Patterns of Pre-Treatment Reward Task Brain Activation Predict Individual Antidepressant Response:

Key Results from the EMBARC Randomized Clinical Trial

Kevin P. Nguyen, BS[a]; Cherise Chin Fatt, PhD[b]; Alex Treacher, BS[a]; Cooper Mellema, BS[a]; Crystal Cooper, PhD[b,c]; Manish K. Jha, MD[b]; Benji Kurian, MD[b]; Maurizio Fava, MD[d]; Patrick J. McGrath, MD[e]; Myrna Weissman, PhD[e]; Mary L. Phillips, MD[f]; Madhukar H. Trivedi, MD*[,b]; and Albert Montillo, PhD*[,a,g,h]

Affiliations:

a. Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX
b. Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX
c. Jane and John Justin Neuroscience Center, Cook Children's Health Care System, Fort Worth, TX
d. Department of Psychiatry, Massachusetts General Hospital, Boston, MA
e. New York State Psychiatric Institute and Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York, NY
f. Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA
g. Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX
h. Advanced imaging Research Center, UT Southwestern Medical Center, Dallas, TX

Address Correspondence to:
Albert Montillo, PhD
Assistant Professor of Bioinformatics
Assistant Professor of Radiology and the Advanced Imaging Research Center
Director, Deep Learning for Precision Health Laboratory
University of Texas Southwestern Medical Center
albert.montillo@utsouthwestern.edu

Madhukar H. Trivedi, M.D.
Professor of Psychiatry
Julie K. Hersh Chair for Depression Research and Clinical Care
Betty Jo Hay Distinguished Chair in Mental Health
Director, Center for Depression Research and Clinical Care
University of Texas Southwestern Medical Center
madhukar.trivedi@utsouthwestern.edu

*Drs. Montillo and Trivedi contributed equally as senior authors to this work
Disclosures are reported at the end of the main text.

Running Title: Reward Task Activation Predicts Antidepressant Response
Keywords: antidepressants, depression, treatment selection, fMRI, deep learning, precision medicine

ABSTRACT

Background: The lack of biomarkers to inform antidepressant selection is a key challenge in personalized depression treatment. This work identifies candidate biomarkers by building deep learning predictors of individual treatment outcomes using reward processing measures from functional MRI, clinical assessments, and demographics.

Methods: Participants in the Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care (EMBARC) study (n = 222) underwent reward processing task-based functional MRI at baseline and were randomized to 8 weeks of sertraline (n = 106) or placebo (n = 116). Subsequently, sertraline non-responders (n = 37) switched to 8 weeks of bupropion. The change in Hamilton Rating Scale for Depression (ΔHAMD) was measured after treatment. Reward processing, clinical measurements, and demographics were used to train treatment-specific deep learning models.

Results: The predictive model for sertraline achieved R^2 of 48% (95% CI 33-61%, $p < 10^{-3}$) in predicting ΔHAMD and number-needed-to-treat (NNT) of 4.86 participants in predicting response. The placebo model achieved R^2 of 28% (95% CI 15-42%, $p < 10^{-3}$) and NNT of 2.95 in predicting response. The bupropion model achieved R^2 of 34% (95% CI 10-59%, $p < 10^{-3}$) and NNT of 1.68 in predicting response. Brain regions where reward processing activity was predictive included the prefrontal cortex and cerebellar crus 1 for sertraline and the cingulate cortex, caudate, orbitofrontal cortex, and crus 1 for bupropion.

Conclusions: These findings demonstrate the utility of reward processing measurements and deep learning to predict antidepressant outcomes and to form multimodal treatment biomarkers.

Clinical trial registration: NCT01407094

INTRODUCTION

The discovery of biomarkers of antidepressant response is crucial to achieving personalized treatment planning in Major Depressive Disorder (MDD). Currently, remission rates for individual antidepressants are typically below 40%(1), and about 33% of patients require > 3-4 drug trials before achieving remission(2). However, biomarkers and predictive tools that optimize antidepressant selection for each patient would reduce the need for multiple drug trials, expedite remission, and enable a precision medicine approach to MDD treatment.

Noninvasive measurements of individual brain activity show promise as pre-treatment markers of treatment outcome. For example, default mode network and hippocampal connectivity in resting-state fMRI has been correlated with sertraline treatment outcome(3). Greater anterior cingulate cortex activation during emotional regulation has been associated with improved response to venlafaxine and fluoxetine(4; 5). Similarly, amygdala activation in response to emotional stimuli has been connected to non-specific treatment outcome(6). Compared to the serotonergic emotion regulation circuit, the dopaminergic reward processing circuity has been less studied for treatment outcome prediction, though one recent study has connected abnormal ventral striatum activity to better response to sertraline vs. placebo(7).

While previous research highlights the potential of neuroimaging to predict treatment response, models with greater statistical complexity are necessary to exploit the richness of fMRI data. This study employs deep learning models which may be more apt to discover the complex, nonlinear association between fMRI measurements and treatment outcome. Deep learning models can also scale to large numbers of multimodal input features, while automatically learning the most informative ones (8; 9). This capacity is exploited to integrate pre-treatment imaging and clinical measurements. Our previous work applied deep learning to predict bupropion outcome (10). The current analysis extends the previous work by developing predictors for three treatments and incorporating data augmentation, which mitigates overfitting and improves prediction accuracy.

This secondary analysis of the Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care (EMBARC) study determines whether pre-treatment reward task-based fMRI can be used to predict treatment-specific outcome. Models are constructed for the sertraline and placebo main treatment groups, as well as the secondary bupropion treatment group. Predictive performance is presented and preliminary composite candidate biomarkers, combining neuroimaging and clinical phenotype features, are identified for each treatment.

METHODS AND MATERIALS

Participants

Details of the EMBARC study design have been previously reported(11). A total of 296 participants with MDD were enrolled with written informed consent and IRB approval across 4 study sites: Columbia University, Massachusetts General Hospital, University of Texas Southwestern Medical Center, and University of Michigan. Inclusion criteria included early onset (before age 30) and chronic (episode duration > 2 years) or recurrent (2+ episodes) disease. Demographics are described in **Table 1**. Further details on inclusion/exclusion criteria (Appendix I.1) and a CONSORT flow diagram (**Fig. S1**) are in the Supplement.

Treatment Protocol and Outcomes

The treatment period included two 8-week stages. In Stage 1, participants were randomized under double-blind conditions into sertraline or placebo treatment arms. Randomization was stratified by study site, baseline depression severity, and disease duration. At week 8, sertraline-treated participants not meeting response criteria (Clinical Global Improvement score less than “much improved”) were crossed over under double-blinded conditions to bupropion treatment in Stage 2. The analyzed samples included participants who received sertraline or placebo in Stage 1 and those from the sertraline arm who received bupropion in Stage 2. Clinical severity was tracked using the 17-item Hamilton Rating Scale for

Depression (HAMD), and the primary outcome is the change in HAMD (ΔHAMD) over the 8-week treatment stage (week 8 minus baseline for sertraline and placebo, week 16 minus week 8 for bupropion). Secondary binary outcomes were defined using standard clinical criteria, including *response* (decrease in HAMD ≥ 50% from pre-treatment) and *remission* (week 8 HAMD ≤ 7). Dosage schedules are described in the Supplement, Appendix I.1; outcome and dosage characteristics for each treatment arm are presented in **Table 1**.

MRI Acquisition

Reward task-based fMRI was acquired at the baseline visit for 8 minutes during a block-design number-guessing task which probes reward processing neural circuitry known to be altered in MDD (paradigm described in Appendix I.2 and **Fig. S2**) (12; 13). This task includes trials in which money is lost for a wrong guess but not gained for a correct guess and trials in which money is gained for a correct guess but not lost for a wrong guess. The participant's differential brain activation is measured between punishing vs. rewarding trials. MRI acquisition details can be found in **Table S1**. Participants whose MRI contained focal signal loss or clipped field-of-view and who did not complete 8 weeks of a given treatment were removed. The analyzed cohort included 106 who completed 8 weeks of sertraline, 116 for placebo, and 37 for bupropion (222 total unique participants).

Data Augmentation

Data augmentation is commonly employed in deep learning to increase performance and avoid overfitting when data is limited (14). This technique generates additional image data, which reflects the natural diversity in the population, by perturbing the original acquired images. The approach used here was developed specifically for 4D fMRI and demonstrated to substantially improve deep learning performance in multiple neuroimaging modeling tasks (15; 16). Full detail of the approach is provided in Supplement, Appendix I.3. Augmentation was applied 10 times for each original fMRI, providing a total of 1060 images for sertraline, 1160 for

placebo, and 370 for bupropion. Importantly, this augmented data was used only during model training and not during evaluation. Additional results in **Table S4** detail the performance improvement achieved with vs. without augmentation.

MRI Preprocessing and Feature Extraction

Original and augmented fMRI were preprocessed using standard steps including skull-stripping, head motion correction, spatial normalization, and spatial smoothing with a 4mm FWHM kernel (see Supplement, Appendix I.4). Three contrast maps were computed for each participant, quantifying brain activation in the initial *anticipation* phase of each number-guessing trial, *reward expectancy* (differential activation in rewarding vs. punishing trials), and *prediction error* (after wrong guesses). Each contrast map was parcellated into 200 functional brain regions using spatially-constrained spectral clustering (13), yielding a total of 600 fMRI features for each participant. See Appendix I.4-8 for preprocessing details. Site effect did not have a significant impact on the results, and there was no association between motion and treatment outcome (**Fig. S3 and Fig. S4**).

Acquisition of Clinical Measurements

In addition to imaging features, 95 pre-treatment clinical measures and demographic features acquired on the same day as imaging were also included as predictor inputs, as machine learning models are suitable for combining such information (8; 9). Demographics consisted of race, ethnicity, age, education, biological sex, and marital status. Clinical measures included total scores and sub-scores for several participant-reported forms and clinician-administered assessments (**Table S2**). These included measurements such as HAMD, MASQ, and QIDS-SR16 (pre-treatment depression severity), CTQ (childhood trauma), ASRM (mania), CSSRS (suicide risk), STAI (anxiety), episode duration, and SAPAS (personality traits). Participant and family psychiatric history items were also included.

Deep Learning Model Training and Evaluation

Feed-forward neural networks were constructed to take fMRI and clinical features as inputs and predict ΔHAMD. A separate model was trained for each treatment: sertraline, bupropion, and placebo. Given the current unavailability of a similar second dataset for external validation, models were evaluated using *nested* cross-validation, which provides the next best estimate of real-world predictive performance (17; 18). For the sertraline and placebo models, 20 outer and 20 inner folds were used, while the bupropion model was validated with 15 outer and 15 inner folds given the smaller original sample size. Technical details of the model architecture (**Fig. S5**), hyperparameter optimization (**Table S3**), and nested cross-validation (Appendix I.9) are provided in the Supplement.

Accuracy metrics for predicting ΔHAMD included R^2 (coefficient of determination) and the root mean squared error (RMSE). The ΔHAMD predictions together with the baseline HAMD values were thresholded to compute predictions of the two binary outcomes: remission and response. Accuracy metrics for these binary outcomes included predictive value (PPV), area under the receiver operating characteristic curve (AUROC), and number-needed-to-treat (NNT). The NNT estimates the number of patients that would need to be treated, based on the model's prediction of remission/response, to achieve one additional success over random treatment assignment. Permutation testing was performed to verify the statistical significance of each model's performance (19), while the most important predictive features learned by the model were identified using the partial derivative method (20; 21) (Appendix I.11-12). In the Results, the 20 most important features are reported for each model; remaining features plateaued in importance.

RESULTS

Prediction of Sertraline Treatment Outcome

For the sertraline treatment arm (*n* = 106, 69% female, mean age 38.4), the mean ΔHAMD during Stage 1 was 7.89 ± 7.16, remission rate was 39%, and response rate was 54% (**Table 1**). The sertraline predictive model achieved a substantial R^2 of 48% (95% confidence interval [CI] of 33-61%) and RMSE of 5.15 in predicting ΔHAMD (**Table 2,** 1st row). NNT was 3.33 and PPV was 69% for predicting remission, and NNT was 4.86 and PPV was 68% for predicting response. Permutation testing confirmed these results to be statistically significant with $p < 0.01$.

Of the 20 most important features for predicting sertraline outcome, half were clinical features (**Fig. 1**). Psychomotor agitation and higher pre-treatment (week 0) symptomatic severity (17-item and 24-item HAMD total) predicted greater improvement (had a positive association with HAMD reduction). Family history of suicide, comorbidities (SCQ[a] total score), and older age at first dysphoric or depressive episode predicted lesser improvement (had a negative association with HAMD reduction). Examining the imaging features, greater improvement was predicted by higher brain activation during reward expectancy in the right inferior frontal gyrus pars triangularis; higher anticipation activation in the right middle occipital gyrus and right middle temporal gyrus; and higher prediction error activation in the right superior temporal gyrus. Lesser improvement was predicted by higher prediction error activation in left crus 1 of the cerebellum; higher reward expectancy activation in the right supramarginal gyrus and right posterior cingulum; and higher anticipation activation in the left superior frontal gyrus.

Prediction of Placebo Treatment Outcome

In the placebo treatment arm (*n* = 116, 63% female, mean age 37.4), the mean ΔHAMD during Stage 1 was 6.70 ± 6.93, not significantly different from the mean ΔHAMD of the

[a] Self-Administered Comorbidity Questionnaire

sertraline arm (p = 0.11). Remission and response rates were 33% and 35% respectively (**Table 1**). The placebo model attained an R^2 of 28% (95% CI of 15-42%) and RMSE of 5.87 for predicting ΔHAMD on held-out test data (**Table 2**, 2nd row), with $p < 0.01$ upon permutation testing. NNT was 2.06 and PPV was 81% for predicting remission, and NNT was 2.95 and PPV was 69% for predicting response (p = 0.02).

Clinical features most important for predicting outcomes (**Fig. 2**) included concurrent panic disorder and hypersomnia (from SCID[a]), older age, and anhedonia (MASQ[b] anhedonic depression score), all of which predicted lesser improvement. Greater improvement was predicted by Asian race and separated marital status (see Appendix II.4 in Supplement for discussion of caveats), NEO[c] openness score, and longer periods spent without dysphoria. The imaging features were distinct from those learned by the sertraline model (**Fig. 1**). Greater improvement was predicted by higher anticipation activation in the right middle occipital gyrus; higher reward expectancy activation in the right supramarginal gyrus; and higher prediction error activation in the left superior temporal gyrus, right inferior occipital gyrus, and left superior parietal lobule. Lesser improvement was predicted by higher prediction error activation in the left cerebellum, right middle temporal gyrus, and left middle frontal gyrus and higher anticipation activation in the left inferior occipital gyrus.

Prediction of Bupropion Treatment Outcome

Given the smaller sample size of the bupropion group (n = 37, 70% female, mean age 37.5), these results should be considered with caution compared to the sertraline and placebo groups. Nevertheless, they are presented here for completeness. Mean ΔHAMD was 5.46 ± 5.57, remission rate was 32%, and response rate was 41% for bupropion-treated participants (**Table 1**). This model achieved R^2 of 34% (95% CI of 10-59%) and RMSE of 4.46 in predicting

[a] Structured Clinical Interview for DSM-5
[b] Mood and Anxiety Symptom Questionnaire
[c] NEO-Five Factor Inventory

ΔHAMD (**Table 2**, 3rd row). NNT was 2.35 and PPV was 75% for predicting remission, and NNT was 1.68 and PPV was 100% for predicting response. These performance values were significant $p < 0.01$ and are an improvement from our previously published bupropion model trained without data augmentation, which demonstrated R^2 of 26% and RMSE of 4.71 (10).

Clinical features predicting greater improvement (**Fig. 3**) included more education and family history of mental illness or depression, while only completing up to high school and concurrent anxious distress (from SCID) predicted lesser improvement. Important imaging features that predicted greater improvement included higher prediction error activation in the right middle cingulate cortex and right caudate; higher anticipation activation in the right posterior cingulate cortex; and higher reward expectancy activation in cerebellum right crus 1 and the right hippocampus. Predictors of lesser improvement included higher anticipation activation in the right cerebellum and right superior frontal gyrus; higher reward expectancy activation in the left caudate, left medial orbitofrontal cortex, and left crus 1 of the cerebellum; and higher prediction error activation in the bilateral lingual gyri.

Patterns of Clinical Improvement Associated with Predicted Outcomes

Fig. 4 compares the change in true HAMD over the treatment period between 3 groups of participants: those with the greatest predicted improvement (top 25%), the least predicted improvement (bottom 25%), and the remaining 50%. For all treatments, the participants with greatest predicted improvement exhibited much faster improvement than others within the treatment group, in addition to starting at higher severity (see **Fig. S6** for plots of unnormalized HAMD). For the *sertraline group* (**Fig. 4a**), participants began demonstrating a marked separation in trajectories at week 3. In the *placebo group*, participants demonstrated a marked separation by week 5 (**Fig. 4b**). The *bupropion group* (**Fig. 4c**), which had the smallest sample size, demonstrated a similar though smaller separation between participants with greatest vs. least predicted improvement.

Treatment Specificity of the Predictive Models

To test whether the models had identified composite predictive biomarkers that are *specific to each treatment*, the model for each treatment was evaluated on participants from the other two treatment arms. In each case, the predictive performance was low (negative R^2), confirming that each model learned treatment-specific predictive features.

Ablation experiments and comparative statistical approaches

Ablation experiments indicated that both the combination of imaging and clinical/demographic features *and* fMRI data augmentation were necessary to achieve the observed performance (**Table S4**). A statistical parametric mapping analysis found no significant fMRI correlates of treatment outcome. Multiple classical machine learning approaches were tested, including an elastic net regressor, using the same clinical and augmented imaging features and same cross-validation method, but these did not explain any variance in the data. These results support the hypothesis that deep learning adds value for moderator identification. See Appendix II.1-2 for details.

DISCUSSION

These results describe the first ever deep learning predictive models for three treatments. All 3 models explained a substantial proportion of the variance in ΔHAMD, and NNT for predicting remission ranged from 2.35 for bupropion to 3.33 for sertraline, equal to or better than the NNTs of many depression treatments(22). Importantly, remission rates were similar across the 3 treatments at the *group* level: 39% for sertraline, 33% for placebo, and 32% for bupropion. This reinforces the need for tools that predict *individual* outcomes and preemptively identify the individuals who will respond. Note that given the cross-over design of the study, where participants were switched to bupropion after failing sertraline in Stage 1, the bupropion results should not be directly compared to the other treatment groups, which is discussed further in the limitations section.

Higher prediction accuracy was achieved than previously published predictors of individual antidepressant outcome. Etkin et al. developed an escitalopram predictor using cognitive and emotional behavioral measures achieving an NNT of 3.8, but their results were limited to cognitively-impaired individuals(23). Gordon et al. developed predictors for escitalopram and venlafaxine with NNTs of 2.7 and 4.6, respectively(24). They also developed a sertraline predictor with NNT of 3.5 and PPV of 43%, lower than the NNT of 3.33 and 62% PPV achieved here. Dunlop et al. used functional connectivity from resting-state fMRI to predict remission on escitalopram or duloxetine with AUROC of 0.72 (25). While this exceeds the AUROC of 0.60 and 0.71 for sertraline and bupropion achieved here, their results were not validated on held-out data for a less biased estimate of real-world performance. Previous analyses of other data modalities from EMBARC have explained similar or less variance in antidepressant outcomes, compared to the 48% for sertraline achieved here. Fonzo et al. explained 24% of the variance in sertraline outcomes using emotion task fMRI (26). Wu et al. explained 36% of the variance for sertraline and 17% for placebo using whole-brain resting-state EEG measurements (27). Pizzagalli et al. explained 40% of the variance in treatment outcomes using anterior cingulate cortex activity from EEG combined with clinical and demographic features, though this was neither antidepressant-specific nor validated on held-out data (28).

The combination of these predictive models presents a possible precision medicine approach for antidepressant selection. In practice, a patient would undergo a pre-treatment fMRI scan, and each model would be applied to provide within seconds a prediction of response to each treatment. The clinician could then prioritize the treatment with the best predicted outcome. Future imaging studies encompassing more treatments would allow the construction of additional treatment-specific models. Additionally, future studies are also needed to externally validate these preliminary models on a separate dataset with compatible data.

Examination of candidate neuroimaging biomarkers

The important imaging features identified by the models can be compared to existing studies of reward processing, MDD pathophysiology, and antidepressant response biomarkers (**Table 3**). In this section, the features are categorized into those with a previously known neurophysiological role in MDD and those that are novel and could warrant further investigation.

Both the sertraline and placebo models identified predictive regions in the prefrontal cortex (PFC), previously implicated in altered reward processing in MDD (29–32). The bupropion model identified other implicated regions, namely the middle cingulate cortex, caudate, and orbitofrontal cortex. For all treatment groups, activation in the frontal regions predicted lesser HAMD improvement.

Interestingly, cerebellar regions were predictive for all models. Though traditionally associated with motor control, recent developments have strongly linked the cerebellum to reward processing and particularly error-based learning (33; 34). For example, functional connectivity between cerebellar crus 1 and the PFC during reward processing was found to be elevated in MDD (35). Here, cerebellar activation during prediction error and reward expectancy was an important predictor for all models. These results suggest that cerebellum, especially crus 1, should be a target of future treatment moderator studies.

Other regions reported to have abnormal connectivity in MDD include the supramarginal gyrus, where reward expectancy activation predicted better sertraline outcome and worse placebo outcome, and the pars triangularis, which was an important feature for the sertraline model (36–38). Additionally, the bupropion model identified important predictive regions associated with emotional processing. Higher reward expectancy activation in the hippocampus predicted greater HAMD improvement. The hippocampus has been implicated in mood dysregulation in MDD through MRI volumetry and PET metabolic activity (29; 39; 40). Another bupropion predictive region was the posterior cingulate cortex, where emotion processing activation was previously shown to be modulated by bupropion (41).

Several regions were learned that have not previously associated with MDD (**Table 3**, bottom row). These included occipital and temporal regions for the sertraline and placebo models and superior parietal lobule for the placebo model. Further fMRI studies with other tasks may help to uncover their exact involvement in treatment response.

Synergy between neuroimaging and clinical features

The predictive clinical and demographic features are discussed in detail in the Supplement, Appendix II.4. Of note, psychomotor agitation was the most important feature for sertraline, predicting greater improvement. Pre-treatment HAMD score, family history of suicide, and comorbidity score were also important features. The placebo model identified concurrent panic disorder, hypersomnia, and older age as important features. For the bupropion model, education level, family history of mental illness, and anxious distress were important. Notably, *this is one of the first studies to synergize imaging measurements with another modality of information, namely clinical assessments, and this combination yielded the highest predictive signals* in the sertraline and placebo models. These findings support further investigations into cross-modal composite treatment moderators.

Limitations

While EMBARC is the largest randomized, placebo-controlled study of antidepressant response with fMRI to date, the cohort may not fully represent the general population. Inclusion criteria included early onset (before age 30), chronic (episode duration > 2 years), or recurrent (2+ episodes) disease which represents a MDD subpopulation. For further real-world performance estimation, the models should be tested on an additional, independent dataset. In the meantime, the rigorous cross-validation employed here provides high confidence that the results will generalize to additional cohorts.

For the sertraline treatment group, there is a potential contribution of placebo effect in the treatment outcome for some participants, given that the overall remission rates of the

sertraline and placebo groups were similar. To fully separate the medication effect from the placebo effect would likely require a fully crossed over study, in which participants received both sertraline and placebo treatments. In the meantime, the discriminant validity of the sertraline model, i.e. the inability of this model to predict on the placebo group, indicates that it did primarily learn a sertraline-specific biomarker.

A limitation for the bupropion treatment group comes from the two-stage crossover study design, which may have biased bupropion recipients to those less responsive to antidepressants and caused carry-over effects due to lack of a washout period. Consequently, the bupropion results should be compared to the sertraline and placebo results with caution. Follow-up studies with wider treatment design are warranted.

The bupropion group was limited to 37 participants, however, other studies have found significant biomarkers with only 10-20 participants(31; 5; 41). Moreover, data augmentation was used to simulate additional fMRI samples, yielding 370 total images for bupropion model training. The 95% confidence intervals were reported for each R^2 accounting for sample size (*before* data augmentation), indicating that at worst, the model still explained 10% of the variance. Furthermore, permutation testing to check the significance of results helped to ensure that findings were not spurious.

Conclusion

This is the first application of deep learning to predict *individual* treatment outcomes from fMRI in MDD. Deep learning models explained a substantial portion of the variance in treatment outcomes, integrating clinical assessments with imaging data. Important predictive clinical features can readily inform clinical treatment decision-making. Examination of the predictive imaging features revealed brain regions in concordance with previous knowledge of MDD neurophysiology as well as regions not previously implicated. Future efforts to develop analogous measurements of reward processing in these regions, such as through behavioral

markers, may form clinically useful tools for treatment selection. This work is an important step towards expediting the selection of appropriate antidepressants and achieving personalized treatment.

ACKNOWLEDGEMENTS

The EMBARC study was supported by the National Institute of Mental Health of the National Institutes of Health under award numbers U01MH092221 (to M.H.T.) and U01MH092250 (to P.J.M., R.V.P., and M.W.). Valeant Pharmaceuticals donated the Wellbutrin XL used in the study. This work was supported by the EMBARC National Coordinating Center at UT Southwestern Medical Center, with M.H.T. as Coordinating PI, and the Data Center at Columbia and Stony Brook Universities.

We thank Thomas Carmody, Andrew Czysz, Tracy Greer, Brittany Mason, Abu Minhajuddin, and Monica Basco for their feedback during the writing of this manuscript.

An early version of this manuscript was posted in on the bioRxiv preprint server at https://doi.org/10.1101/2020.01.28.923821.

CODE AVAILABILITY

To facilitate reuse and extension, we are pleased to provide full source code for these analyses at https://github.com/DeepLearningForPrecisionHealthLab/Antidepressant-Reward-fMRI

DISCLOSURES

M.J. has received contract research grants from Acadia Pharmaceuticals and Janssen Research & Development, educational grant to serve as Section Editor of the Psychiatry & Behavioral Health Learning Network, consultant fees from Eleusis Therapeutics US, Inc, and honoraria for CME presentations from North American Center for Continuing Medical Education and Global Medical Education. B.K. has received grant support from Evotec, Forest Pharmaceuticals, Johnson & Johnson, Naurex, NIMH, Pfizer, Rexahn, and Targacept and is employed by Health Care Service Corporation. M.F. has received research support from Abbot Laboratories Alkermes, American Cyanamid, Aspect Medical Systems, AstraZeneca, Avanir Pharmaceuticals, BioResearch, BrainCells, Bristol-Myers Squibb, CeNeRx BioPharma, Cephalon, Clintara, Cerecor, Covance, Covidien, Eli Lilly, EnVivo Pharmaceuticals, Euthymics Bioscience, Forest Pharmaceuticals, Ganeden Biotech, GlaxoSmithKline, Harvard Clinical Research Institute, Hoffman-LaRoche, Icon Clinical Research, i3 Innovus/Ingenix, Janssen R&D, Jed Foundation, Johnson & Johnson Pharmaceutical Research and Development, Lichtwer Pharma GmbH, Lorex Pharmaceuticals, Lundbeck, MedAvante, Methylation Sciences, NARSAD, National Center for Complementary and Alternative Medicine, Neuralstem, NIDA, NIMH, Novartis, Organon Pharmaceuticals, Pamlab, Pfizer, Pharmacia-Upjohn, Pharmaceutical Research Associates, Pharmavite, PharmoRx Therapeutics, Photothera, Reckitt

Benckiser, Roche Pharmaceuticals, RCT Logic (formerly Clinical Trials Solutions), Sanofi-Aventis US, Shire, Solvay Pharmaceuticals, Stanley Medical Research Institute, Synthelabo, TalMedical, and Wyeth-Ayerst Laboratories; he has served as adviser or consultant to Abbott Laboratories, Acadia, Affectis Pharmaceuticals, Alkermes, Amarin Pharma, Aspect Medical Systems, AstraZeneca, Auspex Pharmaceuticals, Avanir Pharmaceuticals, AXSOME Therapeutics, Bayer, Best Practice Project Management, Biogen, BioMarin Pharmaceuticals, Biovail Corporation, BrainCells, Bristol-Myers Squibb, CeNeRx BioPharma, Cephalon, Cerecor, CNS Response, Compellis Pharmaceuticals, Cypress Pharmaceutical, DiagnoSearch Life Sciences, Dainippon Sumitomo Pharma, Dov Pharmaceuticals, Edgemont Pharmaceuticals, Eisai, Eli Lilly, EnVivo Pharmaceuticals, ePharmaSolutions, EPIX Pharmaceuticals, Euthymics Bioscience, Fabre-Kramer Pharmaceuticals, Forest Pharmaceuticals, Forum Pharmaceuticals, GenOmind, GlaxoSmithKline, Grunenthal GmbH, i3 Innovus/Ingenis, Intracellular, Janssen Pharmaceutica, Jazz Pharmaceuticals, Johnson & Johnson Pharmaceutical Research and Development, Knoll Pharmaceuticals, Labopharm, Lorex Pharmaceuticals, Lundbeck, MedAvante, Merck, MSI Methylation Sciences, Naurex, Nestle Health Sciences, Neuralstem, Neuronetics, NextWave Pharmaceuticals, Novartis, Nutrition 21, Orexigen Therapeutics, Organon Pharmaceuticals, Osmotica, Otsuka Pharmaceuticals, Pamlab, Pfizer, PharmaStar, Pharmavite, PharmoRx Therapeutics, Precision Human Biolaboratory, Prexa Pharmaceuticals, Puretech Ventures, PsychoGenics, Psylin Neurosciences, RCT Logic (Formerly Clinical Trials Solutions), Rexahn Pharmaceuticals, Ridge Diagnostics, Roche, Sanofi-Aventis US, Sepracor, Servier Laboratories, Schering-Plough, Solvay Pharmaceuticals, Somaxon Pharmaceuticals, Somerset Pharmaceuticals, Sunovion Pharmaceuticals, Supernus Pharmaceuticals, Synthelabo, Taisho Pharmaceutical, Takeda Pharmaceutical Company, Tal Medical, Tetragenex Pharmaceuticals, TransForm Pharmaceuticals, Transcept Pharmaceuticals, Vanda Pharmaceuticals, and VistaGen; he has received speaking or publishing fees from Adamed, Advanced Meeting Partners, American Psychiatric Association, American Society of Clinical Psychopharmacology, AstraZeneca, Belvoir Media Group, Boehringer Ingelheim GmbH, Bristol-Myers Squibb, Cephalon, CME Institute/Physicians Postgraduate Press, Eli Lilly, Forest Pharmaceuticals, GlaxoSmithKline, Imedex, MGH Psychiatry Academy/Primedia, MGH Psychiatry Academy/Reed Elsevier, Novartis, Organon Pharmaceuticals, Pfizer, PharmaStar, United BioSource, and Wyeth-Ayerst Laboratories; he has equity holdings in Compellis and PsyBrain; he has a patent for Sequential Parallel Comparison Design, which are licensed by Massachusetts General Hospital to Pharmaceutical Product Development, and a patent application for a combination of ketamine plus scopolamine in major depressive disorder, licensed by Massachusetts General Hospital to Biohaven; and he receives royalties for the MGH Cognitive and Physical Functioning Questionnaire, the Sexual Functioning Inventory, the Antidepressant Treatment Response Questionnaire, Discontinuation-Emergent Signs and Symptoms, the Symptoms of Depression Questionnaire, and SAFER, and from Lippincott, Williams & Wilkins, Wolkers Kluwer, and World Scientific Publishing. P.J.M. has received funding from the Forest Research Laboratories, Naurex Pharmaceuticals (now Allergan), the New York State Department of Mental Hygiene, NIMH, the Research Foundation for Mental Hygiene (New York State), and Sunovion Pharmaceuticals. M.W. has received funding from NARSAD, NIDA, NIMH, the Sackler Foundation, and the Templeton Foundation; and she receives royalties from American Psychiatric Association Press, Multi- Health Systems, Oxford University Press, and Perseus Press. M.L.P. has received funding from NIMH, the Brain and Behavior Research Foundation, the Pittsburgh Foundation, a donation from a private donor, and an honorarium from Sunovion. M.H.T. has served as an adviser or consultant for Abbott Laboratories, Abdi Ibrahim, Akzo (Organon Pharmaceuticals), Alkermes, AstraZeneca, Axon Advisors, Bristol- Myers Squibb, Cephalon, Cerecor, CME Institute of Physicians, Concert Pharmaceuticals, Eli Lilly, Evotec, Fabre Kramer Pharmaceuticals, Forest Pharmaceuticals, GlaxoSmithKline, Janssen Global Services, Janssen Pharmaceutica Products, Johnson & Johnson PRD, Libby, Lundbeck, Meade Johnson, MedAvante, Medtronic, Merck, Mitsubishi Tanabe Pharma Development America, Naurex, Neuronetics, Otsuka Pharmaceuticals, Pamlab, Parke-Davis Pharmaceuticals, Pfizer, PgxHealth, Phoenix Marketing Solutions, Rexahn Pharmaceuticals, Ridge Diagnostics, Roche Products, Sepracor, Shire Development, Sierra, SK Life and Science, Sunovion, Takeda, Tal Medical/Puretech Venture, Targacept, Transcept, VantagePoint, Vivus, and Wyeth-Ayerst Laboratories; he has received grants or research support from the Agency for Healthcare Research and Quality, Cyberonics, NARSAD, NIDA, and NIMH.

All other authors reported no biomedical financial interests or potential conflicts of interests

REFERENCES

1. Dupuy JM, Ostacher MJ, Huffman J, Perlis RH, Nierenberg AA (2011): A critical review of pharmacotherapy for major depressive disorder. *The international journal of neuropsychopharmacology* 14: 1417–1431.
2. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D*, et al.* (2006): Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *American Journal of Psychiatry* 163: 1905–1917.
3. Chin Fatt CR, Jha MK, Cooper CM, Fonzo G, South C, Grannemann B*, et al.* (2019): Effect of Intrinsic Patterns of Functional Brain Connectivity in Moderating Antidepressant Treatment Response in Major Depression. *American Journal of Psychiatry* 177: 143-154.
4. Davidson RJ, Irwin W, Anderle MJ, Kalin NH (2003): The neural substrates of affective processing in depressed patients treated with venlafaxine. *American Journal of Psychiatry* 160: 64–75.
5. Chen C-H, Ridler K, Suckling J, Williams S, Fu CHY, Merlo-Pich E*, et al.* (2007): Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biological psychiatry* 62: 407–414.
6. Canli T, Cooney RE, Goldin P, Shah M, Sivers H, Thomason ME*, et al.* (2005): Amygdala reactivity to emotional faces predicts improvement in major depression. *Neuroreport* 16: 1267–1270.
7. Greenberg T, Fournier JC, Stiffler R, Chase HW, Almeida JR, Aslam H*, et al.* (2019): Reward related ventral striatal activity and differential response to sertraline versus placebo in depressed individuals. *Molecular Psychiatry* 25: 1526–1536.
8. Cao H, Schwarz E (2019): Opportunities and challenges of machine learning approaches for biomarker signature identification in psychiatry. In: Baune BT, Baune Be. *Personalized psychiatry*. Amsterdam: Academic Press, pp 117–126.
9. Kautzky A, Lanzenberger R, Kasper S (2019): Big Data Guided Interventions: Predicting Treatment Response. In: Passos IC, Mwangi B, Kapczinski F. *Personalized psychiatry. Big data analytics in mental health*. Cham, Switzerland: Springer, pp 53–76.
10. Nguyen KP, Chin Fatt C, Treacher A, Mellema C, Trivedi MH, Montillo A (2019): Predicting Response to the Antidepressant Bupropion Using Pretreatment fMRI. *Medical image computing and computer-assisted intervention: PRIME:* 53–62.
11. Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML*, et al.* (2016): Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *Journal of psychiatric research* 78: 11–23.
12. Greenberg T, Chase HW, Almeida JR, Stiffler R, Zevallos CR, Aslam HA*, et al.* (2015): Moderation of the Relationship Between Reward Expectancy and Prediction Error-Related Ventral Striatal Reactivity by Anhedonia in Unmedicated Major Depressive Disorder: Findings From the EMBARC Study. *American Journal of Psychiatry* 172: 881–891.
13. Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000): Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum. *J Neurophysiology* 84: 3072–3077.
14. Shorten C, Khoshgoftaar TM (2019): A survey on Image Data Augmentation for Deep Learning. *J Big Data* 6.
15. Nguyen KP, Chin Fatt C, Treacher A, Mellema C, Trivedi MH, Montillo A (2020): Anatomically-informed data augmentation for functional MRI with applications to deep learning. In: Landman BA, Išgum I. *Medical Imaging: Image Processing*: SPIE, pp 28–33.

16. Nguyen KP, Raval V, Minhajuddin A, Carmody T, Trivedi MH, Dewey RB*, et al.* (2021): The BLENDS Method for Data Augmentation of 4-Dimensional Brain Images. *BiorXiv:* 10.1101/2021.06.02.446748.
17. Varma S, Simon R (2006): Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics* 7: 91.
18. Vabalas A, Gowen E, Poliakoff E, Casson AJ (2019): Machine learning algorithm validation with a limited sample size. *PloS one* 14: e0224365.
19. Ojala M, Garriga GC (2010): Permutation Tests for Studying Classifier Performance. *Journal of Machine Learning Research* 11: 1833–1863.
20. Dimopoulos Y, Bourret P, Lek S (1995): Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Process Lett* 2: 1–4.
21. Olden JD, Joy MK, Death RG (2004): An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* 178: 389–397.
22. Roose SP, Rutherford BR, Wall MM, Thase ME (2016): Practising evidence-based medicine in an era of high placebo response: number needed to treat reconsidered. *The British journal of psychiatry : the journal of mental science* 208: 416–420.
23. Etkin A, Patenaude B, Song YJC, Usherwood T, Rekshan W, Schatzberg AF*, et al.* (2015): A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial. *Neuropsychopharmacology* 40: 1332–1342.
24. Gordon E, Rush AJ, Palmer DM, Braund TA, Rekshan W (2015): Toward an online cognitive and emotional battery to predict treatment remission in depression. *Neuropsychiatric disease and treatment* 11: 517–531.
25. Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS*, et al.* (2017): Functional Connectivity of the Subcallosal Cingulate Cortex And Differential Outcomes to Treatment With Cognitive-Behavioral Therapy or Antidepressant Medication for Major Depressive Disorder. *American Journal of Psychiatry* 174: 533–545.
26. Fonzo G, Etkin A, Zhang Y, Wu W, Cooper C, Chin-Fatt C*, et al.* (2019): Brain Regulation of Emotional Conflict Differentiates Response to Antidepressants Versus Placebo in Depression. *Nature Human Behaviour* 3: 1319--1331.
27. Wu W, Zhang Y, Jiang J, Lucas MV, Fonzo GA, Rolle CE*, et al.* (2020): Antidepressant-Responsive Brain Signature in Major Depression Defined by Electroencephalography. *Nature Biotechnology* 38: 439–447.
28. Pizzagalli DA, Webb CA, Dillon DG, Tenke CE, Kayser J, Goer F*, et al.* (2018): Pretreatment Rostral Anterior Cingulate Cortex Theta Activity in Relation to Symptom Improvement in Depression: A Randomized Clinical Trial. *JAMA Psychiatry* 75: 547–554.
29. Phillips ML, Chase HW, Sheline YI, Etkin A, Almeida JRC, Deckersbach T*, et al.* (2015): Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches. *American Journal of Psychiatry* 172: 124–138.
30. Forbes EE, Christopher May J, Siegle GJ, Ladouceur CD, Ryan ND, Carter CS*, et al.* (2006): Reward-related decision-making in pediatric major depressive disorder: an fMRI study. *Journal of child psychology and psychiatry, and allied disciplines* 47: 1031–1040.
31. Forbes EE, Olino TM, Ryan ND, Birmaher B, Axelson D, Moyles DL*, et al.* (2010): Reward-related brain function as a predictor of treatment response in adolescents with major depressive disorder. *Cognitive, affective & behavioral neuroscience* 10: 107–118.

32. Lener MS, Iosifescu DV (2015): In pursuit of neuroimaging biomarkers to guide treatment selection in major depressive disorder: a review of the literature. *Annals of the New York Academy of Sciences* 1344: 50–65.
33. Doya K (2000): Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology* 10: 732–739.
34. Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004): Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature neuroscience* 7: 887–893.
35. Tepfer LJ, Alloy LB, Smith DV (2021): Family history of depression is associated with alterations in task-dependent connectivity between the cerebellum and ventromedial prefrontal cortex. *Depress. Anxiety* 38: 508–520.
36. Cheng W, Rolls ET, Qiu J, Yang D, Ruan H, Wei D*, et al.* (2018): Functional Connectivity of the Precuneus in Unmedicated Patients With Depression. *Biological psychiatry. Cognitive neuroscience and neuroimaging* 3: 1040–1049.
37. Rolls ET, Cheng W, Du J, Wei D, Qiu J, Dai D*, et al.* (2020): Functional connectivity of the right inferior frontal gyrus and orbitofrontal cortex in depression. *Social cognitive and affective neuroscience* 15: 75–86.
38. Gabbay V, Ely BA, Li Q, Bangaru SD, Panzer AM, Alonso CM*, et al.* (2013): Striatum-based circuitry of adolescent depression and anhedonia. *Journal of the American Academy of Child and Adolescent Psychiatry* 52: 628-41.e13.
39. Campbell S, MacQueen G (2004): The role of the hippocampus in the pathophysiology of major depression. *Journal of Psychiatry and Neuroscience* 29: 417–426.
40. Sapolsky RM (2001): Depression, antidepressants, and the shrinking hippocampus. *Proceedings of the National Academy of Sciences of the United States of America* 98: 12320–12322.
41. Robertson B, Wang L, Diaz MT, Aiello M, Gersing K, Beyer J*, et al.* (2007): Effect of Bupropion Extended Release on Negative Emotion Processing in Major Depressive Disorder. *J. Clin. Psychiatry* 68: 261–267.

FIGURE LEGENDS

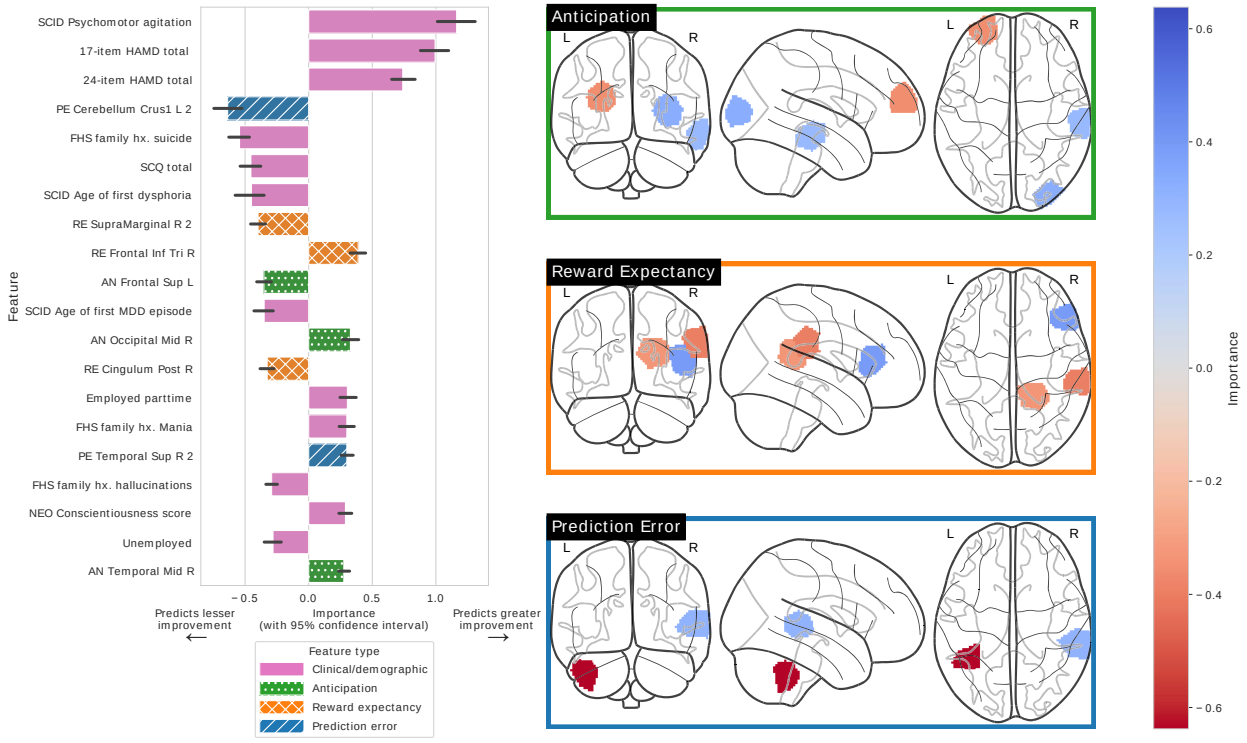

Figure 1. The twenty most important features learned by the sertraline outcome prediction model. Importance was measured as the partial derivative of the model prediction output with respect to the feature. *Left)* Features are ranked by descending importance and colored by feature type. The mean importance and 95% confidence interval over the 20 outer cross-validation folds are shown. *Right)* Imaging features are visualized as colored ROIs in the study-specific brain atlas and overlaid on a “glass brain” view of the MNI brain template (see online version for color figure). Abbreviations: HAMD – Hamilton Rating Scale for Depression, AN – anticipation, RE – reward expectancy, PE – prediction error, L – left, R – right. See Table S2 for abbreviations for clinical measures.

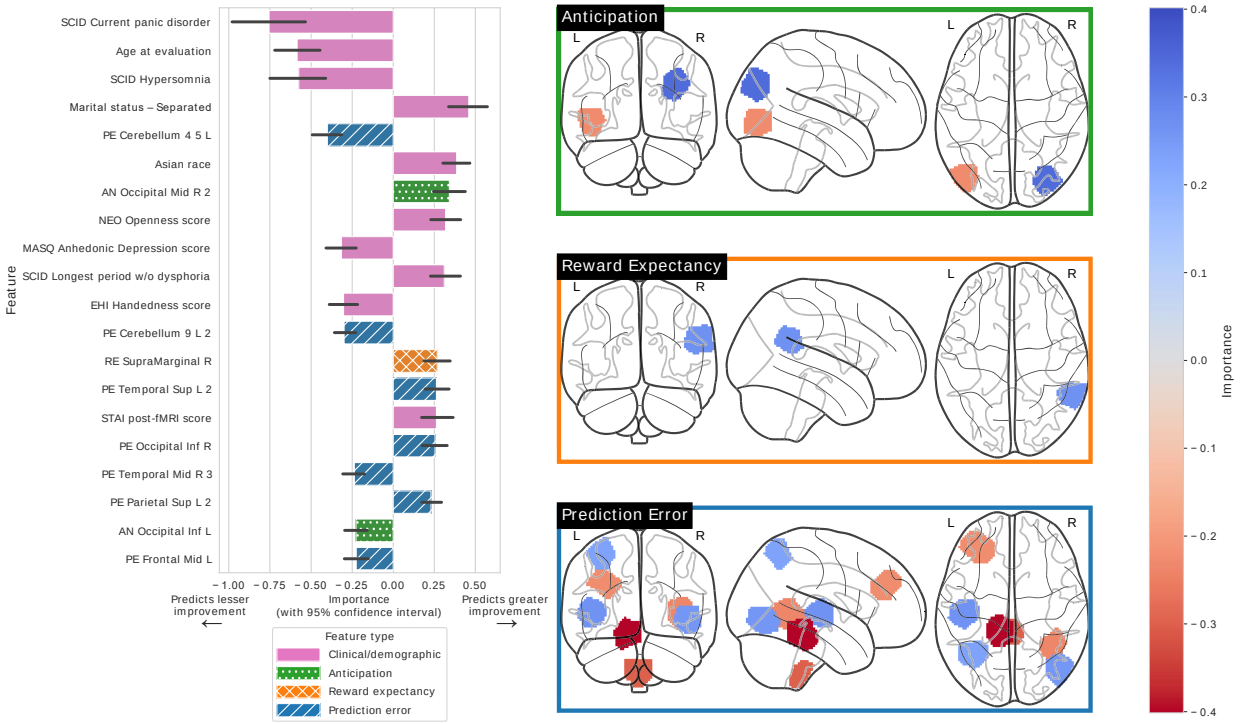

Figure 2. The twenty most important features learned by the placebo outcome prediction model. Importance was measured as the partial derivative of the model prediction output with respect to the feature. *Left)* Features are ranked by descending importance and colored by feature type. The mean importance and 95% confidence interval over the 20 outer cross-validation folds are shown. *Right)* Imaging features are visualized as colored ROIs in the study-specific brain atlas and overlaid on a “glass brain” view of the MNI brain template (see online version for color figure).. See Figure 1 for abbreviation definitions.

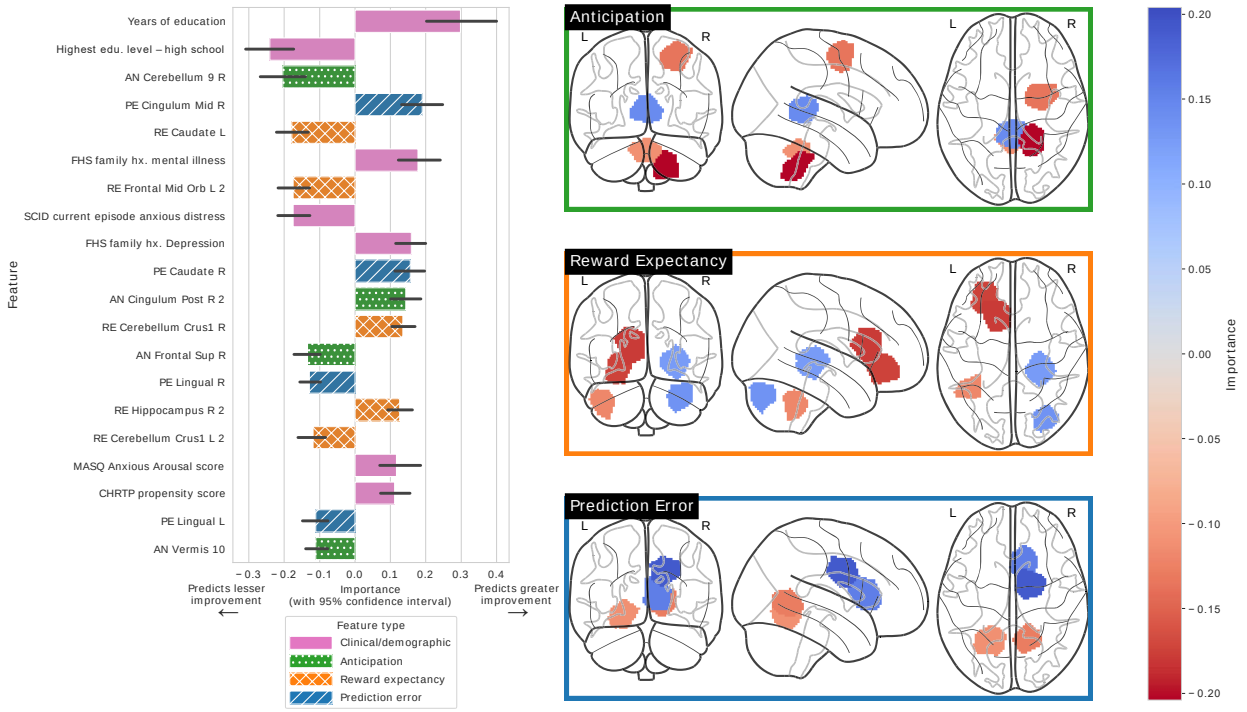

Figure 3. The twenty most important features learned by the bupropion outcome prediction model. Importance was measured as the partial derivative of the model prediction output with respect to the feature. *Left)* Features are ranked by descending importance and colored by feature type. The mean importance and 95% confidence interval over the 20 outer cross-validation folds are shown. *Right)* Imaging features are visualized as colored ROIs in the study-specific brain atlas and overlaid on a “glass brain” view of the MNI brain template (see online version for color figure).. See Figure 1 for abbreviation definitions.

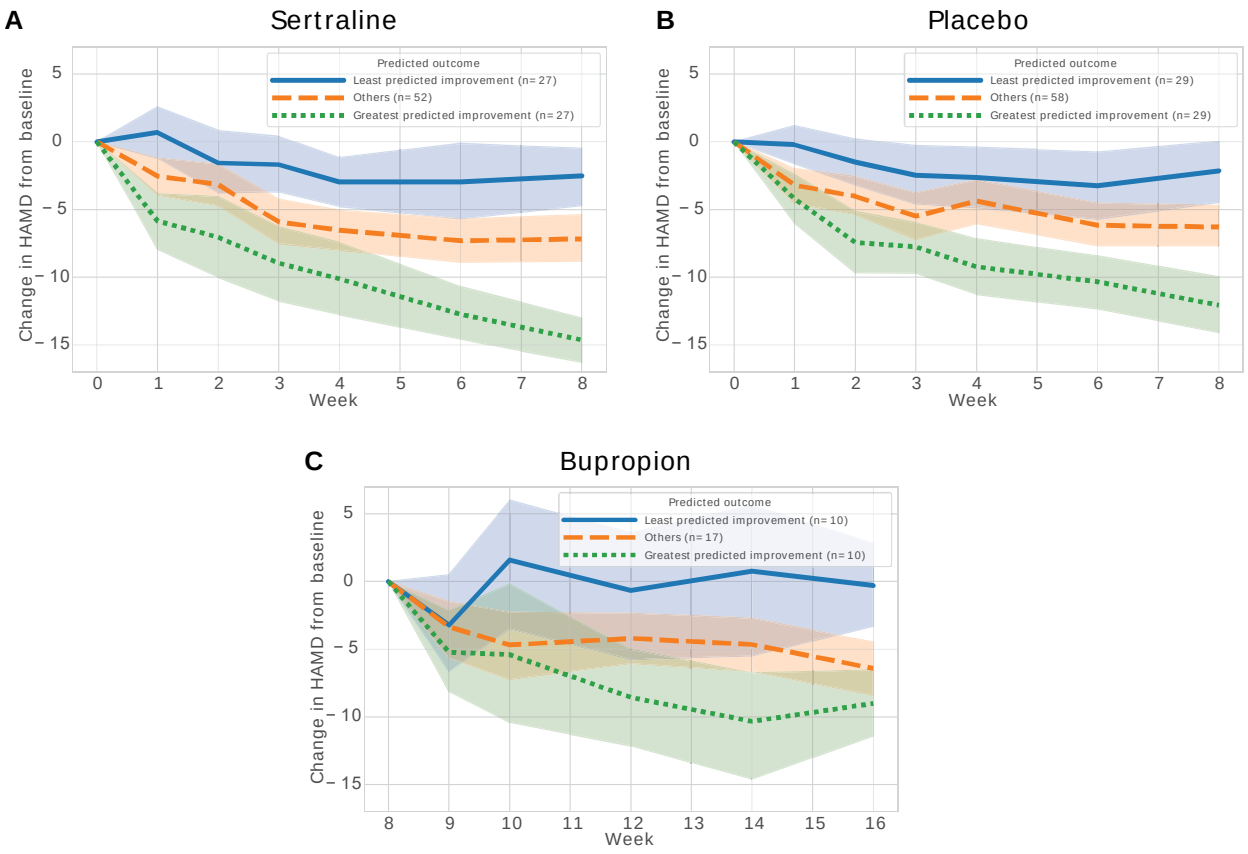

Figure 4. Association between predicted treatment outcomes from baseline imaging measures and longitudinal trajectories of clinical improvement. For each treatment group—a) sertraline, b) placebo, and c) bupropion—participants were ranked by predicted treatment outcome. Participants were then grouped into the 25% with the greatest predicted improvement, the 25% with least predicted improvement, and the remaining 50% (“others”). The trajectories of clinical improvement, measured as the mean change in HAMD score from baseline, over the 8-week treatment period are shown for each group. Steeper slopes indicate a faster rate of symptomatic improvement. The 95% confidence interval is shown as a shaded area around each line. See Figure S6 in the Supplement for plots of absolute HAMD score (not normalized to baseline).

TABLES

Table 1. Demographics, pre-treatment clinical characteristics, and 8-week treatment outcomes for sertraline, placebo, and bupropion treatment groups.

	Sertraline		Placebo		Bupropion	
Participants completing 8-week treatment	126		114		41	
Total participants analyzed	106		116		37	
Demographics						
Female	73	69%	73	63%	26	70%
Race						
White	72	68%	83	72%	24	65%
African American	20	19%	17	15%	8	22%
Asian	5	5%	8	7%	3	8%
Other	9	8%	8	6%	2	5%
Hispanic	19	18%	22	19%	6	16%
Employed	61	58%	69	59%	20	54%
Age	38.38 ± 13.95		37.40 ± 12.80		37.51 ± 14.32	
Clinical characteristics						
Age of first major depressive episode	16.16 ± 5.86		16.57 ± 5.91		16.11 ± 5.77	
Pre-treatment HAMD	18.60 ± 4.45		18.57 ± 4.26		18.00 ± 3.96	
Treatment dose at week 8	139 ± 26 mg		n/a		377 ± 115 mg	
Treatment outcomes						
ΔHAMD (Week 8 – pre-treatment)	7.89 ± 7.16		6.70 ± 6.93		5.46 ± 5.57	

Table 2. Outcome prediction performance for the three treatments investigated. Deep learning models were trained to predict 8-week ΔHAMD. Performance metrics for this target include the coefficient of determination (R^2) and root mean squared error (RMSE). To obtain predictions of remission and response, which are binary variables, model outputs were thresholded post-hoc using the HAMD criteria for remission (HAMD ≤ 7 at week 8) and response (decrease in HAMD ≥ 50%). Performance metrics for remission and response are number-needed-to-treat (NNT), positive predictive value (PPV) and area under the receiver operating characteristic curve (AUROC). Statistical significance of these performance measurements over chance accuracy was measured using permutation testing, and the *p*-values are presented here.

	Prediction target							
	ΔHAMD		**Remission**			**Response**		
Treatment	R^2	RMSE	NNT	PPV	AUROC	NNT	PPV	AUROC
Sertraline	48%	5.15	3.33	0.69	0.60	4.86	0.68	0.62
	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p=0.14$	$p < 0.01$	$p < 0.01$	$p < 0.01$
Placebo	28%	5.87	2.06	0.81	0.65	2.95	0.69	0.67
	$p < 0.01$	$p < 0.01$	$p = 0.02$	$p = 0.02$	$p = 0.06$	$p = 0.02$	$p = 0.02$	$p < 0.01$
Bupropion	34%	4.46	2.35	0.75	0.71	1.68	1.00	0.57
	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p = 0.14$

Table 3. Brain regions identified by the models as containing reward task activation features predictive of treatment outcome. The regions are categorized by neurophysiological roles in MDD as described in previous research. (↑) indicates that higher activation in the region was associated with greater predicted symptomatic (HAMD score) improvement, (↓) indicates association with lesser predicted improvement, and (↕) indicates that the directionality of the association varied among contrasts. Superscripts indicate which reward task contrast feature was learned by the model in a particular region; *AN*: anticipation, *RE*: reward expectancy, *PE*: prediction error.

Reported neurophysiological role in MDD	Sertraline	Placebo	Bupropion
Reward processing activity altered in MDD[29-32]	↓Prefrontal cortex^{AN}	↓Prefrontal cortex^{PE}	↑Middle cingulate cortex^{PE} ↕Caudate^{PE,RE} ↓Orbitofrontal cortex^{RE}
Associated with reward processing, abnormal functional connectivity in MDD[33-35]	↓Cerebellum Crus 1^{PE}	↓Cerebellum Crus 1^{PE}	↕Cerebellum Crus 1RE ↓Cerebellum^{AN}
Associated with abnormal functional connectivity in MDD[36-38]	↓Supramarginal gyrus^{RE} ↑Inferior frontal gyrus, pars triangularis^{RE}	↑Supramarginal gyrus^{RE}	
Associated with mood dysregulation in MDD[29,39,40]			↑Hippocampus^{RE}
Emotion processing activation modulated by bupropion[41]			↑PCC^{AN}
Other regions identified by the predictive model without reported roles in MDD	↑Middle occipital gyrus^{AN} ↑Middle temporal gyrus^{AN} ↑Superior temporal gyrus^{PE}	↑Middle occipital gyrus^{AN} ↕Inferior occipital gyrus^{PE,AN} ↓Middle temporal gyrus^{PE} ↑Superior temporal gyrus^{PE} ↑Superior parietal lobule^{PE}	