Journal of Mood & Anxiety Disorders 11 (2025) 100126

Contents lists available at ScienceDirect

Journal of Mood & Anxiety Disorders

journal homepage: www.journals.elsevier.com/journal-of-mood-and-anxiety-disorders

Individualized functional brain mapping machine learning prediction of symptom-change resulting from selective kappa-opioid antagonism in an anhedonic sample from a Fast-Fail trial

Matthew D. Sacchet [a,b,*], Joseph L. Valenti [a], Poorvi Keshava [a], Shane W. Walsh [a], Moria J. Smoski [c], Andrew D. Krystal [d,1], Diego A. Pizzagalli [b,1]

[a] Meditation Research Program, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[b] Center for Depression, Anxiety and Stress Research, McLean Hospital, Harvard Medical School, Belmont, MA, USA
[c] Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA
[d] Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA

ARTICLE INFO

Keywords:
Individualized brain mapping
Machine learning
Selective kappa-opioid antagonism
Functional brain systems
Treatment prediction
Anhedonia
Transdiagnostic

ABSTRACT

Background: Anhedonia remains a difficult-to-treat symptom and has been associated with poor clinical course transdiagnostically. Here, we applied machine learning models to individualized neural patches derived from fMRI data during the Monetary Incentive Delay Task in anhedonic participants (N = 67) recruited for a clinical trial examining K-opioid receptor (KOR) antagonism in the treatment of anhedonia.
Methods: Nine ensemble models were estimated using cortical, subcortical, and combined cortical subcortical features from individualized functional topographies to predict changes in symptoms of overall psychopathology (anhedonia, depression, anxiety). Analyses were performed on the KOR (N = 33) and placebo (N = 34) group.
Results: Initial models showed that only subcortical data predicting depression and anxiety symptom change had a significant Spearman correlation between veridical and predicted data (*rho* = 0.480 and *rho* = 0.415 respectively). Next, leave-one-out-cross-validation (LOOCV) showed that the best-performing models comprised only the subcortical individualized systems data, which correlated with clinical change for depression and anxiety scores for the KOR group with significantly higher accuracy (*rho* = 0.634 and *rho* = 0.562, respectively) compared to the placebo group (*rho* = 0.294 and *rho* = 0.034, respectively). Further, 25 subcortical neural features were identified based on correlation and ensemble determined importance in driving prediction. Final models for both depression and anxiety showed an overall higher representation of the dorsal attention network. Cortical and combined cortical-subcortical feature data showed no significant improvement in prediction of clinical change between the two groups.
Conclusion: Using an ensemble of machine learning approaches, we identified individual differences in subcortical individualized systems data that predicted clinical change that was specific to KOR antagonism.

Introduction

Despite significant research advances, treatment studies for mental health disorders continue following a rudimentary one-size-fits-all model [1]. This partly originated from the prevalent system of classifying mental disorders as latent constructs, which has led to highly heterogenous clinical and neurobiological presentations of disorders [2], high rates of comorbidity [3] and poor treatment outcomes [4]. For instance, only 30–50 % of patients diagnosed with mood disorders respond favorably to medication [5–8]. Even drugs that show promise in early preclinical studies, do not seem to translate their success in human clinical samples [9,10].

To overcome this crisis the NIMH proposed the Fast-Fail initiative (https://www.nimh.nih.gov/research/research-funded-by-nimh/research-initiatives/fast-fail-trials-fast), which followed a *proof of mechanism* model (POM) that tested drugs on their intended neural treatment target to allow for rapid assessment of "target engagement". Here, we present secondary analyses from one of such study, the Mood and

* Correspondence to: 149 13th St., Charlestown, MA 02129-4522, USA.
E-mail address: meditationadministration@mgh.harvard.edu (M.D. Sacchet).
[1] Equal contribution.

https://doi.org/10.1016/j.xjmad.2025.100126
Received 6 December 2024; Received in revised form 29 April 2025; Accepted 29 April 2025
Available online 9 May 2025

Anxiety Spectrum Disorders (FAST-MAS) study. The goal of this trial was to test K-opioid receptor (KOR) antagonism in the treatment of anhedonia, and specifically whether it would engage a key node within the brain reward system (the nucleus accumbens) [11]. Anhedonia is a complex transdiagnostic phenomenon that encompasses loss of interest, pleasure and motivation to pursue or enjoy otherwise rewarding activities/experiences [12].

The current study leveraged machine learning approaches to characterize individualized neural patches derived from fMRI data obtained during reward anticipation in the Monetary Incentive Delay Task in a transdiagnostic population to predict treatment response to overall symptoms of psychopathology including, depression, anxiety and anhedonia. In line with the FAST-FAIL approach, the aim of this study was to use machine learning approaches to identify individualized functional brain organization patterns related to pharmacological compounds (specifically KOR antagonists) and their therapeutic effects using a data-driven approach. Traditionally, clinical trials take an extensive amount of time and resources to test the efficacy of drugs. Our study overcomes these limitations by proposing a statistical approach that aims to identify brain regions related to the effectiveness of drug compounds through limited human subject data. This approach promises to aid in facilitating outcome prediction and mechanistic pathways, thereby allowing for a more rapid assessment of therapeutic potential of new compounds. Machine learning methods provide valuable opportunities for psychiatric research by allowing researchers to identify robust features and predictive data patterns that correlate to mental health outcomes – such as drug response, symptoms, or behavior – and consider the utility in prediction of future responses for *individual* participants [5,13]. Advanced computational methods are especially useful in facilitating neuroscientific models in psychiatry which require systematic computations to identify distinct structures in brain function [14]. Moreover, machine learning techniques can help identify complex linear and non-linear relationships between multiple variables simultaneously [5]. Finally, ensemble methods present a key advantage in their ability to run prediction analysis with relatively smaller sample sizes [15–18]. This aligns closely with our objective of predicting drug compound outcomes in a resource-conscious manner, providing a more sustainable alternative to prolonged and costly clinical trials. This capability holds promise to both identify precise neural features and associate them with clinical variables.

This study also leveraged recent advances in neuropsychiatric precision-based techniques that highlight an individualized functional brain mapping approach to account for heterogeneity in individuals' brain architecture. Given prior discrepancies in relationships between cortical and subcortical relationships to symptoms of mood disorders [19,20], this study investigated cortical and subcortical regions in isolation and together, in regards to their relationship with symptoms. These approaches were based on previously developed individualized neural mapping algorithms that can essentially converge cortical and subcortical topologies specific to each individual [21–23]. Here, using an iterative optimization method, person-specific functional patches were identified within hypothesized brain regions. Here, 'patches' refer to nodes or distinct regions of the brain that exhibit functional homogeneity and highly correlated activity [24,25]. These patches were then used for subsequent analyses with the aim of developing more precise individualized treatment outcomes and predictions. Toward this aim, we used an ensemble model approach to identify higher performance and high robustness in assessing the dataset. We predicted individualized patch sizes would differentially predict response to KOR antagonism – such as control system patches (particularly in ventromedial and lateral prefrontal cortices and striatum) relating to degree of response to the KOR antagonist (JNJ-67953964). Thus, we predicted that machine learning would link individualized brain features in such areas to treatment outcomes, and that features noted above would contribute most to performance.

Methods and materials

Procedures

This study was conducted as a multi-site, 8-week, double-blind, placebo-controlled, randomized trial with individuals within a mood or anxiety disorder reporting some degree of anhedonia. JNJ-67953964 was administered to drug recipients in 10 mg doses daily. During baseline measurements and at the end of the 8-week period, MRI scans were completed. The original trial adhered to the International Conference on Harmonisation Good Clinical Practice guidelines and was approved by relevant institutional review boards. All participants or their legal representatives provided written informed consent [11]. Please see the published protocol for further details about the FAST-MAS trial [11].

Participants

From a total of 94 participants eligible for the original FAST-MAS trial, we included 67 participants who completed the 8-week treatment trial divided into the drug (N = 33) and placebo (N = 34) groups [11]. Data and clinical self-report measures were collected at the baseline visit ('pre' clinical scores) and clinical self-report was also collected at the end of the 8-week trial ('post' clinical scores) (see Supplemental Table 1).

Measures

Symptom measures

The Snaith Hamilton Pleasure Scale (SHAPS) is a 14-item self-assessment scale that assesses hedonic tone, with higher scores denoting more anhedonia [26].

The 17-item Hamilton Depression Rating Scale (HAM-D) [27] and the 14-item Hamilton Anxiety Rating Scale (HAM-A) [28] were used to assess the severity of depression and anxiety, respectively. A score of 20 or higher on the HAM-D is typically considered to indicate moderate severity of depression [29].

Symptom change for each clinical measure was defined as the difference between the 'post' and 'pre' trial scores.

MRI acquisition and preprocessing

MRI data acquisition and preprocessing was conducted using standard procedures (see Supplement; MRI acquisition and preprocessing for details).

Individualized functional topologies

Iterative parcellation approaches were used to arrive at individualized brain parcellation guided by group-level functional network atlases, flexibly adjusted based on interindividual variability and SNR distributions per subject. With each iteration, the influence from group-level information in determining individual maps was lessened until the final system map comprised individualized neural systems mapping [22]. These individually-derived cortical networks were separated into patches using algorithms and unsmoothed region patches matched to 116 cortical regions extracted from the group-level atlas as referenced [21]. Individual patches were labeled with ROIs based on overlapping vertices ($\geq$20) and nearest-neighbors approaches based on geodesic distance of neural surfaces (for full details, see [21]).

The subcortical individualized mapping, was based on the approach previously established by Greene and colleagues [23]. First the subcortical structures were segmented using FreeSurfer's 'recon-all' segmentation tool, using a downsampled template-derived mask. For each subcortical structure, individual voxel timecourses were first

adjusted to account for signal bleed from nearby cortical regions by regressing out the average timecourse of all ipsilateral cortical voxels within 20 mm of the given structure. Then each voxel of the subcortical regions were correlated using partial correlation coefficients to the average time course of each individualized cortical network, which were derived using methods mentioned above. The highest correlation value to a network was used to 'assign' a cortical network to the subcortical voxels, and these subcortical voxels were then summed to determine subcortical cluster size.

Machine learning: ensemble modeling

Analyses were performed using MATLAB. Analyses were performed using MATLAB. Data analysis was conducted using extensive in-house software, incorporating foundational code from https://github.com/MeilingAva/Homologous-Functional-Regions, using approaches from Wang et al. and Li et al. [21,22].

Feature definition

Nuisance variables (gender, age, site) were regressed from individualized subcortical clusters and cortical patch sizes across all FAST-MAS participants who completed the trial (33 drug and 34 placebo). Previous research has demonstrated the efficacy of ensemble statistical approaches with relatively small sample sizes, validating their application in limited-participant studies. Residualized data were then normalized to *Z*-scale to follow other machine learning methods and to aid in interpretation of residual data [30,31].

Model specification, feature selection, and performance testing

Features from subcortical, cortical, and combined subcortical/cortical datasets were used in conjunction with least squares ensemble models for predicting treatment outcomes [32]. This meant that three separate model types were generated. Choosing to create models with both datasets together and separate allowed for a better approach to understanding the relationships that drove prediction. If one dataset was able to better predict due to its relationship to treatment outcome, this would not be overshadowed by the other dataset worsening prediction. An ensemble model was used for its ability to perform strongly on data that are relatively small in sample sizes and have a higher number of features [33].

Feature selection was conducted by comparing feature data to treatment outcomes using Spearman correlations (as in [34–36]). At each 0.05 step between $p_t = 0.05$ and $p_t = 1.00$, features below the threshold (e.g., $p_t < .05$, $p_t < .10$, $p_t < .15 \ldots p_t < 1.00$), were included in the model, while features above the significance threshold were removed. Using a p_t-value other than 0.05 as determinants for feature selection was conducted given the high number of features and ensemble feature weight determination. This protected against including an arbitrary number of features within the model. Additionally, this method allowed for false discovery rate (FDR) correction on the obtained scores. FDR correction was conducted on all findings for 0.05-steps [37] between $p_t = 0.05$ and $p_{t\,=\,1.00}$, indicating that significant values obtained during cross-validation may perform well on testing data if introduced to the model.

Further, the least squares ensemble model was selected because it uses boosting to achieve improved performance, and has demonstrated strong results previously with neuroimaging data [38]. In an ensemble model, features that improved model accuracy were enhanced and features that lowered model accuracy were penalized in subsequent modeling steps. 100 iterations were conducted to determine the ultimate feature weights and finalize the model. Initial models were created using HAM-D, SHAPS, and HAM-A change-scores without any feature selection to obtain a null baseline comparison to compare against results with feature selection.

Leave-one-out-cross-validation (LOOCV) was conducted to complete feature selection (i.e., within each training dataset) as well as obtain a score of model performance, as assessed across folds against each left-out datapoint (Fig. 1). LOOCV was used to avoid leakage in the validation data affecting feature selection [39]. Once all individuals received a prediction, these scores were compiled, and Spearman correlation was used to determine accuracy between predicted and veridical treatment outcomes. High correlations suggest that the predictive machine learning approach was successful in using only brain imaging data to predict treatment outcome. Only significant positive correlations were analyzed, as the model was gauging the predictive capability of the model.

The model-produced-fit using this selected feature data from the KOR group was then used to test on the placebo group. That is, placebo data points were predicted upon within each LOOCV drug model. The determined predicted placebo scores were compared against the veridical placebo data, and for each fold of the LOOCV drug model, a Spearman correlation score between predicted placebo data and veridical placebo data was calculated. High correlations indicate the model found trends within fMRI data that were not unique to the KOR group and was able to predict well placebo data, while low correlations indicate any predictive results on KOR group were independent of the

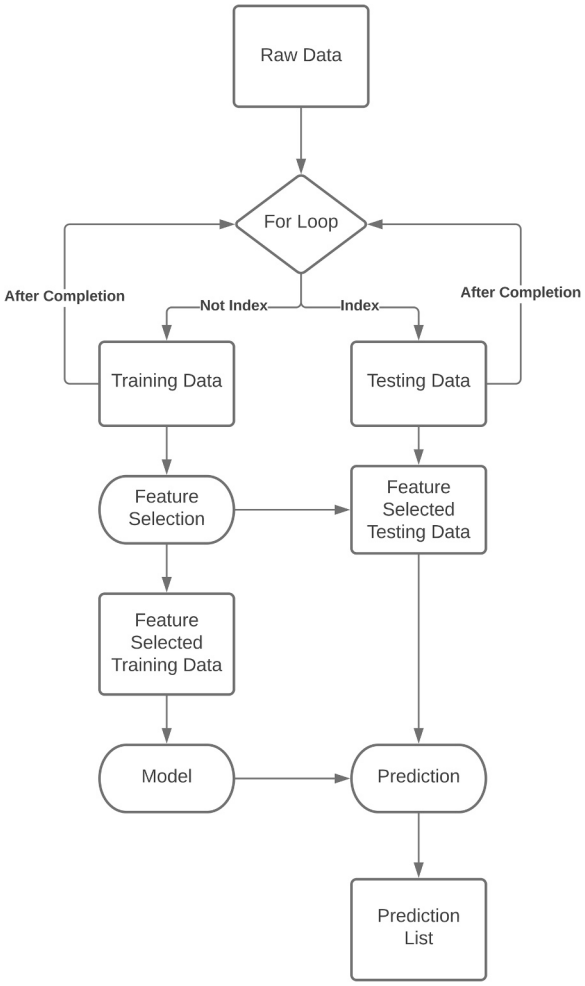


Fig. 1. Study methods overview. Within the drug group, leave one out cross validation (LOOCV) was implemented to determine overall change-scores. Each data point was set aside one at a time as a test point provided on whether the index was selected, while the rest of the model underwent feature selection to then be fit into an ensemble model to predict on the test point. This would then be placed into a prediction list to be used to calculate overall correlation scores and further testing. These data were then cycled through for testing on the next point until all points were evaluated on.

placebo data, relating towards the effectiveness of the drug.

After all folds of the LOOCV drug model were completed, the average of placebo Spearman correlation scores was then computed. This average was used as a prediction score for placebo data for each feature selection model, meaning that if this model was used on placebo data, this would be the outcome when tested on placebo subjects. If there was not a difference between placebo and drug subjects per model, then this implied that the model was predicting a general change effect that was not related to the drug. Testing on placebo data during each step of feature selection provided performance scores that could be used in comparison to drug recipient performance for each 0.05 threshold. This allowed for improvements of the model's performance to be analyzed so that it was only improving prediction on drug recipient subjects.

Statistical comparison of model predictive performance on KOR and placebo groups (i.e., Spearman correlations) was completed by comparing the predicted drug-group and the predicted placebo-group Spearman correlation scores using a Fishers *R-to-Z* test [36,40]. Significant differences in this statistic indicate that the model was only predicted on drug specific change during the trial.

Feature interpretation

Feature interpretation was only conducted if models for a given feature set (cortical, subcortical, or combined) exhibited significant performance after multiple comparisons correction using FDR across p_t-value threshold, and, furthermore, if drug prediction models also significantly performed higher than the placebo outcome prediction, as assessed by the Fishers *R*-to-*Z* test. The p_t-value threshold with the strongest predictive performance that met the qualifications above was used to identify features associated with this top performance.

Important features were first grouped by frequency of selection during feature selection. Using a method similar to LOOCV, the data from each participant was removed one at a time and the correlation threshold was applied on features within the dataset [34,35]. If a feature was present, then it was given a score of 1, otherwise it was given a score of 0. After the count of times a feature was present was tallied, each feature was given a score out of the number of individuals (Nk = 33) to determine the number of folds it was present in. The more times that a feature was present within the model, the higher impact it had on

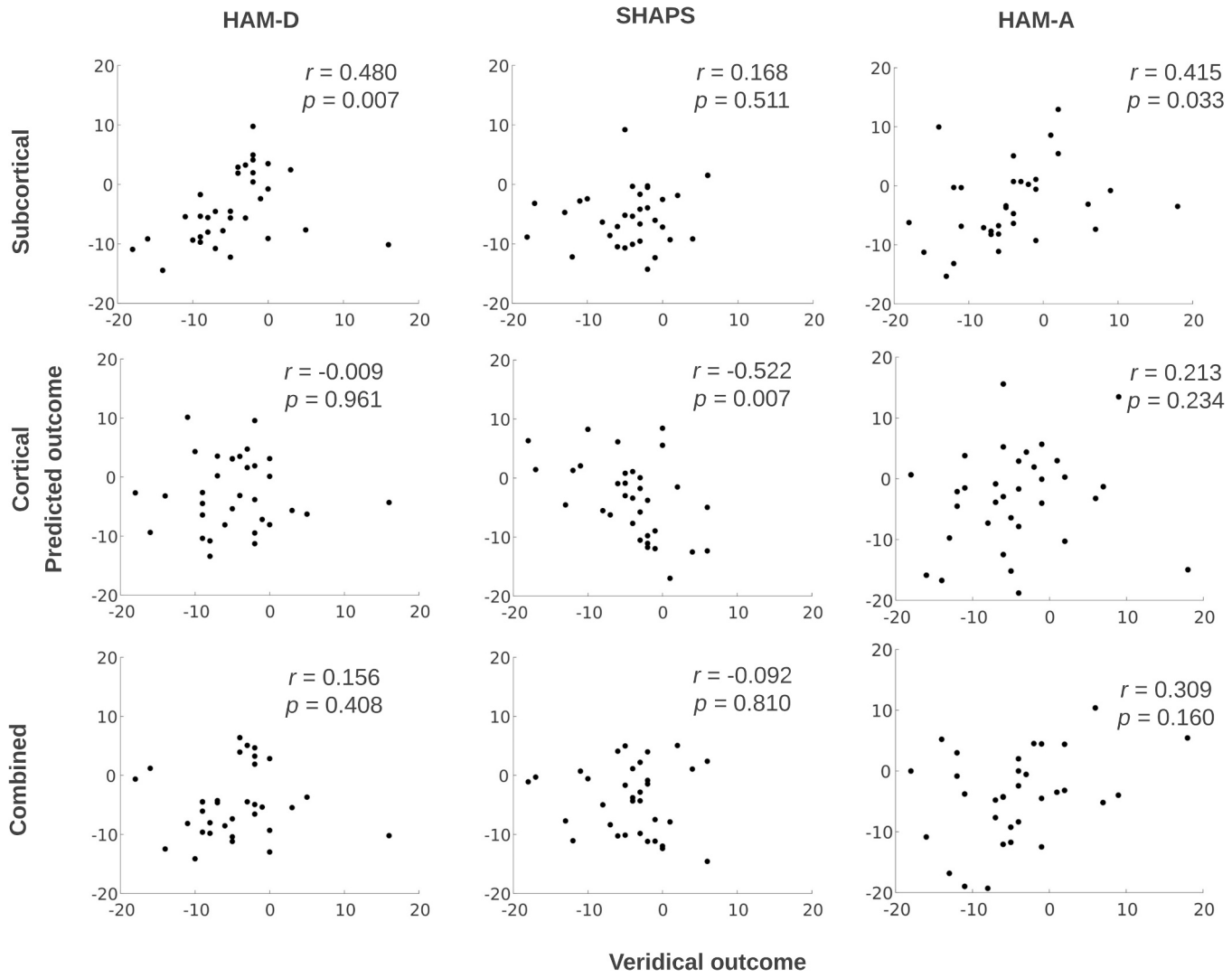


Fig. 2. Prediction of clinical outcomes without feature selection in the KOR group. The y-axis plots the data that used to train ensemble models organized by (i) subcortical only, (ii) cortical only, or (iii) both subcortical and cortical data combined. The x-axis signifies the change in scoring metric that was used for each target variable in the ensemble model, namely Hamilton Depression Rating Scale (HAM-D), Snaith–Hamilton Pleasure Scale (SHAPS) and Hamilton Anxiety Rating Scale (HAM-A). These scatterplots weigh the veridical (actual) outcome of the drug treatment against the predicted (scores generated through our ensemble models) outcome decided by the ensemble model. The scoring metric uses Spearman correlation to judge the similarity between the veridical outcome and the predicted outcome. Each graph is labeled with the spearman correlation score (*r*) and spearman significance value (*p*). The closer this value is to one, the stronger the relationship is.

predictive performance.

It was then decided to analyze the occurrence of features to see if they occurred within 70 % of LOOCV folds (see Supplement: Justification for LOOCV threshold).

After obtaining this information, features were sorted by importance determined by the ensemble model. Specifically, the top 25 most important features were sorted based on location and network affiliation within the brain. Counts of location and network affiliation were then statistically compared based on frequency using ANOVA. Next *post hoc* tests were used to determine if there were significant differences in frequency of occurrence. This analysis revealed which location and network affiliations are most relevant to the predictive modeling in terms of both correlational strength and feature importance.

Results

Assessment of ensemble model performance using subcortical individualized brain systems features to predict JNJ-67953964 treatment outcomes

Without feature selection, ensemble models were able to predict clinical outcomes using subcortical individualized system features (HAM-D $r = 0.480$, $p = 0.004$, FDR-corrected across p_t-value threshold = 0.007; SHAPS $r = 0.168$, $p = 0.360$, FDR-corrected across p_t-value threshold = 0.360; HAM-A $r = 0.415$, $p = 0.016$, FDR-corrected across p_t-value threshold = 0.0327). $r = 0.415$ respectively for the HAM-D, SHAPS, and HAM-A change-scores (Fig. 2).

Next, feature selection was used to improve model performance. During this feature selection, HAM-D change-scores showed the strongest results at the $p_t = 0.45$ threshold when using only subcortical

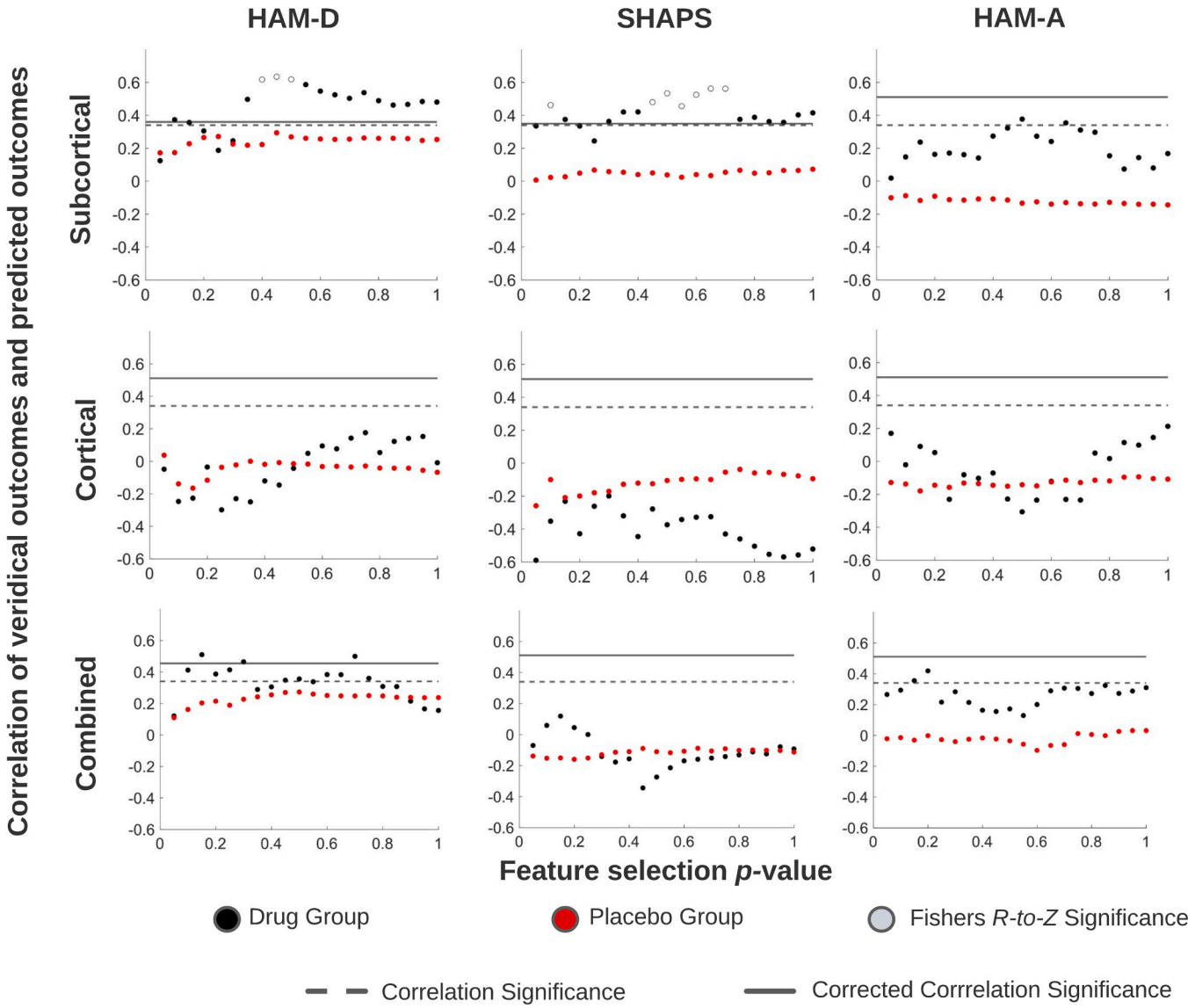


Fig. 3. Prediction of clinical outcomes using feature selection for KOR and placebo groups. The y-axis plots the data that were used to train ensemble models organized by (i) subcortical only, (ii) cortical only, or (iii) both subcortical and cortical data combined. The x-axis represents the change in scoring metric that was used for each target variable in the ensemble model, namely the Hamilton Depression Rating Scale (HAM-D), Snaith–Hamilton Pleasure Scale (SHAPS) and Hamilton Anxiety Rating Scale (HAM-A). Each plot shows the Spearman correlation value of veridical (actual) outcomes (x-axis) against the predicted (scores generated through our ensemble models) outcome for each p_t-value threshold (y-axis). The dashed line represents the uncorrected threshold for correlation significance while the thicker line represents the threshold for correlation significance after correction. The hollow points along the x-axis signify the veridical to predicted Spearman correlation values that were shown to be significantly higher than veridical to predicted Spearman correlation amongst placebo data using a Fisher *R*-to-*Z* test.

individualized systems features, with the strongest Spearman correlation of $r = 0.634$ (uncorrected: $p < 0.001$; corrected: $p = 0.001$). Thresholds between the p_t range of 0.35–1.00 (corrected and uncorrected: $p < 0.01$) as well as $p_t = 0.10$ (uncorrected: $p = 0.033$; corrected: $p = 0.043$) all displayed significant positive correlations after FDR correction (Supplemental Table 2, Fig. 3).

Prediction of SHAPS change-scores using subcortical only features was strongest at the p_t-value threshold of 0.50, with a Spearman correlation of $r = 0.377$ which was significant before (uncorrected: $p = 0.030$), but not after FDR correction (corrected: $p = 0.355$) (Supplemental Table 3, Fig. 3).

Prediction of HAM-A change-scores using subcortical only features was strongest at the p_t-value threshold of 0.65, with a Spearman correlation of $r = 0.562$ which maintained significance after FDR correction (uncorrected: $p < 0.001$; corrected: $p = 0.007$). Thresholds between the ranges of $p_t = 0.30$–1.00 as well as $p_t = 0.10$, 0.15 were significant after FDR correction (corrected and uncorrected: $p < 0.05$) (Supplemental Table 4, Fig. 3).

Subcortical drug ensemble accuracy

Using only subcortical individualized systems data, HAM-D change-score drug prediction performance in the range $p_t = 0.40$–0.50 were significantly more accurate than placebo performance using Fishers *R*-to-*Z* analysis (Supplemental Table 2, Fig. 3). SHAPS drug prediction performance, while not significantly accurate, was significantly more accurate for the KOR compared to placebo outcome predictions at p_t-thresholds of 0.45, 0.50, 0.65, 0.70, and 0.75 (Supplemental Table 3, Fig. 3). HAM-A drug prediction performance in the range $p_t = 0.45$–0.70 as well as $p_t = 0.10$ were significantly more accurate than placebo group prediction scores (Supplemental Table 4, Fig. 3).

Cortical drug ensemble accuracy

Without feature selection, models trained on cortical individualized systems data did not significantly predict any symptom change (see Supplement: Cortical Drug Ensemble Accuracy for details).

Combined subcortical and cortical drug ensemble accuracy

Without feature selection, models trained on combined subcortical and cortical individualized systems data did not significantly predict any symptom change (see Supplement: Combined Subcortical and Cortical Drug Ensemble Accuracy for details).

Feature normalization

All results were repeated under the conditions of using normalized or nonnormalized data during data preprocessing where covariate variables were regressed out and residuals were created and fit in the model. These analytic decisions had no effect on the results.

Subcortical feature abstraction

Since the strongest results were observed at subcortical HAM-D at $p_t = 0.45$ and subcortical HAM-A at $p_t = 0.65$, features were further analyzed at these two threshold levels (Supplemental Table 2, Supplemental Table 4, Fig. 4).

The dorsal attention area within the subcortical HAM-A feature set was the only region that showed significantly greater occurrence compared to the other top 25 neural features ($p = 0.027$) (see Supplement: Subcortical Feature Abstraction for details).

Discussion

The subcortical features dataset proved successful with feature selection and performed well with the training data, while the results from the cortical and cortical and subcortical combined data were not as strong. This is not surprising as depression is a highly complex pathology characterized by alterations in various subcortical regions [41]. For instance, anhedonia, a core symptom of depression [42] is implicated by dysfunctions in the mesolimbic reward circuit which primarily involves connectivity between prefrontal cortex (PFC) and various subcortical structures [20,43–45]. Further studies examining the role of the subcortex in MDD have found atrophy [46,47] and abnormal volumes in subcortical structures which are associated with illness duration and antidepressant treatment outcomes [47]. Therefore, our results showing improved predictive capacity of the subcortical dataset are in line with prior work and suggest that subcortical features may play a substantial role in treatment prediction. However, while these subcortical regions appear to predict treatment outcomes, it is unclear whether these regions are all directly targeted by KOR antagonism. Significant correlations between HAM-D and HAM-A veridical scores and predicted scores demonstrate the successful link between machine learning biomarker outcomes and clinical utility. This connection could potentially lead to promising biomarker-based assessments of depression and anxiety

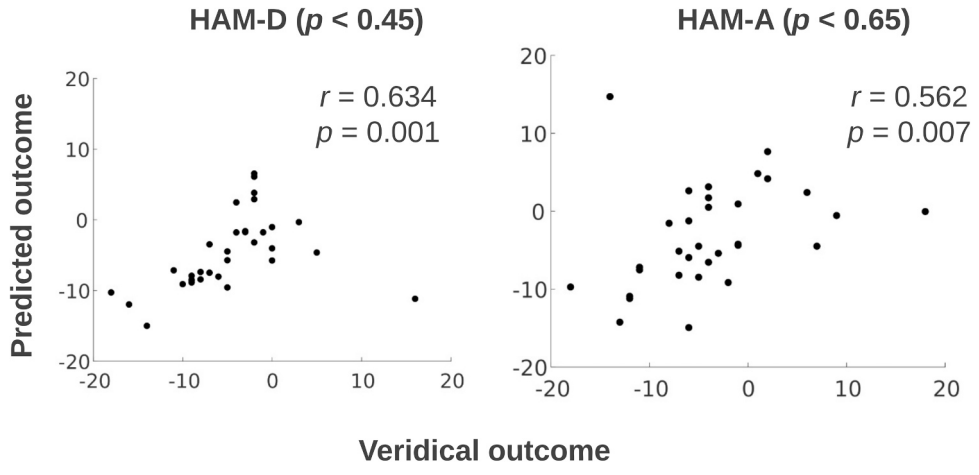


Fig. 4. Highest scoring features and clinical symptom-change in KOR group. From the previous analyses, the highest scoring feature selected models were identified and plotted. Hamilton Depression Rating Scale (HAM-D) (left) and Hamilton Anxiety Rating Scale (HAM-A) (right) outcomes are plotted using subcortical data at the $p = 0.45$ and $p = 0.65$, respectively. Veridical (actual) outcomes are plotted on the x-axis against the predicted (scores generated through our ensemble models) outcome data on the y-axis.

symptoms in clinical populations, and aid in developing efficient biologically meaningful treatment targets in early stages of drug development, in line with the initial goals of the Fast-Fail trial.

The initial subcortical and HAM-D and HAM-A models showed significant correlations between veridical (actual) and predicted (scores generated by our ensemble models) data. Results were strengthened by using various thresholds of correlation-based feature selection. The strongest resulting levels of significance for feature selection were $p_t = 0.45$ and $p_t = 0.65$ for HAM-D and HAM-A models, respectively. However, our model did not predict significant symptom change in anhedonia. However, given that anhedonia is a complex condition involving various mechanisms such as deficits in neurotransmitters [48], altered functional connectivity [49] and anatomical abnormalities [50]; it is possible that our analysis that *solely* relied on functional patches was insufficient to predict changes in this symptom. Additionally, another possible reason why the models did not predict changes in anhedonia scores could be due to the self-report nature of SHAPS. The SHAPS conceptualizes anhedonia as a relatively stable "trait-like" construct, and primarily captures anticipatory or remembered pleasure rather than real-time fluctuations in reward processes. Consequently, our method may not have been sufficiently sensitive to detect changes in self-report based on individualized neural biomarkers. This challenge may also arise from the complexity of neural reward processing, wherein distinct neural systems underlie distinct aspects of the reward process, such as motivation, anticipation, and consummation, which could influence the ability to accurately model and predict changes in self-reported anhedonia over time. This methodology included both linear and nonlinear relationships through the use of correlation and the ensemble model in order to best predict change scores. The improvement based on feature selection and model usage demonstrates the complexity of individualized neural approaches and the need for approaches that explore this complexity. Future research could benefit from expanding on current approaches to feature selection in order to encapsulate both linear and nonlinear patterns within data, either with a thresholding type approach or other methodologies. Expanding the complexity at which data are treated allowed for stronger prediction and increasing the likelihood of a tool that could be used more efficiently to predict treatment success. These results were tested against placebo data, which showed that there were significantly better results of prediction among the $p_t = 0.45$ and $p_t = 0.65$ for HAM-D and HAM-A models. This was crucial in showing that the results were specific to treated individuals. The specificity of these findings to treated individuals, further supports these models as a tool for treatment.

When developing models with cortical feature data and combined cortical-subcortical feature data, there were no significantly improved predictions of clinical change scores between groups. While this does not indicate that cortical regions of the brain are not involved within interactions with KOR, it merely suggests that their activity may not have had a strong enough relationship with change scores to be used as biomarker within these ensemble models.

Given that HAM-A and HAM-D models were significantly correlated to predicted and veridical scores, additional analyses of neural features within these models were conducted. While most feature occurrences did not vary significantly between the models for the various clinical change score predictors, one network was significantly greater in occurrence than others for HAM-A score prediction, specifically the dorsal attention areas. While dorsal regions have been previously identified as a feature discriminating prediction of depressed versus bipolar in support vector machine (SVM) classifiers trained on multimodal MRI data [51], our results pointed towards an implication of the dorsal attention region in the treatment response for anxiety. However, since the study by Jie and colleagues did not exclude for anxiety disorders [51], and given the common co-morbidity of anxiety and depression [52], it may be possible that dorsal attention regions have meaningful implications for anxiety and specifically co-morbid depression and anxiety as opposed to depression alone. The dorsal attention region is primarily involved in top down regulation via allocation of attention to stimuli [53]. A prior study found increased functional connectivity between the dorsal attention network (DAN) and the amygdala in individuals with high trait anxiety [54]. Given amygdala's primary role in processing threat stimuli [55,56], higher connectivity between the amygdala and DAN have suggested an increased attentional bias towards threat related stimuli [54,57], which is well established underlying mechanism of anxiety. Further, trait anxiety is also considered to be a risk factor for depression [58]. Interestingly, in line with our results, individuals co-morbid anxiety and depression show higher threat bias than individuals with depression alone [59]. Other than the dorsal attention areas having significance over other features within the subcortical HAM-A model, no other regions appeared to differ in frequency and makeup at a significant level for the HAM-D or HAM-A features. These results demonstrate the significance of investigating individualized neural systems holistically without the biases of linear group-level influences which may lead to oversimplified conclusions by not accounting for individual variability in neural functional organization. This approach may allow for more likelihood of trial success as it ensures that meaningful systems which may have non-linear relationships to these systems are not overlooked in predicting treatment outcomes.

Despite considerable strengths, we highlight several limitations. While our study did include training and validation data, we did not have a separate hold-out or testing dataset. This was due to a relatively small sample size for such machine learning performance designs and the fact that only one RCT has been conducted so far using a KOR antagonist. We addressed this important limitation by assessing *p*-values and FDR corrections to provide a benchmark of statistical strength. Alternate methods of choosing ideal features in machine learning models could have been explored, such as SelectKBest, however the thresholding approach used in the present study was selected for its ability to filter out arbitrary feature numbers at varying levels. Further, use of LOOCV for cross validation allowed greater amounts of data within the small sample size to be allocated to training rather than testing, but is ultimately associated with high variance and weaker generalization performance [60]. Finally, participant data indicated mild to moderate levels of depression, anxiety, and anhedonia at baseline, which may limit the generalizability of the findings to populations with more severe psychopathology or to those without any psychiatric symptoms.

Future research could include stringent testing-training splits instead of LOOCV approaches to evaluate whether ensemble models may accurately predict similar feature selection outputs. Additionally, while our relatively smaller sample size was adequate for analyses based on previous studies, future research would benefit from validating these findings using a large sample. Future research would also benefit from examining whether our findings replicate in samples with more severe psychopathology. Expanding the range of symptoms severity will help clarify the extent to which these findings generalize across more severe psychopathology and could have meaningful treatment implications. Overall, this study aimed to use individualized data in conjunction with machine learning to further explore the effectiveness of a data-driven approach to identify neural features of importance. The resultant models ran successfully, and in the case of subcortical feature selection, appeared to extract neural features related to affective processing in prediction of drug, and notably *not* placebo responses in depressed and anhedonic participants. The success of these models demonstrates the value of machine learning-based methodologies in creating biomarker-based tools for predictive treatment. Data-driven approaches such as this allow for more multi-faceted and complex relationships between individualized neural markers and treatment prediction may yield more conservative, and notably more accurate treatment targets compared to traditional group-aggregate linear correlational approaches.

Author contributions

MDS and JLV designed the research; MDS, JLV, SWW, contributed new reagents/analytic tools; JLV, SWW analyzed the data; MDS, PK, JLV, DAP drafted the manuscript; MDS, JLV, PK, SWW, MJS, ADK, DAP, contributed to interpretation of data and revisions of the manuscript.

Financial disclosures

Over the past 3 years, Dr. Pizzagalli has received consulting fees from Albright Stonebridge Group, Boehringer Ingelheim, Compass Pathways, Engrail Therapeutics, Neumora Therapeutics (formerly BlackThorn Therapeutics), Neurocrine Biosciences, Neuroscience Software, Otsuka Pharmaceutical, Sage Therapeutics, Sunovion Pharmaceuticals, and Takeda; he has received honoraria from the Psychonomic Society and American Psychological Association (for editorial work) and from Alkermes; he has received research funding from the Brain and Behavior Research Foundation, the Dana Foundation, Millennium Pharmaceuticals, Wellcome Leap MCPsych, and NIMH; he has received stock options from Compass Pathways, Engrail Therapeutics, Neumora Therapeutics, and Neuroscience Software. No funding from these entities was used to support the current work, and all views expressed are solely those of the authors.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Matthew D. Sacchet reports financial support was provided by National Institute of Mental Health. Matthew D. Sacchet reports financial support was provided by Dimension Giving Fund. Matthew D. Sacchet reports financial support was provided by Ad Astra Chandaria Foundation. Matthew D. Sacchet reports financial support was provided by Brain and Behavior Research Foundation. Matthew D. Sacchet reports financial support was provided by BIAL Foundation. Diego A. Pizzagalli reports financial support was provided by National Institute of Mental Health. Diego A. Pizzagalli reports a relationship with Albright Stonebridge Group that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Boehringer Ingelheim that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Compass Pathways that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Engrail Therapeutics that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Neumora Therapeutics that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Neurocrine Biosciences Inc that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Neuroscience Software that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Otsuka pharmaceutical that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Sage Therapeutics that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Sunovion Pharmaceuticals that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Takeda that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Psychonomic Society Inc that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with American Psychological Association that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Alkermes that includes: consulting or advisory. Diego A. Pizzagalli reports a relationship with Brain and Behavior Research Foundation that includes: funding grants. Diego A. Pizzagalli reports a relationship with Dana Foundation that includes: funding grants. Diego A. Pizzagalli reports a relationship with Millennium Pharmaceuticals that includes: funding grants. Diego A. Pizzagalli reports a relationship with Wellcome Leap MCPsych that includes: funding grants. Diego A. Pizzagalli reports a relationship with Compass Pathways that includes: equity or stocks. Diego A. Pizzagalli reports a relationship with Engrail Therapeutics that includes: equity or stocks. Diego A. Pizzagalli reports a relationship with Neumora Therapeutics that includes: equity or stocks. Diego A. Pizzagalli reports a relationship with Neuroscience Software that includes: equity or stocks. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

DAP was partially supported by R37 MH068376. Dr. Sacchet and the Meditation Research Program are supported by the National Institute of Mental Health (Project Number R01MH125850), Dimension Giving Fund, Ad Astra Chandaria Foundation, Brain and Behavior Research Foundation (Grant Number 28972), BIAL Foundation (Grant Number 099/2020), and individual donors.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.xjmad.2025.100126.

References

[1] Purgato M, Singh R, Acarturk C, Cuijpers P. Moving beyond a 'one-size-fits-all' rationale in global mental health: prospects of a precision psychology paradigm. Epidemiol Psychiatr Sci 2021;30:e63.
[2] Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The heterogeneity problem: approaches to identify psychiatric subtypes. Trends Cogn Sci 2019;23:584–601.
[3] Cramer AOJ, Waldorp LJ, Maas HLJ van der, Borsboom D. Comorbidity: a network perspective. Behav Brain Sci 2010;33:137–50.
[4] Newson JJ, Pastukh V, Thiagarajan TC. Poor separation of clinical symptom profiles by DSM-5 disorder criteria. Front Psychiatry 2021;12.
[5] Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 2021;20:154–70.
[6] Nemeroff CB, Owens MJ. Treatment of mood disorders. Nat Neurosci 2002;5:1068–70.
[7] Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006;163:28–40.
[8] Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006;163:1905–17.
[9] Kesselheim AS, Hwang TJ, Franklin JM. Two decades of new drug development for central nervous system disorders. Nat Rev Drug Discov 2015;14:815–6.
[10] Paul SM, Potter WZ. Finding new and better treatments for psychiatric disorders. Neuropsychopharmacology 2023:1–7.
[11] Krystal AD, Pizzagalli DA, Smoski M, Mathew SJ, Nurnberger J, Lisanby SH, et al. A randomized proof-of-mechanism trial applying the 'fast-fail' approach to evaluating κ-opioid antagonism as a treatment for anhedonia. Nat Med 2020;26:760–8.
[12] Pizzagalli DA. Depression, stress, and anhedonia: toward a synthesis and integrated model. Annu Rev Clin Psychol 2014;10:393–423.
[13] Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. JAMA Psychiatry 2017;74:370–8.
[14] Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 2016;19:404–13.
[15] Chung K-H, Chang Y-S, Yen W-T, Lin L, Abimannan S. Depression assessment using integrated multi-featured EEG bands deep neural network models: leveraging ensemble learning techniques. Comput Struct Biotechnol J 2024;23:1450–68.
[16] Jo YT, Joo SW, Shon S-H, Kim H, Kim Y, Lee J. Diagnosing schizophrenia with network analysis and a machine learning method. Int J Methods Psychiatr Res 2020;29:e1818.
[17] Paul R, Tsuei T, Cho K, Belden A, Milanini B, Bolzenius J, et al. Ensemble machine learning classification of daily living abilities among older people with HIV. eClinicalMedicine 2021;35:100845.
[18] Rezaei M, Zare H, Hakimdavoodi H, Nasseri S, Hebrani P. Classification of drug-naive children with attention-deficit/hyperactivity disorder from typical development controls using resting-state fMRI and graph theoretical approach. Front Hum Neurosci 2022;16:948706.
[19] Merz EC, He X, Noble KG. Anxiety, depression, impulsivity, and brain structure in children and adolescents. NeuroImage Clin 2018;20:243–51.
[20] Heller AS. Cortical-subcortical interactions in depression: from animal models to human psychopathology. Front Syst Neurosci 2016;10:20.

[21] Li M, Wang D, Ren J, Langs G, Stoecklein S, Brennan BP, et al. Performing group-level functional image analyses based on homologous functional regions mapped in individuals. PLOS Biol 2019;17:e2007032.
[22] Wang D, Buckner RL, Fox MD, Holt DJ, Holmes AJ, Stoecklein S, et al. Parcellating cortical functional networks in individuals. Nat Neurosci 2015;18:1853–60.
[23] Greene DJ, Marek S, Gordon EM, Siegel JS, Gratton C, Laumann TO, et al. Integrative and network-specific connectivity of the basal ganglia and thalamus defined in individuals. Neuron 2020;105:742–758.e6.
[24] Wang D, Li M, Wang M, Schoeppe F, Ren J, Chen H, et al. Individual-specific functional connectivity markers track dimensional and categorical features of psychotic illness. Mol Psychiatry 2020;25:2119–29.
[25] Bijsterbosch JD, Woolrich MW, Glasser MF, Robinson EC, Beckmann CF, Van Essen DC, et al. The relationship between spatial configuration and functional connectivity of brain regions. eLife 2018;7:e32992.
[26] Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P. A scale for the assessment of hedonic tone the snaith–hamilton pleasure scale. Br J Psychiatry 1995;167:99–103.
[27] Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6:278–96.
[28] Hamilton M. THE assessment of anxiety states by rating. Br J Med Psychol 1959;32:50–5.
[29] Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62.
[30] Plotkin D. Summing it all up. Data steward. Elsevier; 2014. p. 183–209.
[31] Singh D, Singh B. Investigating the impact of data normalization on classification performance. Appl Soft Comput 2020;97:105524.
[32] Kuncheva LI, Rodríguez JJ. Classifier ensembles for fMRI data analysis: an experiment. Magn Reson Imaging 2010;28:583–93.
[33] Liu M, Zhang D, Shen D. Ensemble sparse classification of Alzheimer's disease. NeuroImage 2012;60:1106–16.
[34] Lebois LAM, Li M, Baker JT, Wolff JD, Wang D, Lambros AM, et al. Large-scale functional brain network architecture changes associated with trauma-related dissociation. Am J Psychiatry 2021;178:165–73.
[35] Senan EM, Abunadi I, Jadhav ME, Fati SM. Score and correlation coefficient-based feature selection for predicting heart failure diagnosis by using machine learning algorithms. Comput Math Methods Med 2021;2021:1–16.
[36] Sheshkin DJ. Handbook of parametric and nonparametric statistical procedures. 3rd ed. Boca Raton FL: Chapman & Hall/CRC; 2004.
[37] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat 1995;57.
[38] Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. NeuroImage 2011;55:1519–27.
[39] Kaufman S, Rosset S, Perlich C. Leakage in data mining: formulation, detection, and avoidance. Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining - KDD 11. San Diego, California, USA: ACM Press; 2011. p. 556.
[40] Zar JH. Biostatistical analysis. 4th ed. Prentice Hall: Upper Saddle River NJ; 1999.
[41] Schmaal L, Veltman DJ, van Erp TGM, Sämann PG, Frodl T, Jahanshad N, et al. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA major depressive disorder working group. Mol Psychiatry 2016;21:806–12.
[42] Admon R, Pizzagalli DA. Dysfunctional reward processing in depression. Curr Opin Psychol 2015;4:114–8.
[43] Tzschentke TM. The medial prefrontal cortex as a part of the brain reward system. Amino Acids 2000;19:211–9.
[44] Chau BKH, Jarvis H, Law C-K, Chong TT-J. Dopamine and reward: a view from the prefrontal cortex. Behav Pharmacol 2018;29:569–83.
[45] Haber SN, Knutson B. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacol Publ Am Coll Neuropsychopharmacol 2010;35:4–26.
[46] Price JL, Drevets WC. Neurocircuitry of mood disorders. Neuropsychopharmacology 2010;35:192–216.
[47] Zhou Y-L, Wu F-C, Liu W-J, Zheng W, Wang C-Y, Zhan Y-N, et al. Volumetric changes in subcortical structures following repeated ketamine treatment in patients with major depressive disorder: a longitudinal analysis. Transl Psychiatry 2020;10:264.
[48] Gorwood P. Neurobiological mechanisms of anhedonia. Dialog Clin Neurosci 2008;10:291–9.
[49] Hu Y, Zhao C, Zhao H, Qiao J. Abnormal functional connectivity of the nucleus accumbens subregions mediates the association between anhedonia and major depressive disorder. BMC Psychiatry 2023;23:282.
[50] Auerbach RP, Pisoni A, Bondy E, Kumar P, Stewart JG, Yendiki A, et al. Neuroanatomical prediction of anhedonia in adolescents. Neuropsychopharmacology 2017;42:2087–95.
[51] Jie N-F, Zhu M-H, Ma X-Y, Osuch EA, Wammes M, Théberge J, et al. Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. IEEE Trans Auton Ment Dev 2015;7:320–31.
[52] Hirschfeld RMA. The comorbidity of major depression and anxiety disorders: recognition and management in primary care. Prim Care Companion J Clin Psychiatry 2001;3:244–54.
[53] Koziol LF, Barker LA, Joyce AW, Hrin S. Structure and function of large-scale brain systems. Appl Neuropsychol Child 2014;3:236–44.
[54] He Y, Xu T, Zhang W, Zuo X. Lifespan anxiety is reflected in human amygdala cortical connectivity. Hum Brain Mapp 2015;37:1178–93.
[55] Öhman A. The role of the amygdala in human fear: automatic detection of threat. Psychoneuroendocrinology 2005;30:953–8.
[56] Britton JC, Lissek S, Grillon C, Norcross MA, Pine DS. Development of anxiety: the role of threat appraisal and fear learning. Depress Anxiety 2011;28:5–17.
[57] Mohanty A, Egner T, Monti JM, Mesulam M-M. Search for a threatening target triggers limbic guidance of spatial attention. J Neurosci 2009;29:10563–72.
[58] Wang T, Li M, Xu S, Jiang C, Gao D, Wu T, et al. The factorial structure of trait anxiety and its mediating effect between mindfulness and depression. Front Psychiatry 2018;9.
[59] Ironside M, DeVille DC, Kuplicki RT, Burrows KP, Smith R, Teed AR, et al. The unique face of comorbid anxiety and depression: increased interoceptive fearfulness and reactivity. Front Behav Neurosci 2023;16.
[60] Gao S, Calhoun VD, Sui J. Machine learning in major depression: from classification to treatment outcome prediction. CNS Neurosci Ther 2018;24:1037–52.

Journal of Mood & Anxiety Disorders

Supporting Information for: *Individualized functional brain mapping machine learning prediction of symptom-change resulting from selective kappa-opioid antagonism in an anhedonic sample from a FAST-FAIL trial*

*Matthew D. Sacchet, Joseph L. Valenti, Poorvi Keshava, Shane W. Walsh, Moria J. Smoski, §Andrew D. Krystal, §Diego A. Pizzagalli

*Corresponding author: Matthew Sacchet; 149 13th St., Charlestown, MA 02129-4522; Phone: 617-643-601; Email: sacchetadmin@mgh.harvard.edu.

§Andrew D. Krystal and Diego A. Pizzagalli are co-last authors in this paper.

This PDF file includes:

Supplementary Methods

Participants:

163 participants were screened for the original FAST-MAS trial. From the total screened, 94 were eligible to participate in the study determined by qualifying scores (see next section, Measures). Five participants dropped out of the trial before baseline measurements were acquired. As a result, 89 participants completed baseline measurements. Of this sample, participants were randomized to the KOR (N=45) or the placebo (N=44) group. All participants had a SHAPS score ≥ 20.

From the screened sample, a total of 67 participants completed the 8-week treatment trial and had fMRI data that could be used in the current analyses. This resulted in roughly equivalent group sizes for drug (n=33) and placebo (n=34) groups [1].

MRI Acquisition and Preprocessing

fMRI data were collected during resting-state and the *Monetary Incentive Delay* (MID) task. Resting-state and contrasts of interest used for the MID task in the original study were all aggregated to produce continuous runs to be analyzed, aligning with previous work using individualized systems approaches [2–5].

fMRI data were collected with Gradient-echo echo-planer axial scans using the following parameters: TR/TE: 2000/30ms, flip angle: 70 deg, FOV: 25.6 cm, matrix: 64x64, 32 axial slices, acceleration factor=2, voxel size: 4x4x4 mm.

fMRI data processing was completed using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Registration to high resolution structural and standard space images was conducted with FLIRT [6, 7] with registration from high resolution structural to standard space with FNIRT nonlinear registration [8].

Further preprocessing included motion correction using MCFLIRT [6, 7], slice-timing correction using Fourier-space time-series phase-shifting, non-brain removal using BET spatial smoothing with a FWHM 5mm Gaussian kernel, grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma=45.0s) [7, 9].

Feature Selection Comparison

As a final analysis with pt-value threshold feature selection coupled with the ensemble model feature weight abstraction, a model was generated with the top 25 Spearman-correlated neural features. This was done to determine whether the method of feature selection used, where various steps of correlational significance was used as a filter, would perform better than just using the flat number of significant features, as other papers have used (J. Liu et al. 2021). Twenty-five was chosen as the number of features used to mirror the number of features analyzed during feature interpretation. The models created using the top 25 Spearman-correlated features used LOOCV to create a comparable validation score as the previous feature selection method. The validation scores were compared to the models that performed most highly during feature selection, similar to the metrics that were used to determine what models were analyzed during feature interpretation. This comparison would then have a Fishers R-to-Z test applied to assess if there was a significantly better Spearman correlation determined based on the method of feature selection.

Justification for LOOCV threshold

This threshold was selected for high qualitative occurrence, such that lower presence typically suggested a steep drop among remaining features (Supplemental Fig. 1). All features meeting this threshold were analyzed in an ensemble model in order to abstract the feature weights determined during training. These two steps allowed the extraction of information related to the strengths of the linear and nonlinear relationships of the features and outcomes, assessing how important they are to outcome prediction and model performance.

Supplementary Results

Feature Selection Comparison

Since the best performing models during feature selection were the p_t<0.45 HAM-D model and the p_t<0.65 HAM-A model, these two were used to determine significance of using this method of feature selection. This resulted in two new models generated using subcortical HAM-D and HAM-A data; only using the top 25 correlational features to the respective change scores in the model.

For the new HAM-D model, the Spearman correlation determined by LOOCV was r=0.355 (p=0.042). When compared to the subcortical p_t<0.45 HAM-D model, a Fishers r-to-z test was not significant (p=0.070). For the new HAM-A model, the Spearman correlation determined by LOOCV was r=0.366 (p=0.036). When compared to the subcortical p_t<0.65 HAM-A model, a Fishers R-to-Z test was not significant (p=0.164).

Cortical Drug Ensemble Accuracy

Spearman correlations for cortical data were r=–0.009, r=–0.522, r=0.213 for HAM-D, SHAPS, and HAM-A, respectively (Fig. 2). Cortical data did now show any positive significant Spearman correlation (corrected and uncorrected: p>0.05) (Fig. 3). There were no values in which HAM-D, HAM-A, or SHAPS KOR predictive performance was significantly better than placebo predictive performance using Fisher R-to-Z significance test (Supplemental Tables 5-7, Fig. 3).

Combined Subcortical and Cortical Drug Ensemble Accuracy

Without feature selection, combined subcortical and cortical individualized systems data Spearman correlations were r=0.156, r=–0.093, r=0.310 for HAM-D, SHAPS, and HAM-A, respectively (Fig. 2). The combined data were also not significant. Combined data showed significant positive Spearman correlation in HAM-D predictions against true scores for p_t -values 0.15, 0.30, and 0.70 post FDR correction (p<0.050) (Supplemental Table 8, Fig. 3). None of these were Fisher R-to-Z significant. No SHAPS or HAM-A scores were significant post FDR correction, but HAM-A was significantly better than placebo scores at the p_t=0.20 threshold (p=0.040) (Supplemental Table 9, Supplemental Table 10, Fig. 3).

Subcortical Feature Abstraction

Of the top 25 neural features abstracted from these thresholds (see Supplemental Table 11 and 12), five were shared between the HAM-D and HAM-A abstracted features. These features were the relationships between the (i) right hippocampus and dorsal attention regions, (ii) left nucleus accumbens and dorsal attention regions, (iii) right putamen and dorsal attention regions, (iv) left amygdala and temporal parietal, and (v) right amygdala and limbic regions.

Of the HAM-D selected features, there was a slightly higher percentage of right hemisphere features chosen; this contrasted with HAM-A selected features, which showed a moderately higher percentage of left hemisphere features. HAM-D selected features showed a strong number of putamen, global pallidus, amygdala, and hippocampus regions, while the nucleus accumbens, thalamus, and caudate nucleus regions were less represented (Supplemental Fig. 2). HAM-A selected features appeared qualitatively evenly distributed among the regions, although the putamen and nucleus accumbens regions were more highly represented. These qualitative trends towards differential representation of regions did not appear to be significantly different however, when statistically probed with a one-way ANOVA test for both HAM-D and HAM-A abstracted features (p=0.318, p=0.456). Tests for multiple comparisons did not appear to statistically differentiate the frequencies for region representation either.

The most prominent cortical networks of the HAM-D selected features were dorsal attention areas, limbic regions, and temporal parietal regions. Of the HAM-A selected features, the dorsal attention areas, limbic regions, and temporal parietal were most prominent. Almost all cortical networks were present in both HAM-D and HAM-A change-scores, but the salience/ventral attention A network was not present among either group of selected features, nor was the default network among HAM-D selected features (Supplemental Fig. 2).

While trends may have suggested dissimilar region frequency occurrences, there were low numbers of significant differences in feature occurrence. One exception was within the subcortical HAM-A feature set: the dorsal attention area was significantly greater in occurrence than default, salience ventral attention, and control regions, which were not present in subcortical HAM-A selected features (p=0.027). No other differences in frequency were significantly different for HAM-D comparisons.

Supplementary Discussion

Feature Selection Comparison

Further tests were used to assess whether using the top 25 correlating features were significant in a model predicting clinical change scores. While the algorithmic feature selection performed better than just taking the top 25 correlative features, this was not significant. This lack of significance of subcortical regions/ cortical networks towards the model reflects the inherent nuanced complexity of neural systems involved in the mechanisms underlying psychopathology and indicates the need for further individualized neural approaches to psychopathology. Prediction of treatment response via machine learning models have been performed across a number of studies [10, 11], but this study is the first-to-date, as far as we are aware, of features derived from individual neural network maps rather than collective group-averaged neural datasets.

Cortical and Combined Subcortical and Cortical Drug Ensemble Accuracy

The null findings for improved predictions of clinical symptom change from the cortical and combined data may suggest, that unique information derived from subcortical features within the ensemble models may be relatively more relevant to symptom response across time with respect to KOR antagonists as this dataset was found to be significant both with and without feature selection. Given the high performance at threshold levels of p_t=0.45 and p_t=0.65, subcortical features were further assessed at these two levels which suggested regions encompassing the hippocampus, left nucleus accumbens, right putamen, right and left amygdala, temporal parietal regions, limbic regions, and dorsal attention regions. These regions may provide predictive capacities of drug response in line with hypotheses on likely relationships within affective systems. Further research could help solidify regions such as these as adequate biomarkers towards treatment tools when treating depressive and anxiety symptoms.

Supplementary Tables and Figures

	Total	Drug	Placebo
Count	67	33	34
Mean age in years (SD)	39.8 (13.4)	40.3 (13.6)	39.6 (13.4)
Gender, % female	58.2	60.6	55.9
Race %			
Caucasian	65.7	69.7	61.8
African American	22.3	24.2	20.6
Asian	2.9	0	5.8
American Indian/Alaskan Native	0	0	0
More than one race	7.4	6	8.8
Ethnicity, % Hispanic or Latino	10.5	9.1	11.8
Mean baseline HAM-D (SD)	14.8 (5.3)	14.5 (4.7)	15.0 (5.9)
Mean baseline SHAPS (SD)	34.3 (6.8)	35.1 (8.1)	33.6 (5.2)
Mean baseline HAM-A (SD)	14.8 (6.3)	14.1 (5.2)	15.6 (7.1)
Mean post HAM-D (SD)		9.8 (6.6)	10.9 (7.3)
Mean post SHAPS (SD)		30.9 (7.8)	31.4 (6.8)
Mean post HAM-A (SD)		9.7 (7.4)	10.8 (7.4)

Supplemental Table 1. Study Participant Demographic and Baseline Data. Summary of participant demographic features and Hamilton Depression Rating Scale (HAM-D), Snaith–Hamilton Pleasure Scale (SHAPS) and Hamilton Anxiety Rating Scale (HAM-A) data by mean and standard deviation (SD at baseline and an 8-week follow up.

HAM-D						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	0.1243	0.4908	0.4908	0.1722	0.4241	0.4556
0.1	0.3731	0.0325	0.0433	0.1734	0.1985	0.2647
0.15	0.3565	0.0417	0.0521	0.2277	0.2908	0.3635
0.2	0.3052	0.0841	0.0990	0.2654	0.4328	0.4556
0.25	0.1872	0.2970	0.3126	0.2714	0.3641	0.4284
0.3	0.2450	0.1694	0.1882	0.2258	0.4684	0.4684
0.35	0.4969	0.0033	0.0073	0.2182	0.1033	0.2512
0.4	0.6171	0.0001	0.0009	0.2224	0.0268	0.2512
0.45	0.6341	0.0001	0.0009	0.2938	0.0410	0.2512
0.5	0.6187	0.0001	0.0009	0.2690	0.0404	0.2512
0.55	0.5866	0.0003	0.0017	0.2604	0.0565	0.2512
0.6	0.5470	0.0010	0.0039	0.2562	0.0846	0.2512
0.65	0.5246	0.0017	0.0049	0.2538	0.1035	0.2512
0.7	0.5033	0.0028	0.0071	0.2554	0.1267	0.2512
0.75	0.5377	0.0013	0.0042	0.2624	0.0973	0.2512
0.8	0.4890	0.0039	0.0078	0.2600	0.1471	0.2512
0.85	0.4622	0.0068	0.0097	0.2605	0.1810	0.2586
0.9	0.4662	0.0062	0.0096	0.2586	0.1738	0.2586
0.95	0.4838	0.0043	0.0078	0.2473	0.1411	0.2512
1	0.4803	0.0047	0.0078	0.2531	0.1507	0.2512

Supplemental Table 2: Hamilton Depression Rating Scale (HAM-D) Change Score determined by Subcortical Data Ensemble Model Results by Threshold. Summary of results based on the subcortical data model on HAM-D change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

SHAPS						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	0.0186	0.9182	0.9182	-0.1011	0.3196	0.3196
0.1	0.1470	0.4145	0.5113	-0.0884	0.1778	0.2092
0.15	0.2371	0.1840	0.4089	-0.1173	0.0802	0.1782
0.2	0.1629	0.3651	0.5113	-0.0917	0.1585	0.2050
0.25	0.1706	0.3426	0.5113	-0.1126	0.1326	0.1983
0.3	0.1610	0.3707	0.5113	-0.1151	0.1388	0.1983
0.35	0.1408	0.4346	0.5113	-0.1084	0.1640	0.2050
0.4	0.2736	0.1234	0.3545	-0.1082	0.0642	0.1645
0.45	0.3224	0.0673	0.3545	-0.1145	0.0397	0.1624
0.5	0.3768	0.0306	0.3545	-0.1339	0.0191	0.1624
0.55	0.2731	0.1241	0.3545	-0.1259	0.0561	0.1645
0.6	0.2411	0.1765	0.4089	-0.1393	0.0658	0.1645
0.65	0.3541	0.0432	0.3545	-0.1306	0.0251	0.1624
0.7	0.3107	0.0785	0.3545	-0.1373	0.0364	0.1624
0.75	0.2971	0.0932	0.3545	-0.1393	0.0406	0.1624
0.8	0.1540	0.3922	0.5113	-0.1290	0.1329	0.1983
0.85	0.0736	0.6841	0.7201	-0.1357	0.2059	0.2167
0.9	0.1431	0.4269	0.5113	-0.1405	0.1324	0.1983
0.95	0.0801	0.6577	0.7201	-0.1400	0.1939	0.2154
1	0.1677	0.3508	0.5113	-0.1444	0.1096	0.1983

Supplemental Table 3: Snaith–Hamilton Pleasure Scale (SHAPS) Change Score determined by Subcortical Data Ensemble Model Results by Threshold. Summary of results based on the subcortical data model on SHAPS change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

HAM-A						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	0.3352	0.0565	0.0598	0.0065	0.0908	0.1225
0.1	0.4615	0.0069	0.0220	0.0226	0.0314	0.0966
0.15	0.3746	0.0317	0.0453	0.0265	0.0758	0.1222
0.2	0.3349	0.0568	0.0598	0.0484	0.1209	0.1273
0.25	0.2442	0.1707	0.1707	0.0671	0.2386	0.2386
0.3	0.3628	0.0380	0.0478	0.0578	0.1041	0.1225
0.35	0.4200	0.0150	0.0327	0.0539	0.0621	0.1222
0.4	0.4206	0.0148	0.0327	0.0397	0.0553	0.1222
0.45	0.4799	0.0047	0.0188	0.0498	0.0324	0.0966
0.5	0.5339	0.0014	0.0085	0.0379	0.0147	0.0845
0.55	0.4557	0.0077	0.0220	0.0237	0.0338	0.0966
0.6	0.5252	0.0017	0.0085	0.0398	0.0169	0.0845
0.65	0.5624	0.0007	0.0066	0.0336	0.0093	0.0845
0.7	0.5620	0.0007	0.0066	0.0538	0.0115	0.0845
0.75	0.3754	0.0313	0.0453	0.0660	0.0997	0.1225
0.8	0.3878	0.0258	0.0429	0.0482	0.0794	0.1222
0.85	0.3623	0.0382	0.0478	0.0510	0.0998	0.1225
0.9	0.3573	0.0412	0.0485	0.0651	0.1141	0.1268
0.95	0.4025	0.0202	0.0367	0.0649	0.0789	0.1222
1	0.4149	0.0163	0.0327	0.0727	0.0750	0.1222

Supplemental Table 4: Hamilton Anxiety Rating Scale (HAM-A) Change Score determined by Subcortical Data Ensemble Model Results by Threshold. Summary of results based on the subcortical data model on HAM-A change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

HAM-D						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	-0.0486	0.7881	0.8887	0.0371	0.3689	-0.0486
0.1	-0.2474	0.1652	0.8166	-0.1388	0.3297	-0.2474
0.15	-0.2269	0.2041	0.8166	-0.1654	0.4014	-0.2269
0.2	-0.0356	0.8443	0.8887	-0.1166	0.3751	-0.0356
0.25	-0.2987	0.0913	0.8166	-0.0371	0.1450	-0.2987
0.3	-0.2301	0.1977	0.8166	-0.0222	0.2038	-0.2301
0.35	-0.2500	0.1605	0.8166	-0.0001	0.1594	-0.2500
0.4	-0.1209	0.5027	0.8378	-0.0189	0.3444	-0.1209
0.45	-0.1462	0.4168	0.8378	-0.0084	0.2938	-0.1462
0.5	-0.0434	0.8103	0.8887	-0.0151	0.4559	-0.0434
0.55	0.0486	0.7881	0.8887	-0.0172	0.3986	0.0486
0.6	0.0942	0.6019	0.8887	-0.0327	0.3096	0.0942
0.65	0.0766	0.6716	0.8887	-0.0311	0.3368	0.0766
0.7	0.1425	0.4287	0.8378	-0.0354	0.2423	0.1425
0.75	0.1758	0.3279	0.8378	-0.0290	0.2099	0.1758
0.8	0.0537	0.7668	0.8887	-0.0415	0.3550	0.0537
0.85	0.1223	0.4979	0.8378	-0.0424	0.2594	0.1223
0.9	0.1407	0.4348	0.8378	-0.0422	0.2365	0.1407
0.95	0.1524	0.3970	0.8378	-0.0555	0.2070	0.1524
1	-0.0091	0.9601	0.9601	-0.0679	0.4089	-0.0091

Supplemental Table 5: Hamilton Depression Rating Scale (HAM-D) Change Score determined by Cortical Data Ensemble Model Results by Threshold. Summary of results based on the cortical data model on HAM-D change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

SHAPS						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	-0.5901	0.0003	0.0041	-0.2590	0.0535	-0.5901
0.1	-0.3527	0.0441	0.0735	-0.1003	0.1478	-0.3527
0.15	-0.2317	0.1944	0.2046	-0.2096	0.4638	-0.2317
0.2	-0.4288	0.0128	0.0256	-0.1999	0.1590	-0.4288
0.25	-0.2621	0.1407	0.1563	-0.1800	0.3680	-0.2621
0.3	-0.1997	0.2651	0.2651	-0.1714	0.4544	-0.1997
0.35	-0.3205	0.0690	0.0862	-0.1286	0.2140	-0.3205
0.4	-0.4460	0.0093	0.0232	-0.1214	0.0812	-0.4460
0.45	-0.2787	0.1163	0.1369	-0.1250	0.2654	-0.2787
0.5	-0.3742	0.0319	0.0581	-0.1055	0.1309	-0.3742
0.55	-0.3425	0.0510	0.0785	-0.0990	0.1572	-0.3425
0.6	-0.3286	0.0619	0.0858	-0.0955	0.1689	-0.3286
0.65	-0.3257	0.0643	0.0858	-0.0997	0.1764	-0.3257
0.7	-0.4306	0.0124	0.0256	-0.0549	0.0566	-0.4306
0.75	-0.4606	0.0070	0.0200	-0.0381	0.0362	-0.4606
0.8	-0.5049	0.0027	0.0091	-0.0598	0.0264	-0.5049
0.85	-0.5541	0.0008	0.0041	-0.0564	0.0133	-0.5541
0.9	-0.5697	0.0005	0.0041	-0.0680	0.0119	-0.5697
0.95	-0.5581	0.0007	0.0041	-0.0776	0.0155	-0.5581
1	-0.5220	0.0018	0.0073	-0.0948	0.0294	-0.5220

Supplemental Table 6: Snaith–Hamilton Pleasure Scale (SHAPS) Change Score determined by Cortical Data Ensemble Model Results by Threshold. Summary of results based on the cortical data model on SHAPS change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

HAM-A						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	0.1702	0.3437	0.8592	-0.1288	0.1196	0.1702
0.1	-0.0209	0.9079	0.9270	-0.1376	0.3232	-0.0209
0.15	0.0905	0.6166	0.8680	-0.1799	0.1436	0.0905
0.2	0.0533	0.7684	0.8680	-0.1450	0.2181	0.0533
0.25	-0.2308	0.1962	0.6663	-0.1580	0.3837	-0.2308
0.3	-0.0817	0.6511	0.8680	-0.1320	0.4214	-0.0817
0.35	-0.1029	0.5690	0.8680	-0.1353	0.4489	-0.1029
0.4	-0.0709	0.6952	0.8680	-0.1454	0.3842	-0.0709
0.45	-0.2290	0.1999	0.6663	-0.1511	0.3761	-0.2290
0.5	-0.3069	0.0824	0.6663	-0.1421	0.2484	-0.3069
0.55	-0.2355	0.1870	0.6663	-0.1493	0.3632	-0.2355
0.6	-0.1215	0.5008	0.8680	-0.1251	0.4942	-0.1215
0.65	-0.2313	0.1952	0.6663	-0.1145	0.3189	-0.2313
0.7	-0.2352	0.1876	0.6663	-0.1291	0.3340	-0.2352
0.75	0.0503	0.7812	0.8680	-0.1146	0.2592	0.0503
0.8	0.0166	0.9270	0.9270	-0.1190	0.2975	0.0166
0.85	0.1142	0.5267	0.8680	-0.0964	0.2045	0.1142
0.9	0.0985	0.5855	0.8680	-0.0943	0.2250	0.0985
0.95	0.1451	0.4205	0.8680	-0.1051	0.1630	0.1451
1	0.2129	0.2342	0.6691	-0.1085	0.1021	0.2129

Supplemental Table 7: Hamilton Anxiety Rating Scale (HAM-A) Change Score determined by Cortical Data Ensemble Model Results by Threshold. Summary of results based on the cortical data model on HAM-A change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

HAM-D						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	0.1206	0.5039	0.5039	0.1094	0.4824	0.1206
0.1	0.4117	0.0173	0.0691	0.1617	0.1418	0.4117
0.15	0.5096	0.0024	0.0310	0.2029	0.0820	0.5096
0.2	0.3866	0.0263	0.0700	0.2142	0.2289	0.3866
0.25	0.4134	0.0168	0.0691	0.1884	0.1654	0.4134
0.3	0.4644	0.0065	0.0432	0.2263	0.1436	0.4644
0.35	0.2886	0.1033	0.1292	0.2419	0.4222	0.2886
0.4	0.3049	0.0845	0.1126	0.2543	0.4152	0.3049
0.45	0.3473	0.0477	0.0867	0.2697	0.3688	0.3473
0.5	0.3555	0.0423	0.0846	0.2722	0.3589	0.3555
0.55	0.3379	0.0544	0.0907	0.2593	0.3679	0.3379
0.6	0.3834	0.0276	0.0700	0.2503	0.2813	0.3834
0.65	0.3825	0.0280	0.0700	0.2478	0.2792	0.3825
0.7	0.4992	0.0031	0.0310	0.2474	0.1241	0.4992
0.75	0.3599	0.0397	0.0846	0.2493	0.3168	0.3599
0.8	0.3077	0.0815	0.1126	0.2475	0.3993	0.3077
0.85	0.3069	0.0824	0.1126	0.2391	0.3873	0.3069
0.9	0.2145	0.2307	0.2714	0.2383	0.4609	0.2145
0.95	0.1654	0.3578	0.3975	0.2361	0.3867	0.1654
1	0.1555	0.3877	0.4081	0.2369	0.3704	0.1555

Supplemental Table 8: Hamilton Depression Rating Scale (HAM-D) Change Score determined by Combined Data Ensemble Model Results by Threshold. Summary of results based on the combined data model on HAM-D change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

SHAPS						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	-0.0705	0.6965	0.8194	-0.1388	0.3937	-0.0705
0.1	0.0586	0.7458	0.8287	-0.1530	0.2028	0.0586
0.15	0.1185	0.5114	0.7636	-0.1511	0.1447	0.1185
0.2	0.0447	0.8047	0.8471	-0.1601	0.2104	0.0447
0.25	0.0000	1.0000	1.0000	-0.1518	0.2751	0.0000
0.3	-0.1418	0.4313	0.7636	-0.1304	0.4820	-0.1418
0.35	-0.1776	0.3227	0.7636	-0.1142	0.4001	-0.1776
0.4	-0.1567	0.3839	0.7636	-0.1105	0.4272	-0.1567
0.45	-0.3445	0.0496	0.7636	-0.0898	0.1467	-0.3445
0.5	-0.2743	0.1224	0.7636	-0.1109	0.2532	-0.2743
0.55	-0.2146	0.2303	0.7636	-0.1174	0.3480	-0.2146
0.6	-0.1699	0.3445	0.7636	-0.1078	0.4024	-0.1699
0.65	-0.1595	0.3752	0.7636	-0.0887	0.3893	-0.1595
0.7	-0.1523	0.3974	0.7636	-0.1065	0.4279	-0.1523
0.75	-0.1423	0.4297	0.7636	-0.0918	0.4208	-0.1423
0.8	-0.1317	0.4650	0.7636	-0.1025	0.4540	-0.1317
0.85	-0.1121	0.5345	0.7636	-0.0997	0.4804	-0.1121
0.9	-0.1252	0.4877	0.7636	-0.1020	0.4635	-0.1252
0.95	-0.0791	0.6617	0.8194	-0.1034	0.4619	-0.0791
1	-0.0928	0.6074	0.8098	-0.1130	0.4683	-0.0928

Supplemental Table 9: Snaith–Hamilton Pleasure Scale (SHAPS) Change Score determined by Combined Ensemble Model Results by Threshold. Summary of results based on the combined data model on SHAPS change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

HAM-A						
Feature Selection Threshold	**Correlation Score**	**Correlation Significance**	**Corrected Correlation Significance**	**Placebo Correlation Score**	**Fisher r-z Significance**	**Corrected Fisher r-z**
0.05	0.2653	0.2933	0.3551	0.4186	0.2156	0.2828
0.1	0.1356	0.0976	0.0425	0.0153	0.2282	0.1108
0.15	0.2086	0.2086	0.2086	0.2086	0.3102	0.2086
0.2	-0.0219	-0.0147	-0.0315	-0.0022	-0.0280	-0.0407
0.25	0.1257	0.1080	0.0579	0.0400	0.1674	0.0978
0.3	0.2653	0.2933	0.3551	0.4186	0.2156	0.2828
0.35	0.1356	0.0976	0.0425	0.0153	0.2282	0.1108
0.4	0.2086	0.2086	0.2086	0.2086	0.3102	0.2086
0.45	-0.0219	-0.0147	-0.0315	-0.0022	-0.0280	-0.0407
0.5	0.1257	0.1080	0.0579	0.0400	0.1674	0.0978
0.55	0.2653	0.2933	0.3551	0.4186	0.2156	0.2828
0.6	0.1356	0.0976	0.0425	0.0153	0.2282	0.1108
0.65	0.2086	0.2086	0.2086	0.2086	0.3102	0.2086
0.7	-0.0219	-0.0147	-0.0315	-0.0022	-0.0280	-0.0407
0.75	0.1257	0.1080	0.0579	0.0400	0.1674	0.0978
0.8	0.2653	0.2933	0.3551	0.4186	0.2156	0.2828
0.85	0.1356	0.0976	0.0425	0.0153	0.2282	0.1108
0.9	0.2086	0.2086	0.2086	0.2086	0.3102	0.2086
0.95	-0.0219	-0.0147	-0.0315	-0.0022	-0.0280	-0.0407
1	0.1257	0.1080	0.0579	0.0400	0.1674	0.0978

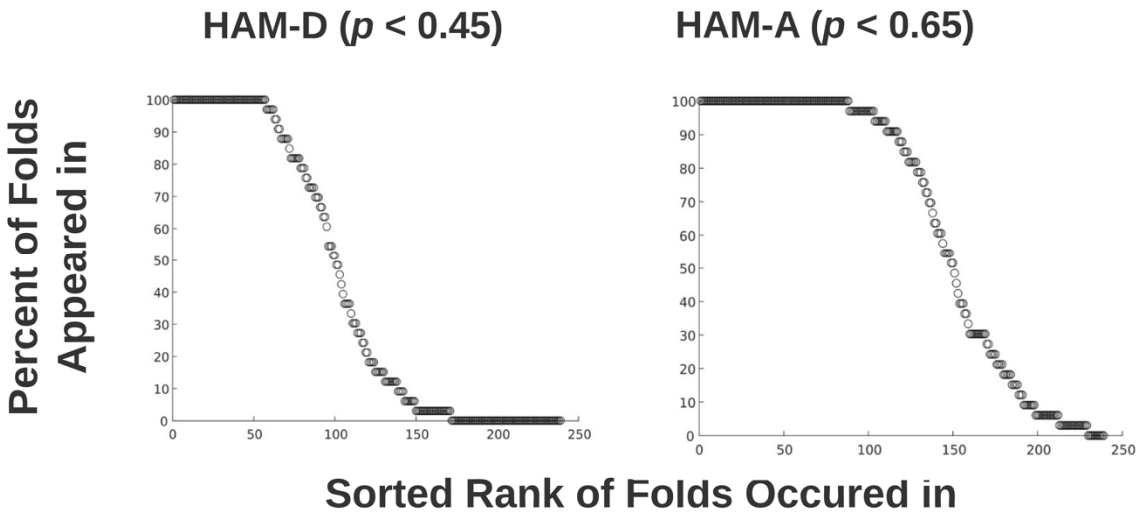
Supplemental Table 10: Hamilton Anxiety Rating Scale (HAM-A) Change Score determined by Combined Data Ensemble Model Results by Threshold. Summary of results based on the combined data model on HAM-A change scores, with each Spearman correlation-based thresholds. Correlation scores and significance scores are determined by Spearman correlation of veridical data against predicted data determined by the determined ensemble model.

HAM- D		
Feature	**% of Model Occurrence**	**Importance**
Left Putamen – Dorsal Attention B	100	0.033957759
Left Hippocampus – Somatomotor B	100	0.019215166
Right Caudate Nucleus – Limbic A	87.88	0.010810147
Right Putamen – Default C	100	0.008783307
Right Hippocampus – Dorsal Attention B	100	0.006860286
Left Amygdala – Central Visual	81.82	0.004269485
Left Amygdala – Dorsal Attention B	100	0.003669994
Right Amygdala – Default A	100	0.001511568
Right Amygdala – Limbic B	84.85	0.001199421
Left Thalamus – Default A	100	0.001146271
Right Putamen – Temporal Parietal	100	0.000754547
Left Nucleus Accumbens – Dorsal Attention A	100	0.000538508
Right Thalamus – Dorsal Attention B	100	0.000507222
Right Putamen – Dorsal Attention A	100	0.000447356
Right Global Pallidus – Ventral Attention B	100	0.000284192
Left Global Pallidus – Control B	100	0.000232368
Left Hippocampus – Temporal Parietal	100	0.000208981
Right Global Pallidus – Somatomotor A	100	0.000195525
Left Global Pallidus – Default C	100	0.000127003
Left Caudate Nucleus – Dorsal Attention B	100	1.03E-04
Left Amygdala – Temporal Parietal	100	4.67E-05
Left Putamen – Dorsal Attention A	100	3.99E-05
Right Putamen – Control B	87.88	2.60E-05
Right Hippocampus – Dorsal Attention A	100	1.58E-05
Right Global Pallidus – Limbic B	96.97	1.55E-05

Supplemental Table 11: Hamilton Depression Rating Scale (HAM-D) 25 Key Abstracted Features. List of the key features within the best performing HAM-D subcortical model. Key features were defined as being in 70% of LOOCV models after feature selection and the top 25 ensemble weighting. Features are described by their region-network named pair, percent of models appeared in, and ensemble model determined importance. Features are ranked by ensemble importance.

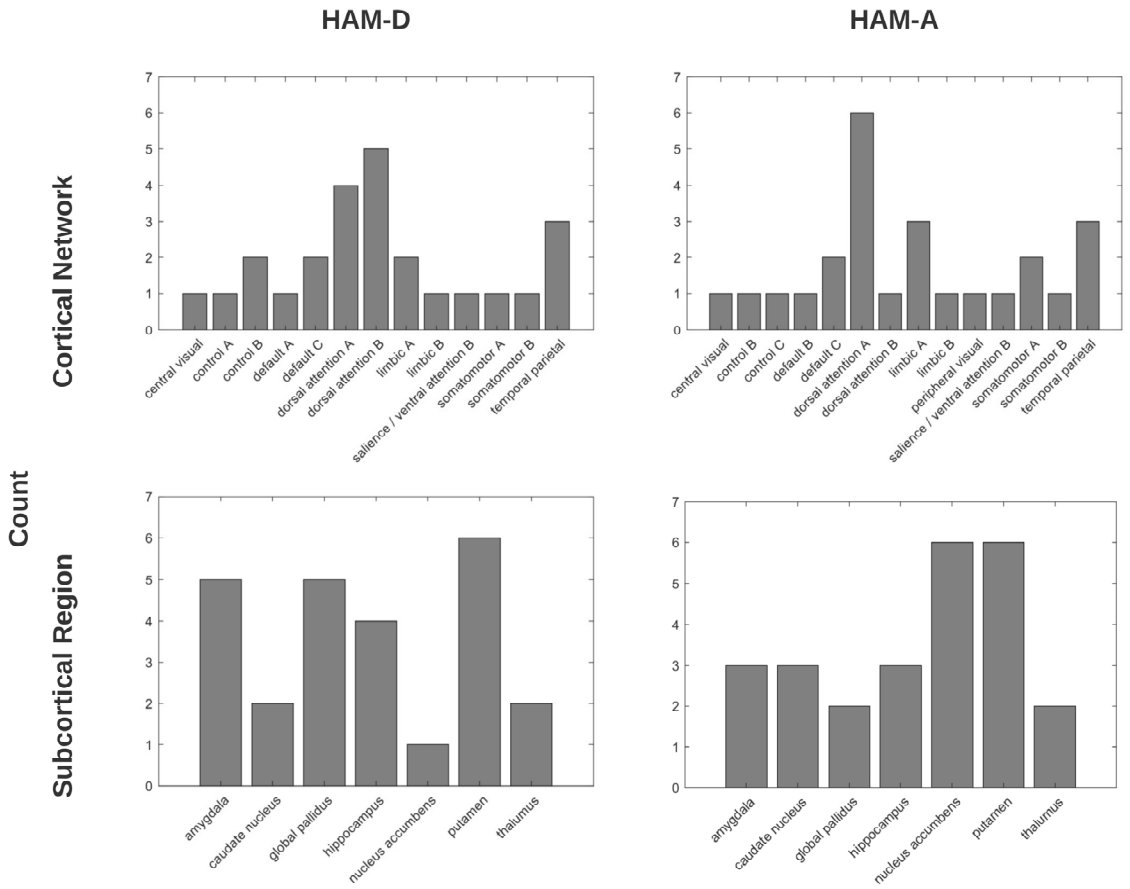
HAM - A		
Features	**% of Model Occurrence**	**Importance**
Right Hippocampus – Dorsal Attention A	100	0.051589
Left Nucleus Accumbens – Dorsal Attention A	100	0.034451
Left Putamen – Peripheral Vision	100	0.015294
Left Nucleus Accumbens – Temporal Parietal	96.97	0.013756
Right Putamen – Dorsal Attention A	100	0.006324
Left Hippocampus – Somatomotor A	100	0.004929
Right Thalamus – Dorsal Attention A	96.97	0.002749
Right Nucleus Accumbens – Limbic A	96.97	0.001101
Left Nucleus Accumbens – Ventral Attention B	100	0.00095
Right Putamen – Limbic A	81.82	0.000948
Left Amygdala – Temporal Parietal	100	0.000476
Right Amygdala – Somatomotor A	100	0.000306
Left Thalamus – Temporal Parietal	100	0.000299
Left Nucleus Accumbens – Dorsal Attention B	87.88	0.000261
Right Global Pallidus – Default C	93.94	0.000183
Right Caudate Nucleus – Default C	90.91	0.000166
Right Caudate Nucleus – Control B	75.76	0.000129
Left Putamen – Dorsal Attention A	100	0.000125
Left Global Pallidus – Limbic B	100	0.000108
Left Putamen – Somatomotor B	100	7.93E-05
Left Putamen – Central Visual	96.97	6.41E-05
Right Amygdala – Limbic A	100	4.73E-05
Right Hippocampus – Default B	100	4.18E-05
Right Nucleus Accumbens – Dorsal Attention A	100	2.96E-05
Left Caudate Nucleus – Control C	100	2.76E-05

Supplemental Table 12: Hamilton Anxiety Rating Scale (HAM-A) 25 Key Abstracted Features. List of the key features within the best performing HAM-A subcortical model. Key features were defined as being in 70% of LOOCV models after feature selection and the top 25 ensemble weighting. Features are described by their region-network named pair, percent of models appeared in, and ensemble model determined importance. Features are ranked by ensemble importance.



Supplemental Figure 1: Feature Occurrence for KOR group during feature selection

From the peak performing HAM-D and HAM-A subcortical feature selected models ($p < 0.45$, $p<0.65$), the occurrence of what percent of folds each feature was selected to be in during feature selection.



Supplemental Figure 2. Feature Selection

From the peak performing Hamilton Depression Rating Scale (HAM-D) and Hamilton Anxiety Rating Scale (HAM-A) subcortical feature selected models ($p < 0.45$, $p<0.65$), the top 25 features relevant to each peak correlational threshold feature selected model were abstracted. Features were determined to be most relevant by feature selection occurrence, how many folds did each feature appear in when the correlational threshold was applied ($n>70\%$ of folds), and the abstracted feature importance generated by the ensemble model. The cortical networks (top) represent the sum of all the subcortical voxels won by that particular cortical network. The subcortical structures (bottom) represent the location of the voxel regardless of their cortical network label. All plots illustrate frequency of occurrence, and are not differentiated based on left or right hemisphere.

References:

1. Krystal AD, Pizzagalli DA, Smoski M, Mathew SJ, Nurnberger J, Lisanby SH, et al. A randomized proof-of-mechanism trial applying the 'fast-fail' approach to evaluating κ-opioid antagonism as a treatment for anhedonia. Nat Med. 2020;26:760–768.
2. Gordon EM, Laumann TO, Adeyemo B, Gilmore AW, Nelson SM, Dosenbach NUF, et al. Individual-specific features of brain systems identified with resting state functional correlations. NeuroImage. 2017;146:918–939.
3. Wang D, Buckner RL, Fox MD, Holt DJ, Holmes AJ, Stoecklein S, et al. Parcellating cortical functional networks in individuals. Nat Neurosci. 2015;18:1853–1860.
4. Li M, Wang D, Ren J, Langs G, Stoecklein S, Brennan BP, et al. Performing group-level functional image analyses based on homologous functional regions mapped in individuals. PLOS Biol. 2019;17:e2007032.
5. Greene DJ, Marek S, Gordon EM, Siegel JS, Gratton C, Laumann TO, et al. Integrative and Network-Specific Connectivity of the Basal Ganglia and Thalamus Defined in Individuals. Neuron. 2020;105:742-758.e6.
6. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. 2001;5:143–156.
7. Jenkinson M, Bannister P, Brady M, Smith S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage. 2002;17:825–841.
8. Andersson JL, Jenkinson M, Smith S. Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. FMRIB Anal Group Univ Oxf. 2007;2:e21.
9. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp. 2002;17:143–155.
10. Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Bürger C, et al. Prediction of Individual Response to Electroconvulsive Therapy via Machine Learning on Structural Magnetic Resonance Imaging Data. JAMA Psychiatry. 2016;73:557–564.
11. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med. 2017;23:28–38.