0.46, -0.362±0.37	-0.147±0.54, -0.393±0.37	-0.135±0.4, -0.383±0.34
Baseline Aperiodic offset	-0.366±0.68, -0.917±0.72	-0.339±0.56, -0.648±0.79	-0.381±0.64, -0.94±0.71	-0.471±0.65, -0.901±0.67	-0.263±0.45, -0.73±0.75

Table 4: Significant features for specific biotypes- Electrophysiological feature comparison for responders and non-responders in a subset of population with specific symptom severity, with relatively higher symptom score than the population median. The above table highlights the feature group from the top 20 that are significantly different between the responders and non-responders. The cells highlighted with green has a greater mean for SC than NC and vice versa for red cells. The cells depict the mean and std dev of NC and SC datapoints. p-value<0.05 was defined as significant.

DISCUSSION

Criticism against watchful waiting. Current medicine and supportive therapies including current or magnetic stimulation or those activities facilitating cognitive restructuring demand constant and close monitoring of the patient even while the course of therapy (off-sessions) to reliably predict the treatment outcome. However, due to increasing patient to clinician ratio and subjective evaluation, it is hard to successfully predict the efficacy of the treatment plan through the conventional in-person interviews. The process makes the patients suffer from unpleasant side-effects, including Psychosis, Seizures, and even gut dysfunctions, with extreme being suicidality. More than a third of patients even drop-out from the treatment plan not able to withstand the side-effects by 30 days of treatment. If the depression isn't managed after two different medication types, the patient is likely to be called treatment-resistant. This process of "watchful waiting" or "sequential treatment" is heavily critized in the clinical community for the amount of resources utilized by this process and yet only a minority (less than 1/5th) of patients benefit from the process according to studies. It is crucial to arrive at the correct treatment plan in the shortest time possible for any patient, for which firstly a prediction of current treatment outcome is necessary sooner than the current standards of 4-6 weeks (Bauer et al., 2007; Bockting et al., 2008; Fochtmann & Gelenberg, 2005; Pigott et al., 2010; Quitkin et al., 2003).

Our results suggest that though changes in frontal brain measures of connectivity can be indicative of treatment response after 30 days, gut-brain connectivity features were more significantly predicting early within 2 weeks of treatment onset. Interestingly, we observe that the trackable features for precision medicine should be chosen based on patient biotype (symptom presentation profile) whether they present relatively higher issues with regards to anxiety, sleep, appetite, negative thoughts, or social behavior.

Both periodic and aperiodic activations predict treatment outcome. Frontal changes in various bands have been markers of depressive symptoms and its treatment response according to numerous studies. Some specific markers include increased theta cordance in depressed individuals especially in the frontal region (de la Salle et al., 2020). However, there has been contradictory viewpoints on the oscillatory mechanistic underpinnings because when separating the theta power into its aperiodic and periodic components, the aperiodic one reflected the depression measures relatively (Smith et al., 2023). This opens up a debate on the processes mediating the severity of depression and their responses to intervening treatment. We teased apart the periodic components of oscillation to specifically understand the role of theta mechanism in predicting response to medicine. Interestingly, periodic theta power in the left frontal and central region reduced post medication as mentioned in previous literature. On the other hand, aperiodic parameters such as slope and offset played a crucial role in categorizing someone as a responder/non-responder highlighting the importance of using aperiodic parameters as a biomarker for treatment outcome prediction. Aperiodic offset and aperiodic exponent are a measure of global

signal power and E/I ratio. A greater offset and a lower exponent value indicates balanced power between bands and increased global power. A steeper exponent indicates greater lower frequency power when compared to higher frequency power and is commonly observed in state of relaxation. Antidepressants such as SSRI, SNRI are said to increase the global band power of the signal.

Frontal beta asymmetry can early predict treatment outcome. The slowness of brain activity, in other words the power of lower frequency spectrum was related to treatment response by many studies even in alpha band. The frontal alpha asymmetry between the left and right hemispheres were in particular a signature of depression severity. Studies suggest that asymmetry in the brain facilitated a differentiating response to positive versus negative affect, with more attenuation to positive affect and increased response to negative affect (Davidson, 2003; Harmon-Jones et al., 2010). However, the frontal alpha asymmetry was not consistently observed to be indicative of treatment response, and rather as a marker of depression experience (van der Vinne et al., 2019). The positive or negative affective stimulus induced emotional responses were also separately studied by few scientists, to investigate the elicited emotional sensitivity to depression severity and treatment response, but there weren't any robust results (Kaviani et al., 2004; Kołodziej et al., 2021). In our study, we did not observe strong effect of the frontal alpha power for predicting response. Interestingly, we observed frontal beta asymmetry to be sensitive to treatment response and the beta power in the frontal left region relative to the right region increased significantly for depressed individuals who responded to medication. Some studies specifically looked into reward processing and inhibitory control mechanisms for understanding the severity (Oh et al., 2023; Yitzhak et al., 2023) for response to treatment (Brandt et al., 2021): Not-surprisingly, evidences point to the significant presence of reward information in the frontal beta oscillations, suggesting their potential role for early predicting of outcome (Koloski et al., 2024).

Brain connections and its gut modulatory indices hold important information on treatment effects. Recently, the role of connectivity has garnered more attention (Elam et al., 2021), and a reduced connectivity were found in people experiencing depression (Huang et al., 2023; Roemer et al., 1992). More structural level investigation also suggested the possibility of a short-range excitation and long-range reduction in connections in depression (Fingelkurts et al., 2007). In our study, we find a significant contribution of the connectivity features, in all the theta, alpha and beta bands, for predicting the treatment response.

Our study for the first time rigorously tests the role of gut-brain coupling in explaining depression severity and response to treatment. Interestingly, we find it to be predicting the response to treatment as early as within two weeks from the treatment onset. The role of gut dysfunction in depression has been investigated by numerous studies, however to our understanding, the gut dysfunction exhibited as abnormal motility of the stomach and the intestine has not been explored to strategizing the treatment for depression. Our study strongly suggests that looking into the gut-brain coupling can assist with personalizing treatment strategy in an optimal fashion. Importantly, it highlights that a scalable electrogastrography tool could be used for the purpose of computing the gut-brain coupling. In our population, we observed that the brain-gut coupling in tachygastric frequency reduces significantly for responders because of medication.

Furthermore, our study also proposes that longitudinal design that can track the plasticity in neural circuits are able to explain the depression measures better than the cross-sectional design of studying and developing of prediction models based on the baseline time sample in silo. This is

in lines with many studies that find greater sensitivity for a longitudinal marker in depression (Bares et al., 2015; Schwartzmann et al., 2024).

Simple cognitive tasks as a probe for desired neural circuit activations. Another important observation of our study includes the sensitivity of cognitive control by administration of a sensitive task paradigm for acquiring markers of treatment response. In addition to the resting state eyes open and eyes closed paradigms, we find highly predictive markers during photic flicker presentation and hyperventilating paradigms. This flicker effect has been shown to facilitate various cognitive processes based on the frequency of stimulus, and even the processing of external stimulus or thoughts based on the phase coupling with the flicker onset (Thut et al., 2011; Williams et al., 2006). The effect has been also found to be differentiating depression in lines with our observation.

Limitation. Our study has many limitations. Our sample was limited to a part of India. The study does not account for the role of genetics (Shadrina et al., 2018), social and cultural factors (Kupferberg & Hasler, 2023) influencing the onset and progression of mental health issues, and designing of treatment strategies. Our future plan is to extend the study to investigating multiple national and international sites for validation of the identified markers of prediction. Moreover, many clinicians administer multimodal treatment, that is medicine in combination with alternative treatments, repetitive transcranial magnetic stimulation, as an intervention. Our future work will also include understanding of the physiological differences due to treatment modality for predicting outcome.

Acknowledgements

We thank Jyoti Mishra, Dhakshin Ramanathan, Venkatasubramanian, Srinivasa Chakravarthy, for some insightful discussions.

REFERENCE

1. Arevalo-Rodriguez, I., Smailagic, N., i Figuls, M. R., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O. L., Cosp, X. B., & Cullum, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, *3*.

2. Athreya, A. P., Brückl, T., Binder, E. B., John Rush, A., Biernacka, J., Frye, M. A., Neavin, D., Skime, M., Monrad, D., Iyer, R. K., Mayes, T., Trivedi, M., Carter, R. E., Wang, L., Weinshilboum, R. M., Croarkin, P. E., & Bobo, W. V. (2021). Prediction of short-term antidepressant response using probabilistic graphical models with replication across multiple drugs and treatment settings. *Neuropsychopharmacology*, *46*(7), 1272–1282. https://doi.org/10.1038/s41386-020-00943-x

3. Bares, M., Novak, T., Kopecek, M., Brunovsky, M., Stopkova, P., & Höschl, C. (2015). The effectiveness of prefrontal theta cordance and early reduction of depressive symptoms in the prediction of antidepressant treatment outcome in patients with resistant depression: Analysis of naturalistic data. *European Archives of Psychiatry and Clinical Neuroscience*, *265*(1), 73–82. https://doi.org/10.1007/s00406-014-0506-8

4. Bauer, M., Bschor, T., Pfennig, A., Whybrow, P. C., Angst, J., Versiani, M., & Möller, H. J. (2007). World federation of societies of biological psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders in primary care. *World Journal of Biological Psychiatry*, *8*(2), 67–104. https://doi.org/10.1080/15622970701227829

5. Bockting, C. L. H., ten Doesschate, M. C., Spijker, J., Spinhoven, P., Koeter, M. W. J., Schene, A. H., & DELTA study group. (2008). Continuation and maintenance use of antidepressants in recurrent depression. *Psychotherapy and Psychosomatics*, *77*(1), 17–26. https://doi.org/10.1159/000110056

6. Cao, Z., Lin, C.-T., Ding, W., Chen, M.-H., Li, C.-T., & Su, T.-P. (2019). Identifying Ketamine Responses in Treatment-Resistant Depression Using a Wearable Forehead EEG. *IEEE Transactions on Biomedical Engineering*, *66*(6), 1668–1679. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/TBME.2018.2877651

7. Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., & Knight, R. T. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, *23*(12), 1655–1665.

8. Drossman, D. A. (2016). Functional gastrointestinal disorders: History, pathophysiology, clinical features, and Rome IV. *Gastroenterology*, *150*(6), 1262–1279.

9. Fochtmann, L. J., & Gelenberg, A. J. (2005). Guideline Watch: Practice Guideline for the Treatment of Patients With Major Depressive Disorder, 2nd Edition. *FOCUS*, *3*(1), 34–42. https://doi.org/10.1176/foc.3.1.34

10. Goyal, O., Nohria, S., Dhaliwal, A. S., Goyal, P., Soni, R. K., Chhina, R. S., & Sood, A. (2021). Prevalence, overlap, and risk factors for Rome IV functional gastrointestinal disorders among college students in northern India. *Indian Journal of Gastroenterology*, *40*(2), 144–153. https://doi.org/10.1007/s12664-020-01106-y

11. Hamilton, M. (1960). The Hamilton Depression Scale—Accelerator or break on antidepressant drug discovery. *Psychiatry*, *23*, 56–62.

12. Hasanzadeh, F., Mohebbi, M., & Rostami, R. (2019). Prediction of rTMS treatment response in major depressive disorder using machine learning techniques and nonlinear features of EEG signal. *Journal of Affective Disorders*, *256*, 132–142. https://doi.org/10.1016/j.jad.2019.05.070

13. Jaworska, N., de la Salle, S., Ibrahim, M.-H., Blier, P., & Knott, V. (2019). Leveraging Machine Learning Approaches for Predicting Antidepressant Treatment Response Using Electroencephalography (EEG) and Clinical Data. *Frontiers in Psychiatry*, *9*. https://doi.org/10.3389/fpsyt.2018.00768

14. John Rush, A., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., George Niederehe, M., Thase, M. E., Lavori, P. W., Lebowitz, B. D., McGrath, P. J., Rosenbaum, J. F., Sackeim, H. A., Kupfer, D. J., Luther, J., & Maurizio Fava, M. (2006). *Acute and Longer-Term Outcomes in Depressed Outpatients Requiring*

*One or Several Treatment Steps: A STAR*D Report* (Am J Psychiatry, Vol. 163, Issue 11, pp. 1905–1917). www.star-d.org

15. Keitner, G. I., Solomon, D. A., & Ryan, C. E. (2008). STAR*D: Have We Learned the Right Lessons? *American Journal of Psychiatry*, *165*(1), 133–133. https://doi.org/10.1176/appi.ajp.2007.07081375

16. Knott, V., Mahoney, C., Kennedy, S., & Evans, K. (2001). EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Research: Neuroimaging*, *106*(2), 123–140.

17. Kołodziej, A., Magnuski, M., Ruban, A., & Brzezicka, A. (2021). No relationship between frontal alpha asymmetry and depressive disorders in a multiverse analysis of five studies. *eLife*, *10*, e60595. https://doi.org/10.7554/eLife.60595

18. Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613.

19. Lee, H. S., Baik, S. Y., Kim, Y.-W., Kim, J.-Y., & Lee, S.-H. (2020). Prediction of Antidepressant Treatment Outcome Using Event-Related Potential in Patients with Major Depressive Disorder. *Diagnostics*, *10*(5), Article 5. https://doi.org/10.3390/diagnostics10050276

20. Leuchter, A. F., Cook, I. A., Hunter, A. M., & Korb, A. S. (2009). *The current treatment paradigm for Major Depressive Disorder A new paradigm for the prediction of antidepressant treatment response* (Dialogues Clin Neurosci, Vol. 11, pp. 435–446). www.dialogues-cns.org

21. Lieber, A. L., & Prichep, L. S. (1988). Diagnosis and subtyping of depressive disorders by quantitative electroencephalography: I. Discriminant analysis of selected variables in untreated depressives. *The Hillside Journal of Clinical Psychiatry*, *10*(1), 71–83.

22. Lydiard, R. B. (2001). Irritable bowel syndrome, anxiety, and depression: What are the links? *Journal of Clinical Psychiatry*, *62*, 38–47.

23. Masand, P. S., Kaplan, D. S., Gupta, S., Bhandary, A. N., Nasra, G. S., Kline, M. D., & Margo, K. L. (1995). Major depression and irritable bowel syndrome: Is there a relationship? *The Journal of Clinical Psychiatry*, *56*(8), 363–367.

24. Overs, J., Morgan, S., Apputhurai, P., Tuck, C., & Knowles, S. R. (2024). Comparing the prevalence and association between anxiety, depression and gastrointestinal symptoms in gastroparesis versus functional dyspepsia: A systematic review and meta-analysis. *Journal of Psychosomatic Research*, *183*, 111834. https://doi.org/10.1016/j.jpsychores.2024.111834

25. Pigott, H. E., Leventhal, A. M., Alter, G. S., & Boren, J. J. (2010). Efficacy and effectiveness of antidepressants: Current status of research. *Psychotherapy and Psychosomatics*, *79*(5), 267–279. https://doi.org/10.1159/000318293

26. Quitkin, F. M., Petkova, E., McGrath, P. J., Taylor, B., Charles Beasley, Mp., Stewart, J., Amsterdam, J., Fava, M., Rosenbaum, J., Reimherr, F., Fawcett, J., Chen, Y., & Klein, D. (2003). *Article When Should a Trial of Fluoxetine for Major Depression Be Declared Failed?* (Am J Psychiatry, Vol. 160, pp. 4–4). http://ajp.psychiatryonline.org

27. Sadat Shahabi, M., Shalbaf, A., & Maghsoudi, A. (2021). Prediction of drug response in major depressive disorder using ensemble of transfer learning with convolutional neural network based on EEG. *Biocybernetics and Biomedical Engineering*, *41*(3), 946–959. https://doi.org/10.1016/j.bbe.2021.06.006

28. Smith, S. E., Ma, V., Gonzalez, C., Chapman, A., Printz, D., Voytek, B., & Soltani, M. (2023). Clinical EEG slowing induced by electroconvulsive therapy is better described by increased frontal aperiodic activity. *Translational Psychiatry*, *13*(1), 1–10. https://doi.org/10.1038/s41398-023-02634-9

29. Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

30. Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, *5*, 1–13. https://doi.org/10.1186/1477-7525-5-63

31. Trivedi, M. H., John Rush, A., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Louise Ritz, M., Grayson Norquist, M., Howland, R. H., Lebowitz, B., McGrath, P. J., Shores-Wilson, K., Biggs, M. M., Balasubramani, G., & Fava, M. (2006). *Article Evaluation of Outcomes With Citalopram for Depression Using Measurement-Based Care in STAR*D: Implications for Clinical Practice STAR*D Study Team* (Am J Psychiatry, Vol. 163, pp. 1–1). http://ajp.psychiatryonline.org

32. van der Vinne, N., Vollebregt, M. A., van Putten, M. J. A. M., & Arns, M. (2019). Stability of frontal alpha asymmetry in depressed patients during antidepressant treatment. *NeuroImage: Clinical*, *24*, 102056. https://doi.org/10.1016/j.nicl.2019.102056

33. Warden, D., Madhukar Trivedi, M. H., Wisniewski, S. R., Davis, L., Nierenberg, A. A., Gaynes, B. N., Sidney Zisook, M., Hollon, S. D., Balasubramani, G., Howland, R., Fava, M., Stewart, J. W., & John Rush, A. (2007). *Predictors of Attrition During Initial (Citalopram) Treatment for Depression: A STAR*D Report* (Am J Psychiatry, Vol. 164).

34. Zamani, M., Alizadeh-Tabari, S., & Zamani, V. (2019). Systematic review with meta-analysis: The prevalence of anxiety and depression in patients with irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics*, *50*(2), 132–143. https://doi.org/10.1111/apt.15325

SUPPLEMENTARY MATERIALS

1) ***Questionnaires question-wise differences for responders vs non-responders***

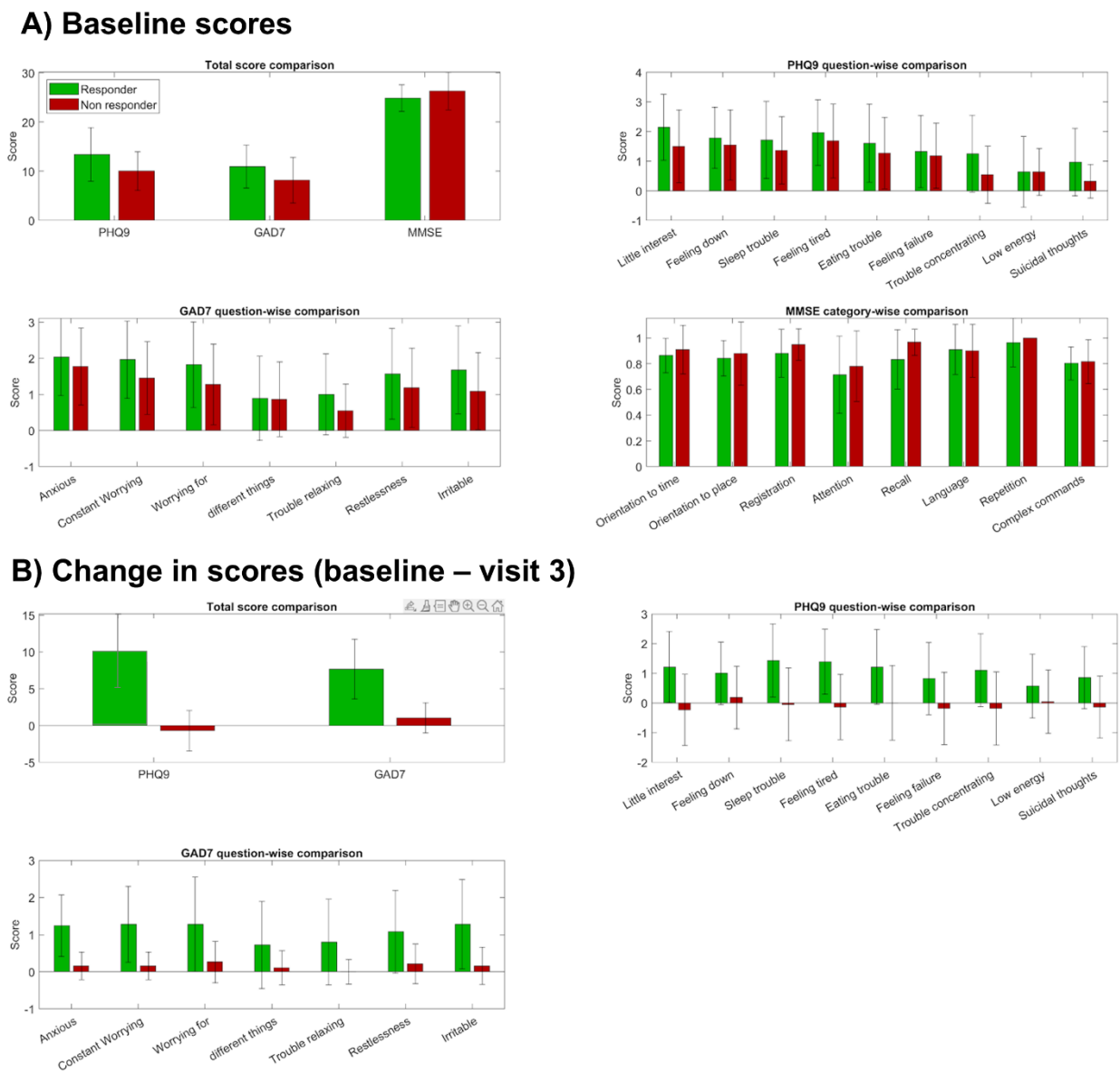
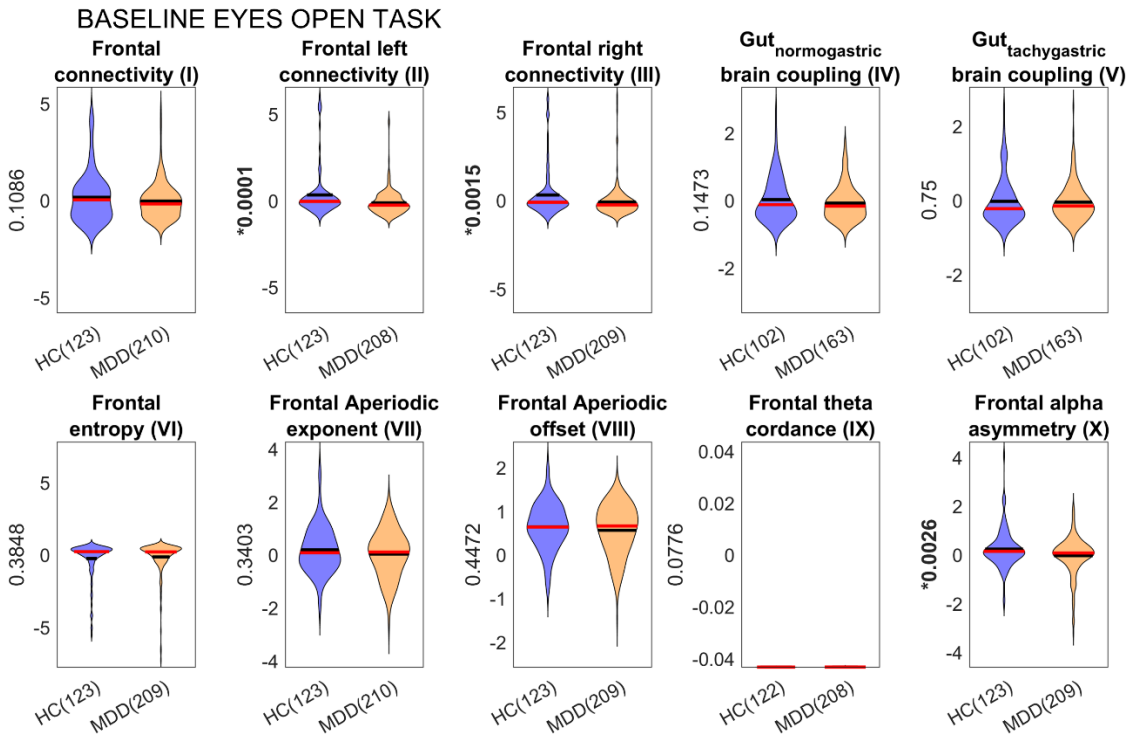
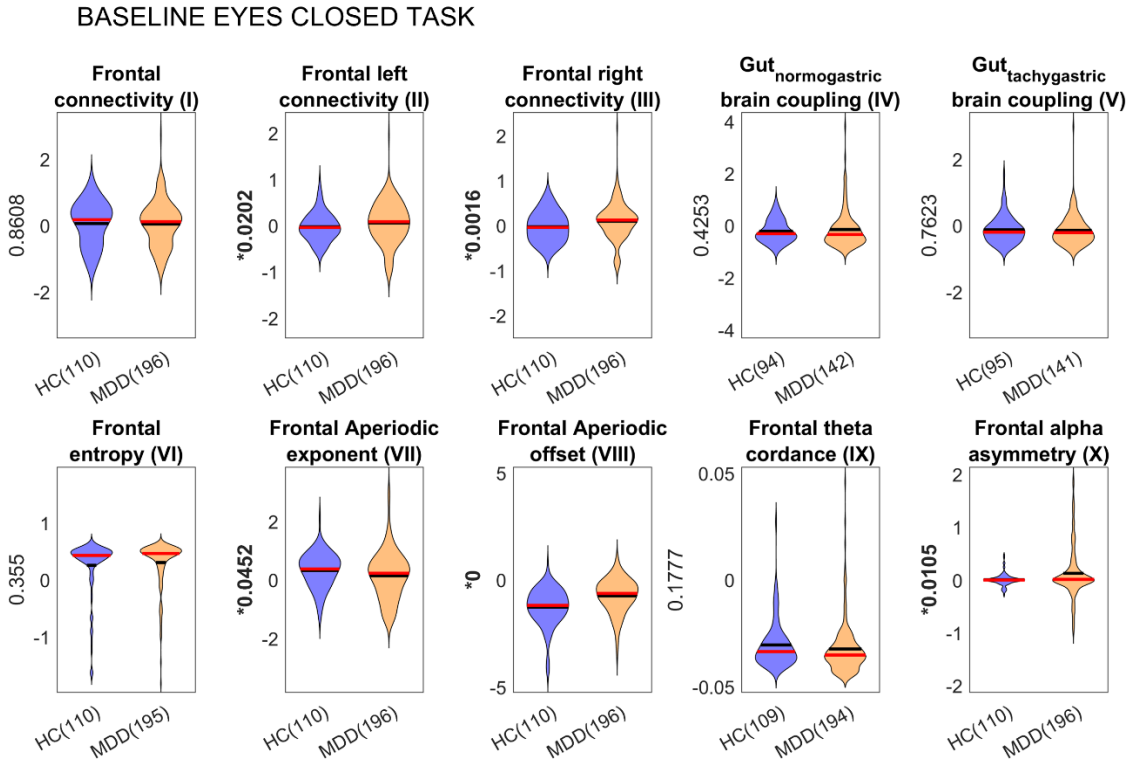


Figure I: Question-wise comparison for responders and non-responders.

From the above figure, we observe that there is no significant difference between the scores of responders and non-responders at the baseline. Post treatment, responders show significant decrease in severity across all symptoms when compared to non-responders.

2) *Feature comparison between participant groups*





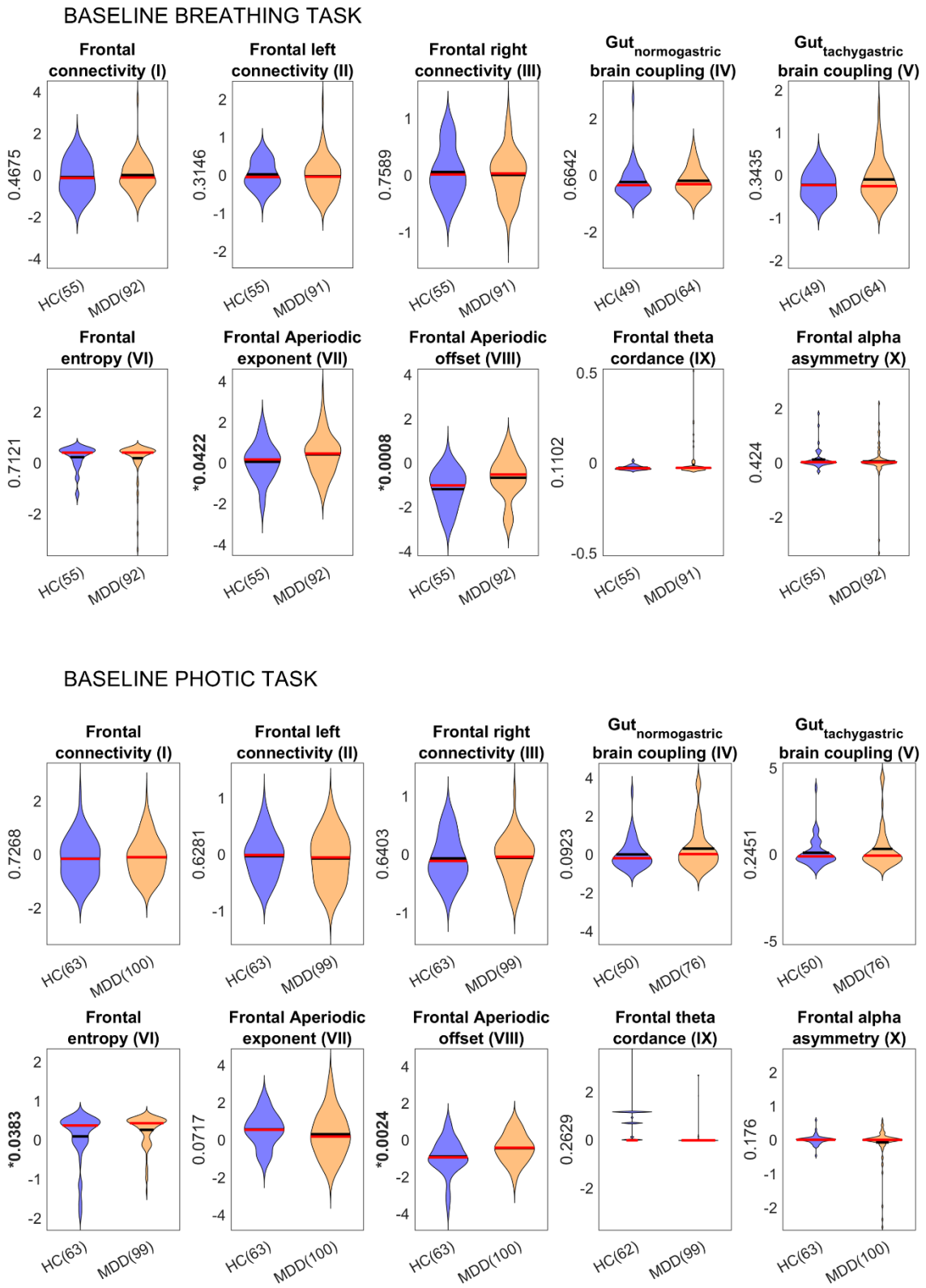
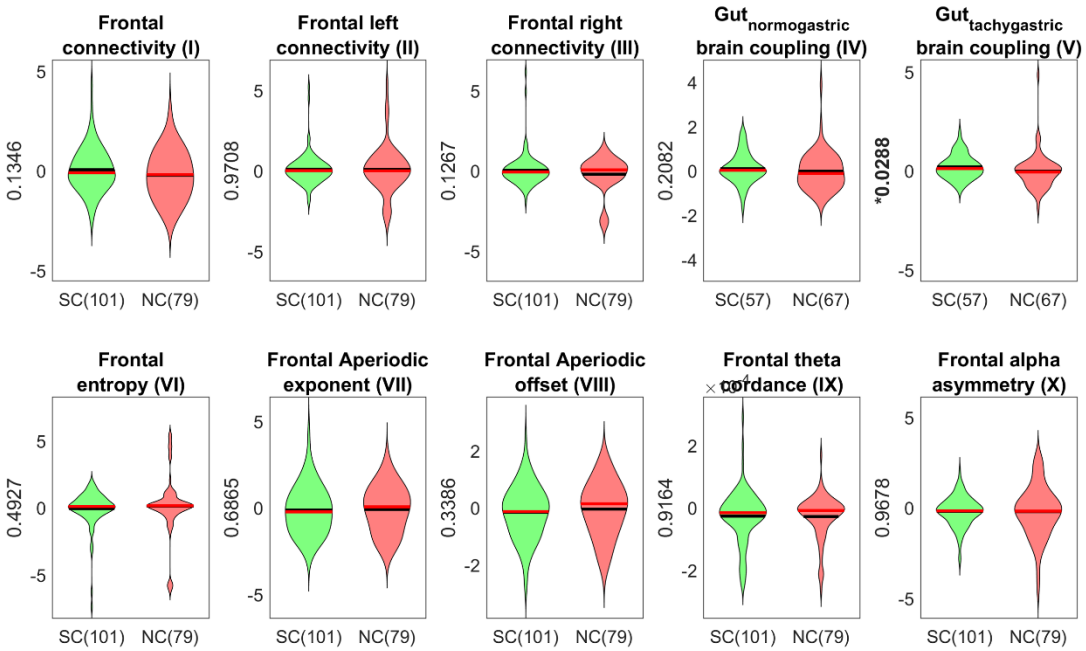
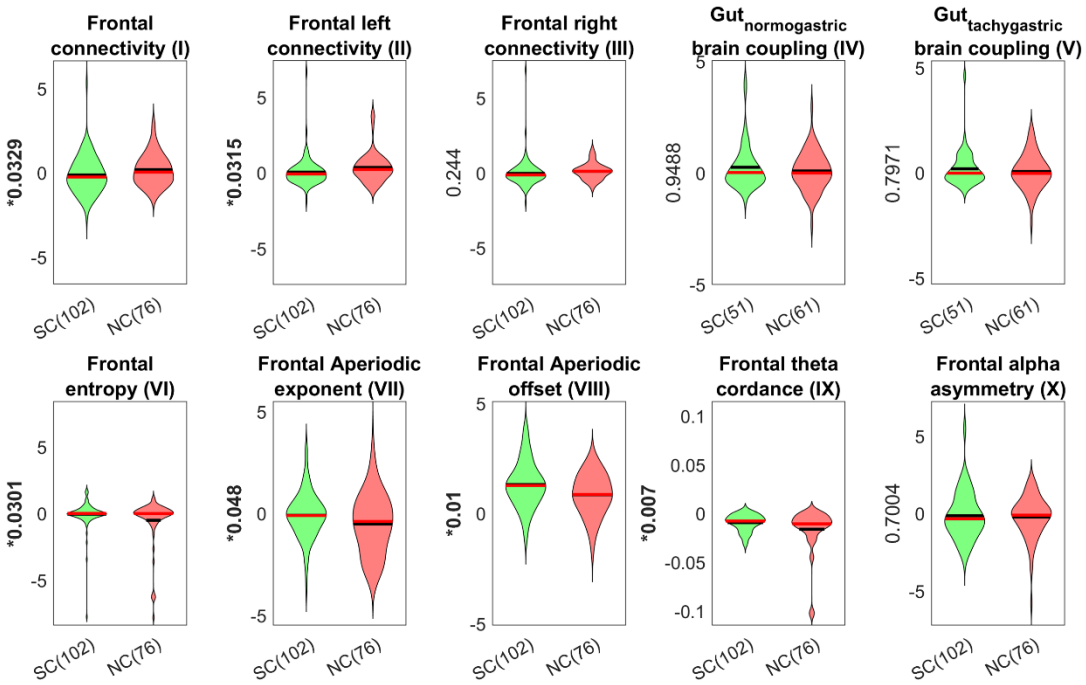


Figure II) Baseline summary feature comparison between healthy and depressed population.The p-values are uncorrected.

VISIT 2 – BASELINE EYES OPEN TASK



VISIT 2 - BASELINE EYES CLOSED TASK



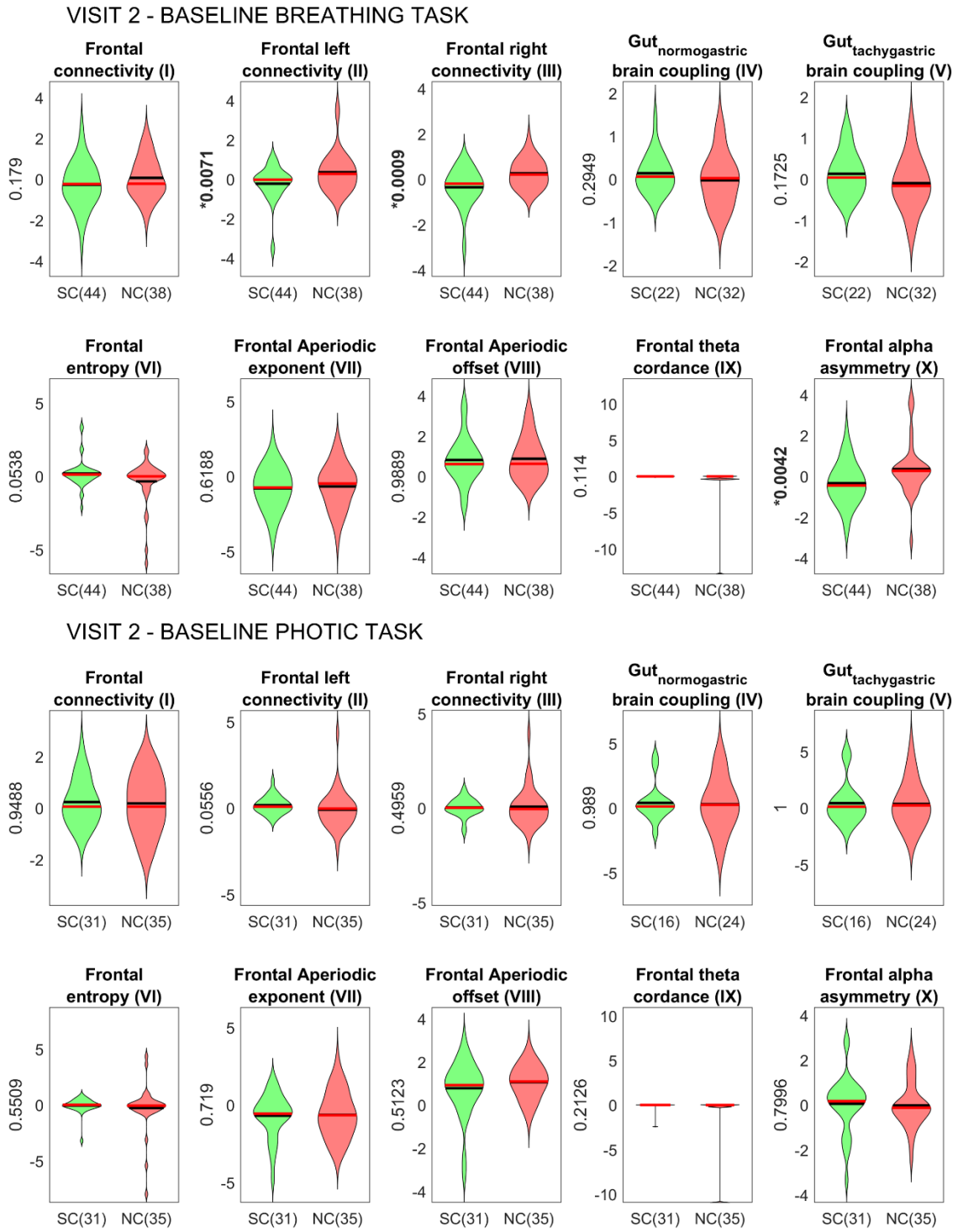
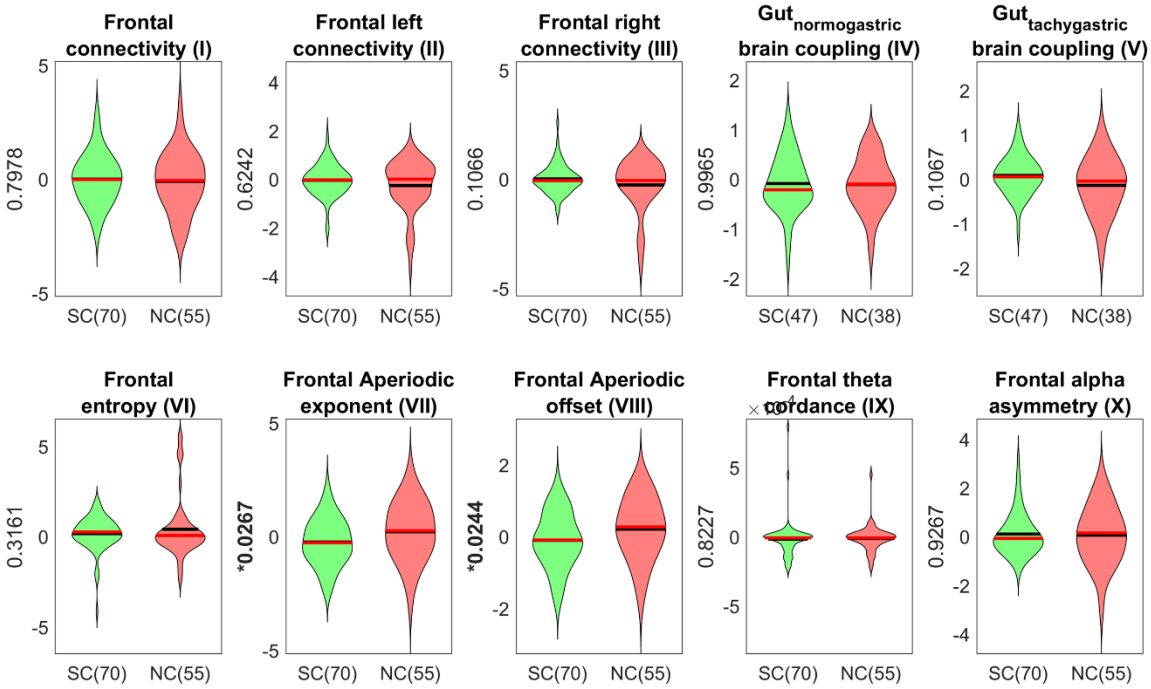
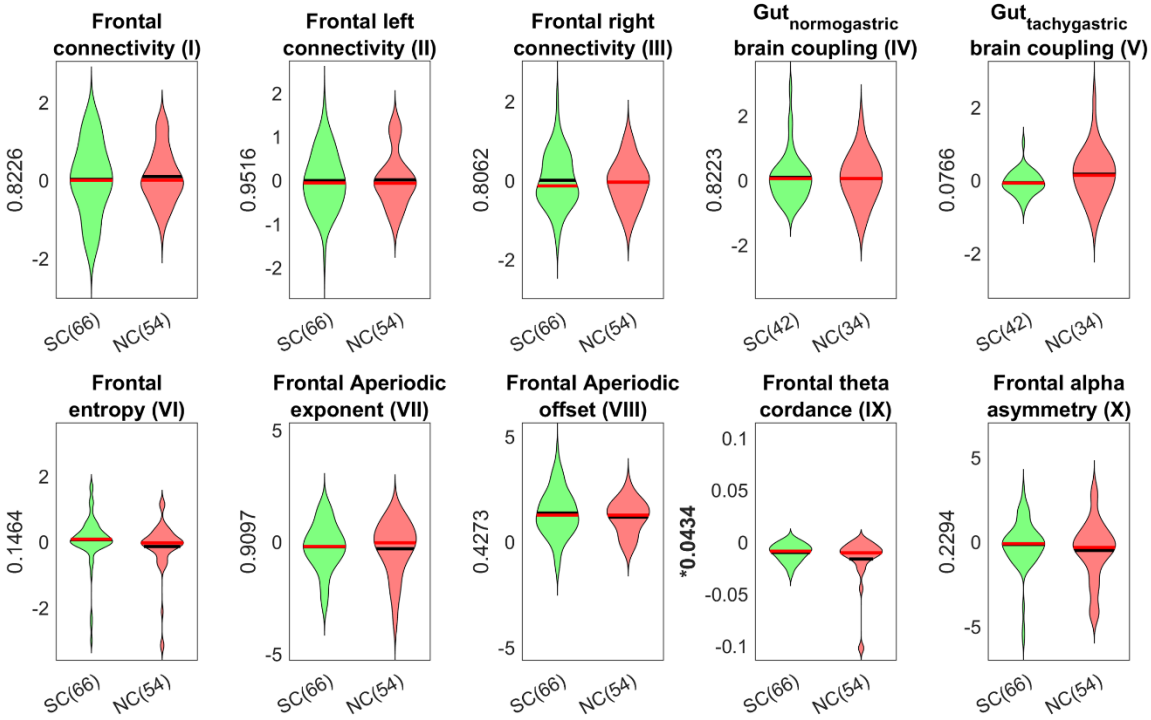


Figure III) Change in feature value at visit 2 from the baseline visit for responders and non-responders.

VISIT 3 – BASELINE EYES OPEN TASK



VISIT 3 - BASELINE EYES CLOSED TASK



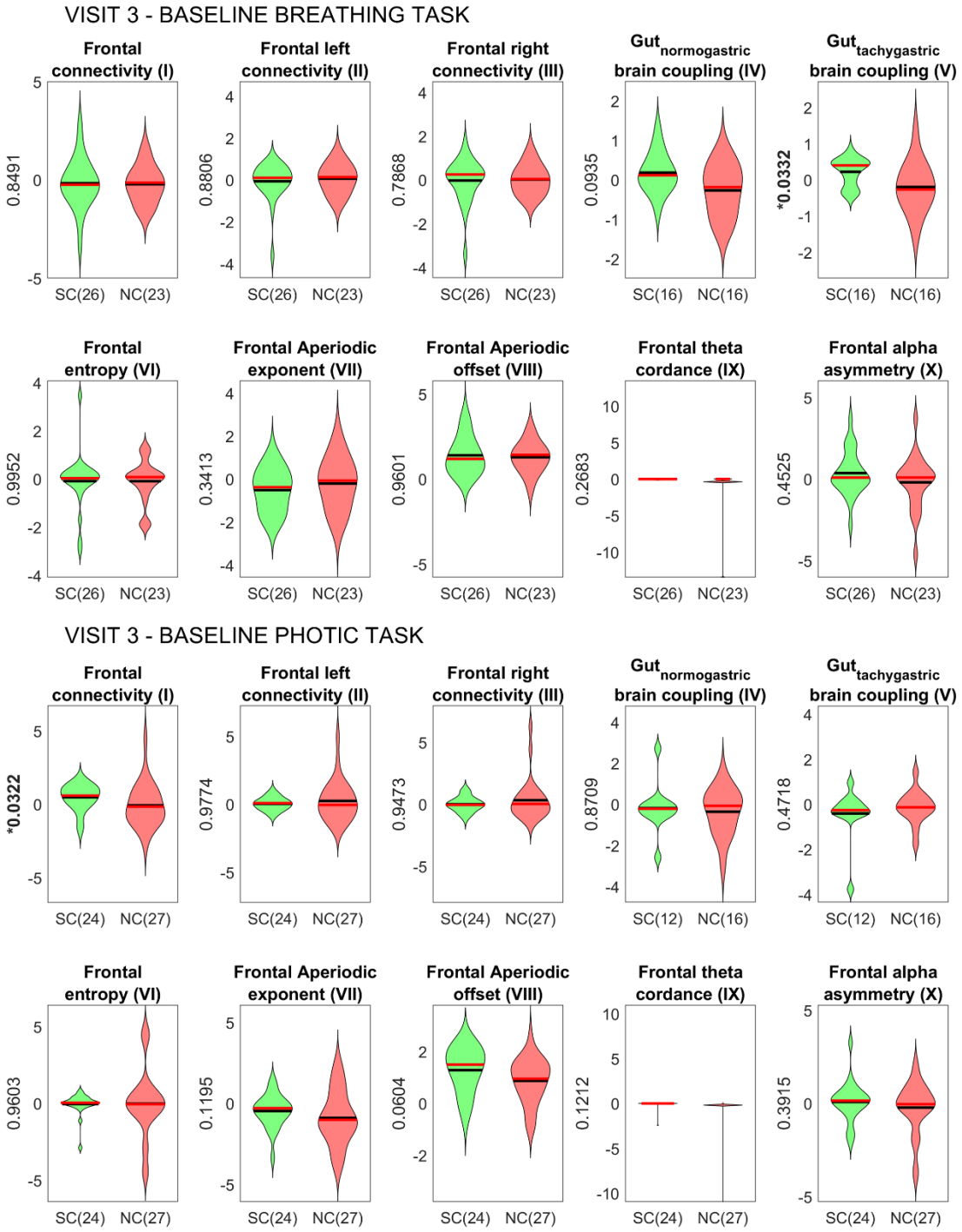


Figure IV) Change in feature value at visit 3 from the baseline visit for responders and non-responders.

From the above 3 figures (II,III,IV), we can observe that frontal connectivity with other regions, aperiodic parameters and frontal alpha asymmetry during eyes closed task are sensitive to presence of depression and also to treatment response initially. However, in the final visit, these features fail to distinguish responders and non-responders. None of the other features show reliability across more than 1 comparison for a particular task. Some of the features are distinguishing depressed population from healthy, while several other features are sensitive to treatment outcome depending on the timepoint they are probed.

3) *Final list of features, groups and their importance*

GROUP	FEATURE	SHAPLEY VALUE (NC)	SHAPLEY VALUE (SC)
-	age	5.08E-05	-1.16E-05
-	gender	-0.00018	-2.60E-05
-	handedness	-0.0004	2.90E-05
-	trauma	0.001623	0.002374
-	Patient/healthy	0.019194	0.002625
Social symptoms	Phq q1 v1	0.015715	0.011284
Negative thoughts (self)	Phq q2 v1	0.000317	0.000627
Sleep	Phq q3 v1	0.000407	-0.00169
Sleep	Phq q4 v1	-0.00048	-5.70E-05
Appetite	Phq q5 v1	-6.63E-05	-0.00034
Negative thoughts (self)	Phq q6 v1	0.002472	0.003175
Social symptoms	Phq q7 v1	0.004395	0.004879
Social symptoms	Phq q8 v1	0.000426	-0.00057
Negative thoughts (self)	Phq q9 v1	0.012225	0.009761
Anxiety	Gad q1 v1	-0.00026	-0.00018
Anxiety	Gad q2 v1	0.008919	0.006164
Anxiety	Gad q3 v1	0.001075	-0.00014

Anxiety	Gad q4 v1	-0.0003	2.63E-05
Anxiety	Gad q5 v1	0.002266	0.002683
Anxiety	Gad q6 v1	0.001378	0.000807
Anxiety	Gad q7 v1	0.001451	0.000319
Cognition	mmse category 1	0.000848	0.000434
Cognition	mmse category 2	0.000145	0.000244
Cognition	mmse category 3	2.77E-05	0.00072
Cognition	mmse category 4	0.001439	-0.00032
Cognition	mmse category 5	0.000982	0.00063
Cognition	mmse category 6	-0.00037	0.000279
Cognition	mmse category 7	0	0
Cognition	mmse category 8	-3.03E-05	9.72E-05
-	Average of base Entropy Fl T1 base Entropy Fr T1 base Entropy Cl T1 base Entropy Cr T1 base Entropy Ol T1 base Entropy Or T1	-4.82E-05	0.00052
Baseline_eyes_open_global_coherence	Average of base Coh th Fl-Cl T1 base Coh th Fr-Cl T1 base Coh th Cl-Ol T1 base Coh th Cl-Or T1 base Coh alp Fl-Cl T1 base Coh alp Fr-Cl T1 base Coh alp Cl-Ol T1 base Coh bet Fl-Cl T1 base Coh bet Fr-Cl T1 base Coh bet Cl-Ol T1 base Coh bet Cl-Or T1 base Coh bet Cr-Or T1	0.001456	0.001042
Baseline_eyes_open_global_coherence	Average of base Coh th Fl-Cr T1 base Coh th Fr-Cr T1 base Coh th Cl-Cr T1 base Coh th Cr-Ol T1 base Coh th Cr-Or T1 base Coh alp Fl-Cr T1 base Coh alp Fr-Cr T1 base Coh alp Cr-Or T1 base Coh bet Fl-Cr T1 base Coh bet Fr-Cr T1 base Coh	0.00058	0.00135

	bet Cl-Cr T1 base Coh bet Cr-Ol T1		
Baseline_eyes_open_global_coherence	Average of base Coh th Fl-Ol T1 base Coh th Fl-Or T1 base Coh th Fr-Ol T1 base Coh th Fr-Or T1 base Coh bet Fl-Ol T1 base Coh bet Fl-Or T1 base Coh bet Fr-Ol T1 base Coh bet Fr-Or T1	-0.00338	-0.00254
Baseline_eyes_open_global_coherence	Average of base Coh th Fr-Fr T1 base Coh alp Fr-Fr T1 base Coh bet Fr-Fr T1	-0.00183	-0.00125
Baseline_eyes_open_global_coherence	Average of base Coh th Ol-Or T1 base Coh alp Ol-Or T1 base Coh bet Ol-Or T1	0.008656	0.00677
Baseline_eyes_open_global_coherence	Average of base Coh alp Fl-Ol T1 base Coh alp Fr-Ol T1 base Coh alp Fr-Or T1	0.000636	5.65E-05
Baseline_Aperiodic_offset	Average of base Ap off Fl T2 base Ap off Fr T2 base Ap off Cl T2 base Ap off Cr T2 base Ap off Ol T2 base Ap off Or T2	0.014236	0.008498
Baseline_relative_power	Average of base Relative Pow th Fl T2 base Relative Pow th Fr T2 base Relative Pow th Cr T2 base Relative Pow th Fl T3 base Relative Pow th Fr T3 base Relative Pow th Cl T3 base Relative Pow th Cr T3 base Relative Pow th Ol T3 base Relative Pow th Or T3 base Relative Pow th Fl T4 base Relative Pow th Fr T4 base Relative Pow th Cl T4 base Relative Pow th Ol T4 base Relative Pow th Or T4	0.002633	5.33E-05
Baseline_relative_power	Average of base Relative Pow th Cl T2 base Relative Pow th Ol T2 base Relative Pow th Cr T4	0.0043	-0.00048

Baseline_relative_power	Average of base Relative Pow alp Fl T2 base Relative Pow alp Fr T2 base Relative Pow alp Cl T2 base Relative Pow alp Cr T2 base Relative Pow alp Ol T2 base Relative Pow alp Or T2 base Relative Pow alp Fl T3 base Relative Pow alp Fr T3 base Relative Pow alp Cl T3 base Relative Pow alp Cr T3 base Relative Pow alp Ol T3 base Relative Pow alp Or T3 base Relative Pow alp Fl T4 base Relative Pow alp Fr T4 base Relative Pow alp Cl T4 base Relative Pow alp Cr T4 base Relative Pow alp Ol T4 base Relative Pow alp Or T4 base alp asymmetry occipital T4	0.002251	0.00329
Baseline_relative_power	Average of base Relative Pow bet Fl T2 base Relative Pow bet Fr T2 base Relative Pow bet Cl T2 base Relative Pow bet Cr T2 base Relative Pow bet Ol T2 base Relative Pow bet Or T2 base Relative Pow bet Fl T3 base Relative Pow bet Fr T3 base Relative Pow bet Cl T3 base Relative Pow bet Cr T3 base Relative Pow bet Ol T3 base Relative Pow bet Or T3 base Relative Pow bet Fl T4 base Relative Pow bet Fr T4 base Relative Pow bet Cl T4 base Relative Pow bet Cr T4 base Relative Pow bet Ol T4 base Relative Pow bet Or T4	-0.00128	-0.00085
Baseline_Aperiodic_offset	Average of base Ap off Fl T3 base Ap off Fr T3 base Ap off Cl T3 base Ap off Cr T3 base Ap off Ol T3 base Ap off Or T3	0.005292	0.002566

Baseline_breathing_theta_periodic_power	Average of base Periodic Pow th Cl T3 base Periodic Pow th Cr T3 base Periodic Pow th Ol T3 base Periodic Pow th Or T3	0.0167	0.003866
Baseline_Aperiodic_offset	Average of base Ap off Fl T4 base Ap off Fr T4 base Ap off Cl T4 base Ap off Cr T4 base Ap off Ol T4 base Ap off Or T4	0.000801	-0.00033
Baseline_absolute_power	Average of base Absolute Pow th Fl T4 base Absolute Pow th Fr T4 base Absolute Pow th Cl T4 base Absolute Pow th Cr T4 base Absolute Pow th Ol T4 base Absolute Pow th Or T4 base Absolute Pow alp Fl T4 base Absolute Pow alp Fr T4 base Absolute Pow alp Cr T4 base Absolute Pow alp Ol T4 base Absolute Pow alp Or T4 base Absolute Pow bet Fl T4 base Absolute Pow bet Fr T4 base Absolute Pow bet Cl T4 base Absolute Pow bet Cr T4 base Absolute Pow bet Ol T4 base Absolute Pow bet Or T4	0.000671	0.000101
Baseline_eyes_open_global_coherence	base Coh th Cr-Cr T1	0.004501	0.001367
Baseline_eyes_open_global_coherence	base Coh alp Fl-Fr T1	0.003714	0.002794
Baseline_eyes_open_global_coherence	base Coh alp Cl-Cr T1	-0.00044	-0.00048
Baseline_eyes_open_global_coherence	base Coh alp Cl-Or T1	0.016104	0.012517
Baseline_eyes_open_global_coherence	base Coh alp Cr-Ol T1	-0.00167	-0.00089
-	base Entropy Ol T2	-0.00046	0.000148
Baseline_eyes_closed_central_left_coherence	base Coh th Cl-Or T2	0.003177	-0.00067

Baseline_eyes_closed_non_central_left_coherence	base Coh bet Fl-Fl T2	0.002368	-0.00271
-	base PLI Fl-Fr T2	0.005308	0.003767
-	base PLI Ol-Ol T2	0.003825	0.005516
Baseline_breathing_low_frequency_coherence	base Coh alp Fl-Cl T3	0.014864	0.013338
Baseline_breathing_low_frequency_coherence	base Coh alp Fr-Cr T3	0.003845	0.003262
Baseline_breathing_low_frequency_coherence	base Coh alp Cr-Ol T3	-0.0002	-0.00061
Baseline_breathing_central_tachy_PAC	base tachy PAC bet top Cr T3	0.004111	0.001112
Baseline_photic_central_left_coherence	base Coh th Fr-Cl T4	0.000875	0.001967
Baseline_photic_non_central_left_coherence	base Coh alp Fr-Fr T4	-0.00098	-3.34E-05
Baseline_photic_non_central_left_coherence	base Coh alp Fr-Cr T4	0.004286	0.001904
Baseline_photic_non_central_left_coherence	base Coh alp Cr-Ol T4	0.009384	0.002897
Baseline_photic_non_central_left_coherence	base Coh bet Or-Or T4	0.003627	-0.00329
Longitudinal_Aperiodic_offset	Average of Ap exp Fl T1 Ap off Fl T1	-0.00082	-0.00091
Longitudinal_Aperiodic_offset	Average of Ap off Ol T1 Ap off Or T1	-0.00061	-0.00093
Longitudinal_eyes_open_fronto-central_left_beta_periodic_power	Average of Periodic Pow bet Fl T1 Periodic Pow bet Cl T1	0.002073	0.001369
-	Average of Periodic Pow bet Cr T1 Periodic Pow bet Ol T1 Periodic Pow bet Or T1	0.003686	0.003711

Longitudinal_fronto-central_left_absolute_power	Average of Absolute Pow th Cl T1 th cordance Cl T1	0.00011	6.06E-05
Longitudinal_eyes_open_global_coherence	Average of Coh th Cr-Ol T1 Coh bet Cr-Ol T1	0.003978	0.001764
Longitudinal_eyes_open_global_coherence	Average of Coh th Ol-Or T1 Coh alp Ol-Or T1 Coh bet Ol-Or T1	-0.0003	0.001782
-	Average of Relative Pow bet Fl T2 Relative Pow bet Fr T2 Relative Pow bet Cl T2 Relative Pow bet Cr T2 Relative Pow bet Ol T2 Relative Pow bet Or T2 Relative Pow bet Fr T3 Relative Pow bet Cl T3 Relative Pow bet Cr T3 Relative Pow bet Ol T3 Relative Pow bet Or T3	0.000907	-0.00232
Longitudinal_eyes_closed_central_left_coherence	Average of Coh th Cl-Cl T2 Coh bet Cl-Cl T2	0.004102	-0.00147
Longitudinal_fronto-central_left_absolute_power	Average of Absolute Pow th Fl T3 Absolute Pow alp Fl T3 Absolute Pow alp Cl T3 Absolute Pow alp Cr T3 Absolute Pow alp Ol T3 Absolute Pow alp Or T3 Absolute Pow bet Fl T3 Absolute Pow bet Fr T3 Absolute Pow bet Cl T3 Absolute Pow bet Cr T3	0.004003	0.003358
Longitudinal_photic_non_central_left_coherence	Average of Coh th Fl-Fl T4 Coh bet Fl-Fl T4	0.006108	0.010807
Longitudinal_photic_non_central_left_coherence	Average of Coh th Fl-Cr T4 Coh th Cr-Or T4	0.005073	0.003881
Longitudinal_Aperiodic_offset	Ap exp Ol T1	0.010496	0.004676
Longitudinal_eyes_open_fronto_left_beta_relative_power	Relative Pow bet Fl T1	0.001331	-0.00235
Longitudinal_eyes_open_global_coherence	Coh bet Cr-Cr T1	0.004614	0.001228

-	th asymmetry occipital T1	-0.0014	-0.00107
Longitudinal_eyes_closed_central_left_coherence	Coh th Cl-Ol T2	0.010509	0.002536
Longitudinal_eyes_closed_non_central_left_coherence	Coh th Cr-Cr T2	0.002934	0.000754
Longitudinal_eyes_closed_non_central_left_coherence	Coh alp Fl-Fr T2	0.009293	0.006183
Longitudinal_eyes_closed_non_central_left_coherence	Coh alp Fl-Ol T2	0.002592	0.004229
Longitudinal_eyes_closed_non_central_left_coherence	Coh alp Fl-Or T2	0.001896	0.001337
Longitudinal_eyes_closed_non_central_left_coherence	Coh alp Fr-Or T2	0.016035	0.012844
Longitudinal_eyes_closed_central_left_coherence	Coh alp Cl-Cl T2	0.003974	-0.00133
Longitudinal_eyes_closed_non_central_left_coherence	Coh alp Or-Or T2	0.004425	-0.00252
Longitudinal_eyes_closed_central_left_coherence	Coh bet Cl-Ol T2	0.008804	-0.00046
Longitudinal_eyes_closed_non_central_left_coherence	Coh bet Cr-Cr T2	0.003715	-6.80E-06
Longitudinal_eyes_closed_non_central_left_coherence	Coh bet Cr-Ol T2	0.001588	-0.00275
-	PLI Fl-Or T2	0.002502	0.000891
-	PLI Fr-Fr T2	0.000541	-0.00061
Longitudinal_normo_PAC	normo PAC alp top Fl T2	0.002504	0.001151
Longitudinal_normo_PAC	normo PAC alp top Ol T2	0.000525	-0.0009
-	tachy PAC th top Fl T2	-4.27E-05	0.001114
-	tachy PAC alp top Fl T2	0.006698	0.006527
-	th asymmetry occipital T2	0.004749	0.001275

Longitudinal_low_frequency_central_asymmetry	alp asymmetry central T2	0.023277	0.019103
Longitudinal_high_frequency_frontal_asymmetry	beta asymmetry frontal T2	0.022208	0.009193
Longitudinal_breathing_fronto-central_left_theta_periodic_power	Periodic Pow th Cl T3	0.04386	0.029372
Longitudinal_breathing_fronto_left_beta_relative_power	Relative Pow bet Fl T3	0.003048	-0.00226
Longitudinal_breathing_low_frequency_coherence	Coh th Cr-Cr T3	0.000249	-0.00165
Longitudinal_breathing_low_frequency_coherence	Coh alp Fl-Cr T3	-0.00456	0.000362
Longitudinal_breathing_high_frequency_coherence	Coh bet Fr-Cr T3	0.001184	-0.00212
Longitudinal_breathing_high_frequency_coherence	Coh bet Fr-Ol T3	0.017992	0.012467
Longitudinal_normo_PAC	normo PAC brd top Or T3	0.001228	0.001127
Longitudinal_normo_PAC	normo PAC alp top Or T3	0.003027	0.006878
-	tachy PAC alp top Or T3	0.013614	0.012587
Longitudinal_breathing_central_tachy_PAC	tachy PAC bet top Cl T3	0.001239	0.000692
Longitudinal_breathing_central_tachy_PAC	tachy PAC bet top Cr T3	0.00763	0.005899
-	tachy PAC bet top Or T3	-0.00103	-0.00022
Longitudinal_high_frequency_frontal_asymmetry	beta asymmetry frontal T3	0.028463	0.011418
Longitudinal_photic_central_left_coherence	Coh th Cl-Cr T4	0.001768	2.00E-05
Longitudinal_photic_central_left_coherence	Coh alp Cl-Or T4	0.007817	0.004112
Longitudinal_photic_non_central_left_coherence	Coh alp Ol-Or T4	0.007304	0.002176

Longitudinal_photic_central_left_coherence	Coh bet Fl-Cl T4	0.00318	0.003703
Longitudinal_photic_central_left_coherence	Coh bet Fr-Cl T4	0.00918	0.007004
Longitudinal_photic_central_left_coherence	Coh bet Cl-Ol T4	0.000331	-0.00228
Longitudinal_photic_non_central_left_coherence	Coh bet Ol-Or T4	0.003007	0.001013
-	PLI Cl-Cl T4	-0.00325	2.08E-05
-	PLI Cr-Ol T4	-0.00039	0.000568
Longitudinal_low_frequency_central_asymmetry	th asymmetry central T4	0.008169	0.000334

Table A: List of selected features and their groups for feature importance. T1- eyes open, T2- eyes closed, T3- breathing, T4- photic administration tasks.

Feature importance

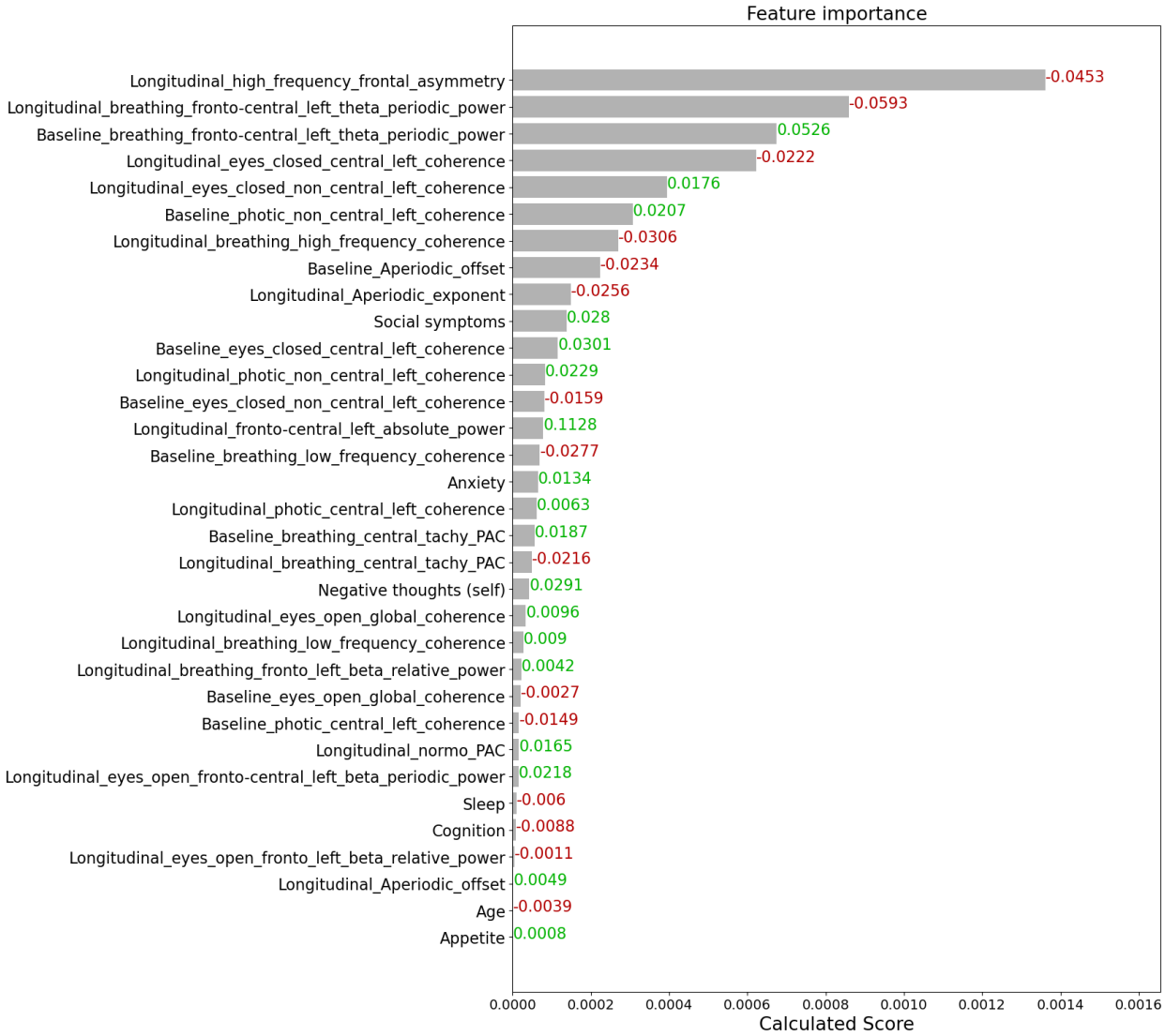


Figure V: Feature importance based on calculated score and grouping.

The slope depicts the effect of changing the feature value on shapley value. Red color depicts that increasing the feature value decreases the marginal contribution towards response and vice versa for green.

4) *Symptom wise electrophysiological feature*

Significant reduction in coupling of the tachygastric gut rhythm with the beta signals of central brain region, and also in overall beta coherence during breathing task, were seen in responders who had manifested higher severity than population median in appetite symptoms, negative thought about self, or social symptoms. Similarly, frontal beta asymmetry and periodic theta power in left fronto-central regions while breathing task showed significant reduction in responders with relatively higher appetite, sleep and social issues in the population. Responders with sleep issues displayed significant increase in coherence during photic administration task. Interestingly, responders with anxiety showed significant decrease in coherence in central left region, while it increased in the other regions during resting state eyes-closed task. Overall, we observed different electrophysiological biomarkers to be predictive of treatment outcome depending on their psychotype and physiological symptom presentation (**Table B**).

Symptom group	**Electrophysiology feature**	**mean_NC**	**mean_SC**	**corrected p_value**
Appetite	Baseline breathing left fronto-central theta periodic power	-0.160 (± 0.432)	0.700 (± 1.146)	0.034
	Baseline photic left non central coherence	-0.542 (± 0.294)	-0.104 (± 0.365)	<0.001
	Longitudinal breathing central tachy PAC	0.282 (± 0.596)	-0.176 (± 0.379)	0.042
	Longitudinal breathing beta coherence	0.192 (± 0.115)	-0.278 (± 0.328)	<0.001
	Longitudinal beta frontal asymmetry	0.259 (± 0.712)	-0.612 (± 1.121)	0.033
	Longitudinal breathing left fronto-central theta periodic power	0.235 (± 0.715)	-0.969 (± 0.823)	<0.001
Anxiety	Longitudinal eyes closed left non central coherence	-0.323 (± 0.330)	0.377 (± 0.776)	<0.001
	Longitudinal eyes closed left central coherence	0.261 (± 0.291)	0.063 (± 1.238)	0.041
Negative thoughts self	Longitudinal breathing central tachy PAC	0.356 (± 0.611)	-0.143 (± 0.395)	0.020
	Longitudinal eyes closed left central coherence	1.154 (± 1.946)	0.032 (± 1.259)	0.031
	Longitudinal breathing beta coherence	0.772 (± 0.793)	-0.054 (± 0.469)	<0.001
Sleep	Longitudinal photic non central beta coherence	-0.196 (± 0.461)	0.222 (± 0.449)	0.003

	Longitudinal photic central beta coherence	-0.070 (± 0.299)	0.248 (± 0.525)	0.040
	Longitudinal beta frontal asymmetry	0.105 (± 0.833)	-0.588 (± 1.037)	0.006
	Longitudinal breathing left fronto-central theta periodic power	0.078 (± 0.753)	-0.825 (± 0.748)	<0.001
Social symptoms	Longitudinal breathing central tachy PAC	0.291 (± 0.565)	-0.130 (± 0.390)	0.039
	Longitudinal eyes closed left central coherence	0.723 (± 1.763)	-0.106 (± 1.150)	0.024
	Longitudinal breathing beta frequency coherence	0.501 (± 0.559)	-0.039 (± 0.414)	<0.001
	Longitudinal beta frontal asymmetry	0.315 (± 0.512)	-0.480 (± 0.790)	<0.001
	Longitudinal breathing left fronto central theta periodic power	0.370 (± 0.910)	-0.753 (± 0.505)	<0.001

Table B: Significant features for specific biotypes- Electrophysiological feature comparison for responders and non-responders in a subset of population with specific symptom severity, with relatively higher symptom score than the population median. The above table highlights the feature group from the top 20 that are significant after Bonferonni correction for multiple comparisons, their corrected p-value, the mean feature value and N of SC and NC groups exhibiting significant change and no change in mental health for the specific subset of cohort.