Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Machine Learning Prediction of Quality of Life Improvement During Antidepressant Treatment of Patients With Major Depressive Disorder:

A STAR*D and CAN-BIND-1 Report

Tejas Phaterpekar, MDS; John-Jose Nunez, MD, MSc; Emma Morton, PhD; Yang S. Liu, PhD; Bo Cao, PhD; Benicio N. Frey, MD, PhD; Roumen V. Milev, MD, PhD; Daniel J. Müller, MD, PhD; Susan Rotzinger, PhD; Claudio N. Soares, MD, PhD; Valerie H. Taylor, MD, PhD; Rudolf Uher, MD, PhD; Sidney H. Kennedy, MD; and Raymond W. Lam, MD

Abstract

Background: Quality of life (QoL) is an important patient-centric outcome to evaluate in treatment of major depressive disorder (MDD). This work sought to investigate the performance of several machine learning methods to predict a return to normative QoL in patients with MDD after antidepressant treatment.

Methods: Several binary classification algorithms were trained on data from the first 2 weeks of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (n = 651, conducted from 2001 to 2006) to predict week 9 normative QoL (score ≥ 67, based on a community normative sample, on the Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form [Q-LES-Q-SF]) after treatment with citalopram. Internal validation was performed using a STAR*D holdout dataset, and external validation was performed using the Canadian Biomarker Integration Network in Depression-1 (CAN-BIND-1) dataset (n = 175, study conducted from 2012 to 2017) after treatment with escitalopram. Feature importance was calculated using SHapley Additive exPlanations (SHAP).

Results: Random Forest performed most consistently on internal and external validation, with balanced accuracy (area under the receiver operator curve) of 71% (0.81) on the STAR*D dataset and 69% (0.75) on the CAN-BIND-1 dataset. Random Forest Classifiers trained on Q-LES-Q-SF and Quick Inventory of Depressive Symptomatology–Self-Rated variables had similar performance on both internal and external validation. Important predictive variables came from psychological, physical, and socioeconomic domains.

Conclusions: Machine learning can predict normative QoL after antidepressant treatment with similar performance to that of prior work predicting depressive symptom response and remission. These results suggest that QoL outcomes in MDD patients can be predicted with simple patient-rated measures and provide a foundation to further improve performance and demonstrate clinical utility.

Trial Registration: ClinicalTrials.gov identifiers NCT00021528 and NCT01655706

J Clin Psychiatry 2024;85(1):23m14864

Author affiliations are listed at the end of this article.

Major depressive disorder (MDD) is not only a common mental illness—affecting up to 1 in 5 people in their lifetime—but also one that has been shown to negatively impact all domains of well-being and functioning.[1–3] Individuals with MDD also have demonstrated more impairment in quality of life (QoL) than those with chronic medical conditions like diabetes and hypertension.[4] QoL is an important outcome to consider in treatment of MDD because patients generally value QoL over symptom relief.

Several studies have focused on the relationship of QoL and depressive symptom improvement in MDD. While symptom severity and QoL are generally negatively correlated, studies show only a partial association between these outcomes.[5,6] For example, QoL typically improves with symptomatic reduction following treatments for MDD,

Scan Now See supplementary material for this article at Psychiatrist.com

Editor's Note
We encourage authors to submit papers for consideration as a part of our Early Career Psychiatrists section. Please contact Joseph F. Goldberg, MD, at jgoldberg@psychiatrist.com.

Clinical Points

- Quality of life serves as an important patient-centric outcome. So far, no one has investigated using machine learning to predict whether quality of life will improve with antidepressant treatment.
- This work shows that quality of life improvement with antidepressant treatment can be predicted using early depression symptoms and quality of life impairment.
- With further research and development of machine learning models that predict patient-centric outcomes, clinicians will have access to individualized tools that can supplement treatment decision-making.

but approximately half of patients who reach symptom remission still have lower QoL than community norms.[7,8] Other studies found both QoL and functioning impairments to be significant predictors of depressive relapse.[5,9] Therefore, both symptom and QoL outcomes are important for a full assessment of treatment response and recovery.[10]

Although many effective, evidence-based treatments for MDD are available, clinically useful predictors of treatment outcome have yet to be identified. Hence, treatment selection is still primarily a trial-and-error process that often takes weeks to months before an effective treatment is found.[11] Machine learning offers the potential for a more personalized approach by using an individual's unique attributes to predict treatment outcomes and guide clinician decision-making.

Foundationally, supervised machine learning involves training a statistical model on a labeled dataset. Throughout the training process, the models learn relationships between the features and a chosen target variable and can then make predictions on new, unseen data (test data). To date, work predicting the outcomes of treating MDD using machine learning has focused primarily on predicting remission, response, and treatment resistance, as assessed by symptom scales. We are not aware of studies using machine learning to predict treatment outcomes that extend beyond depressive symptoms, such as QoL, well-being, or functioning. Additionally, machine learning prediction studies in MDD used small datasets and were not externally validated.[12,13] Despite its limited use in the literature, external validation is important to establish generalizability, the ability of a model to perform consistently on new, independent data.[14] This will be necessary before models can be used clinically.

In this study, we investigated the performance of binary classification algorithms in predicting QoL outcomes following antidepressant treatment in patients with MDD using a dataset from the Sequenced Treatment Alternative to Relieve Depression (STAR*D) study,[15] and we describe important features of the best performing predictive models. We also conducted an external validation of the models on a dataset from the Canadian Biomarker Integration Network in Depression-1 study (CAN-BIND-1).[16] We hypothesized that machine learning models for QoL prediction would be able to perform with accuracy similar to that of prior work predicting depressive symptom treatment outcomes.

Table 1.

A Comparison of Scales Utilized in Creating Overlapping Features From the STAR*D and CAN-BIND-1 Datasets

Variable	STAR*D Source	CAN-BIND-1 Source
Demographics	• Demographics form	• Demographics form
Functional Impairment	• Q-LES-Q-SF • WSAS • WPAI	• Q-LES-Q-SF • SDS • LEAPS
Psychiatric History	• Psychiatric history	• MINI
Depressive Symptoms	• QIDS-SR	• QIDS-SR

Abbreviations: CAN-BIND-1 = Canadian Biomarker Integration Network in Depression-1, LEAPS = Lam Employment Absence and Productivity Scale, MINI = Mini-International Neuropsychiatric Interview, Q-LES-Q-SF = Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self Report, SDS = Sheehan Disability Scale, STAR*D = Sequenced Treatment Alternatives to Relieve Depression, WPAI = Work Productivity and Activity Impairment, WSAS = Work and Social Adjustment Scale.

METHODS

Datasets

STAR*D dataset. The STAR*D study (NCT00021528, conducted from 2001 to 2006) utilized several levels of treatment to investigate strategies for MDD patients who did not achieve symptom remission with an initial antidepressant.[15] The study recruited 4,000 adults, aged 18–75 years, who met *DSM-IV* criteria for single or recurrent nonpsychotic MDD (Supplementary Table 1). This study uses data from Level 1, when participants received open-label citalopram 20–60 mg/d. If a patient was deemed unremitted by clinical judgment and use of the 16-item Quick Inventory of Depressive Symptomatology (QIDS) score after 14 weeks, they moved to Level 2 where they were randomized to either antidepressant switch or augmentation.[17] QoL was assessed using scores from the self-report Quality of Life Enjoyment and Satisfaction Questionnaire Short Form (Q-LES-Q-SF), scores for which were collected at the start, midpoint, and exit of each level. We obtained STAR*D data from the National Institute of Mental Health Data Archive (NDA), where it is publicly available. All patients gave written consent for this study.

CAN-BIND-1 dataset. The CAN-BIND-1 study (NCT01655706, conducted from 2012 to 2017) recruited 211 participants from 6 Canadian outpatient centers, aged 18–60 years, who met *DSM-IV-TR* criteria for MDD (Supplementary Table 1).[18,19] We used data from the first 8 weeks, when participants were given open-

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

label escitalopram 10–20 mg/d. Participants received aripiprazole augmentation for a further 8 weeks if they did not achieve remission.[20] A self-reported Q-LES-Q-SF score was collected for each participant at baseline, week 8, and week 16. All participants provided written consent and ethics approval was obtained at each recruitment center. We obtained data from members of the CAN-BIND team, and the dataset is publicly available on Brain-CODE (https://www.braincode.ca/).

Target Generation

The Q-LES-Q-SF is a 14-item self-report instrument which captures enjoyment and satisfaction, with regards to multiple areas of daily roles and functioning.[21] Q-LES-Q-SF scores were recorded in both the STAR*D (week 9) and CAN-BIND-1 (week 8) trials, allowing its use as an end-study target for external validation. All raw total scores on the Q-LES-Q-SF were converted into a percentage of the maximum total score using an established formula.[22] If there were 5 or more missing items from the questionnaire, then these participants were excluded from the study.[23] Prior literature has used normative Q-LES-Q-SF cut-offs from a community sample of 529 people to make comparisons.[24] A normative cutoff was established as 1 standard deviation below the community norm of 78.3% (SD = 11). Using this information, we created a binary classification target, separating impaired (< 67) and non-impaired (≥ 67) QoL.

Data Preprocessing

The STAR*D and CAN-BIND-1 datasets consisted of both sociodemographic and clinical variables: demographics, psychiatric history, medication history, hospital/clinic visits, side effects history, functioning, and diagnostic scales. Features (ie, numerical inputs for model training) were prepared in a pipeline similar to that in a previous study,[25] using the Python 3 programming language. An automated pipeline utilizing common data-handling packages, Pandas and NumPy, was developed to handle data cleaning, aggregation, and processing.[26,27] We included engineered features representing depression subscales and percent changes in depression scores as used in prior work.[25,28] Our final STAR*D dataset consisted of 480 features, derived from the baseline and week 2 visits.

For external validation, only the 100 features that overlapped between the STAR*D and CAN-BIND-1 datasets could be used to train and evaluate the models. This overlapping feature set included the self-rated QIDS (QIDS-SR) and Q-LES-Q-SF scales, demographics, functional measures, and history of concurrent psychiatric disease (summarized in Table 1; full descriptions in Supplementary Appendix 1).

To further understand which features are useful for QoL improvement prediction, we created STAR*D and CAN-BIND-1 datasets with features from only QIDS-SR, from only Q-LES-Q-SF, from both QIDS-SR and Q-LES-Q-SF, and from neither QIDS-SR nor Q-LES-Q-SF (Supplementary Table 2).

Participant Selection

For the STAR*D dataset, participants were included if they had a baseline Q-LES-Q-SF score and a Level 1 Q-LES-Q-SF score between weeks 4 and 9. If there were multiple Level 1 Q-LES-Q-SF scores within the eligible time window, the last outcome carried forward (LOCF) was used. Participants who went into follow-up before week 4 due to intolerance were excluded from the study. Any data after week 9 were also excluded to narrow the focus on the effect of the Level 1 drug, citalopram, and to allow for comparison with the first 8 weeks of CAN-BIND-1. Finally, patients were excluded if their baseline Q-LES-Q-SF scores were already within 1 standard deviation of the community norm (≥ 67).

For the CAN-BIND-1 dataset, participants were included if they had both baseline and week 8 Q-LES-Q-SF values, with baseline scores required to be below 67. Additionally, if a participant was missing 5 or more items from the 14 total items of the Q-LES-Q-SF, then they were also excluded. Supplementary Appendix 2 provides further details on participant selection.

Training and Evaluation of Predictive Models

We used Scikit-learn to train and evaluate models using various classifiers: L2 Logistic Regression (LR), Elastic Net, Random Forest (RF), Gradient-Boosted Decision Tree (GBDT), Support Vector Classifier (SVC), K Nearest Neighbors (KNN), and Dummy Classifier.[29] For each classifier, Scikit-learn's GridSearchCV function was used to identify optimal hyperparameters. Scikit-learn's MinMax scaling was used to normalize the data between 0 and 1. Due to the imbalanced nature of the data, intrinsic class balance options were used for every model except for KNN, which lacked this option. Feature selection reduces the complexity of datasets and potentially improves model performance when using the STAR*D dataset.[30,31] In line with prior work, we used Elastic Net feature selection, using 10-fold cross validation to optimize for l1:l2 ratio. For the full feature set (k = 480) and overlapping feature set (k = 100), for which k is the number of features, independent cross-validated Elastic Net was utilized to isolate 61 and 30 features with nonzero coefficients, respectively.

The STAR*D dataset was randomly split, with stratification of the target variable, into an 80% training set and a 20% holdout set. All training and optimization were done with 10-fold cross validation using only the training set. The optimized models were trained on the STAR*D training data and evaluated on the STAR*D holdout set. Each classifier was assessed for balanced accuracy, accuracy, area under the receiver

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Table 2.

Characteristics of the Cohorts Used in This Study After Participant Selection, From the STAR*D and CAN-BIND-1 Trials[a]

Characteristic	STAR*D (N = 651)	CAN-BIND-1 (N = 175)
Female:male	352:299 (54.1:45.9)	110:65 (62.9:37.1)
Married/domestic partnership	301 (46.2)	49 (28.0)
Never married/divorced/separated/widowed	350 (53.8)	126 (72.0)
Working/student	175 (26.9)	118 (67.4)
Unemployed/disabled/retired	122 (18.7)	42 (24.0)
Any substance use disorder	47 (7.2)	8 (4.6)
Any anxiety disorder	113 (17.4)	87 (49.7)
Years of education, mean (SD)	13.6 (3.0)	13.7 (2.3)
Hours worked over last 2 weeks if employed, mean (SD)*	67.0 (32.6)	55.3 (27.6)
Hours missed from illness over last 2 weeks if employed, mean (SD)*	10.9 (18.3)	5.4 (10.9)
Baseline QIDS-SR total score, mean (SD)	15.8 (4.2)	16.1 (4.1)
Week 2 QIDS-SR total score, mean (SD)	11.3 (4.9)	12.2 (5.0)
Baseline Q-LES-Q-SF total score, mean (SD)*	38.9 (14.0)	35.0 (12.4)

[a]Values are shown as n (%) unless otherwise noted.
*$P < .05$, with Bonferroni correction (n = 6).
Abbreviations: CAN-BIND-1 = Canadian Biomarker Integration Network in Depression-1, Q-LES-Q-SF = Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self Report, STAR*D = Sequenced Treatment Alternatives to Relieve Depression.

operator curve (AUC), sensitivity/recall, specificity, F1 score, precision, positive predictive value, and negative predictive value. For external validation, models were trained on the entire STAR*D dataset (training + holdout) and externally validated on the unseen CAN-BIND-1 dataset, using the same metrics and hyperparameters.

To better interpret which features were positive or negative predictors for the Random Forest models, global feature importance was calculated using SHapley Additive exPlanations (SHAP) values.[32] A representative Random Forest model with similar performance to the 100-run mean was selected, and we inspected beeswarm plots for each of the features with highest mean SHAP values. Features are positive predictors if higher values correspond to the prediction of improving quality of life, and vice versa.

Statistical Analysis

We performed 100 independent evaluations for each classifier and feature set combination, for both internal and external validation. Statistical significance of mean balanced accuracy and mean AUC differences across models was determined using 2-tailed *t* tests and calculated *P* values at the 95% confidence level.

Data Availability

The Python 3 code in this work is available publicly in a GitHub repository, alongside our trained models. The raw and processed STAR*D dataset from this project was obtained from National Institute of Mental Health Data Archive (NDA) (https://nda.nih.gov/edit_collection.html?id=2148). The CAN-BIND-1 dataset is available from Brain-CODE (https://www.braincode.ca/), based at the Ontario Brain Institute. Both datasets have requirements for data stewardship and appropriate use.

Table 3.

Performance of Machine Learning Models When Predicting Normative Quality-Of-Life Using Overlapping Features Found in the STAR*D and CAN-BIND-1 Trials[a]

Model	Mean Balanced Accuracy, % (AUC)	
	STAR*D Holdout Set (k = 100)	CAN-BIND-1 External Validation (k = 100)
Logistic Regression	78% (0.83)	61% (0.69)
Elastic Net Regression	79% (0.83)	61% (0.69)
Random Forest	71% (0.81)	69% (0.75)
SVC	72% (0.79)	62% (0.67)
GBDT	62% (0.75)	63% (0.72)
KNN	64% (0.64)	50% (0.50)

[a]The 100 features (k) found in both CAN-BIND and STAR*D are used for these evaluations. Training and evaluation are repeated 100 times to obtain mean scores.
Abbreviations: AUC = area under the receiver operator curve, CAN-BIND-1 = Canadian Biomarker Integration Network in Depression-1, GBDT = Gradient-Boosted Decision Tree, KNN = K Nearest Neighbors, STAR*D = Sequenced Treatment Alternatives to Relieve Depression, SVC = Support Vector Classifier.

RESULTS

Participant Selection

For the STAR*D dataset, applying selection criteria led to 651 subjects, with 125 (19%) achieving a Q-LES-

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Table 4.
Top 15 Features for Random Forest Models Based on SHAP Values[a]

Rank	Full Features (k = 480)[b]	Overlapping Features (k = 100)[c]
1	QIDS-SR week 2 total (CR)	QIDS-SR week 2 sad mood (SR)
2	QIDS-SR week 2 sad mood (CR)	QIDS-SR week 2 total (SR)
3	QIDS-SR percent change (CR)	Q-LES-Q-SF baseline total
4	IDS-C5	QIDS-SR week 2 suicidality (SR)
5	QIDS-SR week 2 sad mood (SR)	QIDS-SR percent change (SR)
6	QIDS-SR week 2 energy/fatigue (CR)	QIDS-SR week 2 psychomotor slowing (SR)
7	QIDS-SR week 2 total (SR)	QIDS-SR week 2 negative outlook (SR)
8	Short Form Health Survey—physical component	Q-LES-Q-SF baseline housing situation
9	QIDS-SR week 2 suicidality (SR)	QIDS-SR week 2 psychomotor domain (SR)
10	IDS-C5—percent change	Q-LES-Q-SF baseline physical mobility
11	QIDS-SR week 2 psychomotor slowing (SR)	Q-LES-Q-SF baseline sex drive
12	Q-LES-Q-SF baseline score	QIDS-SR week 2 energy/fatigue (SR)
13	QIDS-SR week 2 concentration/decision-making (SR)	Q-LES-Q-SF baseline economic status
14	Clinic Visit week 2—currently on STAR*D meds?	QIDS-SR week 2 early morning insomnia (SR)
15	Short Form Health Survey—pain interfering with normal work?	QIDS-SR week 2 concentration/decision-making (SR)

[a]Feature rankings are based on the absolute SHAP values; we interpreted a representative model's SHAP beeswarm plot to color positive predictors orange and negative predictors blue. Plots are available in Supplementary Figures 2 and 3.
[b]All 480 features extracted from STAR*D.
[c]The 100 overlapping features between STAR*D and data from CAN-BIND-1.
Abbreviations: AUC = area under the receiver operator curve, CAN-BIND-1 = Canadian Biomarker Integration Network in Depression-1, CR = clinician-rated, IDS-C5 = Very Quick Inventory of Depressive Symptomatology, k = number of features, Q-LES-Q-SF = Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self Report, SHAP = SHapley Additive exPlanations, SR = self-rated, STAR*D = Sequenced Treatment Alternatives to Relieve Depression.

Q-SF score within 1 SD of the community norm after citalopram treatment. A sizeable number of STAR*D participants were missing Q-LES-Q-SF data from weeks 4–8, which we verified were not available from the NDA. After application of selection criteria, the CAN-BIND-1 dataset consisted of 175 subjects, with 39 (22%) achieving Q-LES-Q-SF score within 1 SD of the community norm after escitalopram treatment. Table 2 shows characteristics of the cohorts after selection (flowchart shown in Supplementary Figure 1).

Internal and External Validation

We performed internal validation using a STAR*D holdout set, and external validation with the CAN-BIND-1 dataset. Table 3 shows evaluation results on the STAR*D holdout and CAN-BIND-1 external validation sets, after training several models with compatible overlapping features (k = 100), with additional metrics in Supplementary Tables 3 and 4. Random Forest had the most consistent performance on both internal (balanced accuracy: 71%, AUC: 0.81) and external validation (balanced accuracy: 69%, AUC: 0.75) (Supplementary Tables 5–7). Use of either the full set of features (k = 480) or the Elastic Net feature selection did not result in numerical improvements to model performance on either internal or external validation (Supplementary Tables 8 and 9).

Feature Importance

We investigated the feature importance of the most consistently performing model, Random Forests. The 15 most important features in predicting QoL outcome are shown in Table 4. A mixture of QIDS-SR and Q-LES-Q-SF features make up the top 15 features in the overlapping feature set used for external validation. When using the full STAR*D feature set, top features included 3 features not from QIDS-SR or Q-LES-Q-SF: features for physical health, whether participants discontinued their initial STAR*D antidepressant at 2 weeks, and whether pain is interfering at work. Feelings of sadness, suicidality, negative outlook, and psychomotor slowing all ranked within the top 15 important features. Several baseline Q-LES-Q-SF–specific questions were considered important in making future QoL predictions, centered on housing situation, physical mobility, sexual drive, and economic status. Several top-ranking features also reflect early changes in QIDS-SR symptom scores at week 2 of treatment. According to SHAP beeswarm plots (Supplementary Figures 2 and 3), higher baseline

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Table 5.

Performance of Random Forest Models When Predicting Normative Quality-Of-Life Using Different Feature Sets[a]

Feature Set[b]	Mean Balanced Accuracy, % (AUC)	
	STAR*D Internal Holdout	CAN-BIND-1 External Validation
Full STAR*D (k = 480)	72% (0.78)	...
Overlapping (k = 100)	71% (0.81)	69% (0.75)
QIDS-SR only (k = 47)	69% (0.78)	66% (0.73)
Q-LES-Q-SF only (k = 16)	67% (0.77)	56% (0.64)
QIDS-SR + Q-LES-Q-SF (k = 63)	72% (0.81)	68% (0.75)
No QIDS-SR or Q-LES-Q-SF full STAR*D (k = 391)	64% (0.72)	...
No QIDS-SR or Q-LES-Q-SF overlapping (k = 37)	65% (0.71)	55% (0.58)

[a]Training and evaluation are repeated 100 times to obtain mean scores.
[b]*Full STAR*D* refers to all 480 features we extracted from STAR*D. *Overlapping* refers to the 100 overlapping features between STAR*D and CAN-BIND-1.
Abbreviations: AUC = area under the receiver operator curve, CAN-BIND-1 = Canadian Biomarker Integration Network in Depression-1, k = number of features, Q-LES-Q-SF = Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self Report, SR = self-rated, STAR*D = Sequenced Treatment Alternatives to Relieve Depression.

Q-LES-Q-SF values were positive predictors of QoL outcomes, while higher QIDS-SR scores, corresponding to greater symptom severity, were negative predictors.

Specific Clinical Scale Importance

To investigate the relative importance of the QIDS-SR and Q-LES-Q-SF scales, we evaluated predictive performance when using additional datasets containing only these features, or without these features (Table 5). On internal validation, using only QIDS-SR + Q-LES-Q-SF features resulted in a balanced accuracy (72%) similar to that of the overlapping (71%) and full feature (72%) sets. Using QIDS-SR or Q-LES-Q in isolation led to decreased performance, as did using all features other than features from these scales (Supplementary Table 10). Similar patterns in performance are observed with external validation, as well as with other models (Supplementary Tables 11–13).

DISCUSSION

This work investigates several machine learning models in their ability to predict whether a MDD patient undergoing antidepressant therapy would achieve normative QoL, as defined by a Q-LES-Q-SF score within 1 SD of an established community norm.[6,33] Machine learning serves to maximize prediction accuracy on future data, a task that is not attainable with standard regression statistics, and allows these predictions to be made on an individual level.[34] On external validation, the best performing model, Random Forest, predicted normative QoL with a balanced accuracy of 69% and AUC of 0.75. This performance is similar to that found in prior work predicting the effect of antidepressants on depression symptom severity using similar methodology and datasets.[25,28] The best performing model for predicting normative QoL achieved numerically higher performance than the prediction of symptom response (defined as a 50% reduction from baseline in depression symptom rating scales), for which the best models in this prior work achieved balanced accuracies of up to 65% and AUC of 0.70. However, the performance of our models was numerically lower than that from prior work predicting depressive symptom remission (defined as endpoint scores within remission range), for which the best models achieved balanced accuracies of up to 77% and AUC of 0.83. This may suggest that improvement in QoL is more difficult to predict than improvement in depression symptoms, which may stem from the relatively delayed treatment response of QoL compared to depression symptoms.[35] As QoL is a holistic outcome that broadly assesses physical, social, and environmental outcomes, future work may consider predicting different domain-specific subscores for more nuanced predictions.

Similar to prior work, the Random Forest classifiers performed best for external validation, with other models more prone to overfitting.[25] While the STAR*D and CAN-BIND-1 datasets differ on study year, country, medication used (citalopram vs escitalopram), and other demographics, such differences are expected if models such as these are to be used clinically, and so our results suggest some external validity.

Understanding feature importance in machine learning models is important for clinical use[13,36,37] and can be helpful in hypothesis generation and better understanding of the relationships between features and clinical outcomes. We investigated feature importance for our best performing model, Random Forests. For models trained with all STAR*D dataset features, as well as the overlapping features between STAR*D and CAN-BIND-1, most of the top 15 important features were from Q-LES-Q-SF and QIDS-SR items. High scores on week 2 QIDS-SR features were particularly important, suggesting that lack of early depressive symptom improvement can help predict later QoL outcomes. This finding is consistent with literature showing that early symptom improvement at week 2 can also predict depression response and remission.[38,39]

Interestingly, the remaining top 15 model features pertain to Q-LES-Q-SF items (physical mobility, housing situation, sexual drive, economic status) and are reflective of physical, social, and environmental domains. Prior work on the CAN-BIND-1 cohort has identified symptom reduction as a significant predictor for certain QoL domains, particularly physical and psychological.[40] However, approximately half of the patients in the study who reached clinical remission still maintained some

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

degree of QoL impairment.[7] Our Random Forest models did not focus exclusively on symptomatic reduction, but rather included features from several domains pertaining to QoL, suggesting a protective effect of these domains for eventual restoration of normative QoL.

We investigated the importance of QIDS-SR and Q-LES-Q-SF features by evaluating our models using standalone feature sets. Using QIDS-SR or Q-LES-Q-SF feature sets in isolation led to worse performance, but the use of QIDS-SR + Q-LES-Q-SF together (k = 63) led to a performance comparable to that of the models using all available features (k = 480). This finding suggests that both scales elicit non-redundant information that is useful for this prediction. This result aligns well with previous studies that have noted only a moderate correlation between depressive symptoms and QoL.[5,6] Together, these results suggest that the models using QIDS-SR + Q-LES-Q-SF together are sufficient to predict normative QoL or, at least, do not benefit from the additional clinical features we used.

This work lays a foundation for eventually using predictive models in clinical practice to help guide the treatment of depression. Predictive models as investigated in our study may give patients and their clinicians individualized insight into predicted QoL outcomes, in addition to symptom-based outcomes, to help select the most effective medication. With refinement of these predictive models and investigations of clinical utility, these tools may offer a more proactive, individualized approach to treatment decisions.

Limitations

There are several limitations to our study. The sample size of the external validation dataset is relatively small, which can limit generalizability. A stringent threshold for normative QoL was chosen based on community norms, which may not be attainable within an 8-week treatment period, as improvement in QoL may take longer than improvement in depressive symptoms. There are also differences in settings, patient demographics, clinical features, and treatments used (citalopram vs escitalopram) between the STAR*D and CAN-BIND-1 datasets, although these differences may allow for greater generalizability of our models. While our model performance is in line with recent predictive modeling work in depression treatment,[25,28] further performance improvement may be needed for clinical application.

CONCLUSION

Accurate prediction of QoL outcomes with antidepressant treatment is important because QoL has only partial association with depressive symptom outcomes. Machine learning models, especially Random Forest classifiers, predicted normative QoL after 8 weeks of antidepressant therapy with performance similar to that of prior work predicting other depression treatment outcomes, using both internal and external validation sets. Baseline depressive symptoms and QoL features were most important in making these predictions, reinforcing the non-redundancy of depression with QoL for these predictions. In fact, the use of a feature set involving only depressive symptoms and QoL measures had performance similar to that of using all clinical features. Together, these results suggest that the Q-LES-Q-SF and QIDS-SR, when used together, are particularly important for predicting QoL relative to the other clinical features used. Given its importance for patients, QoL should be included with other outcomes in predictive modeling to personalize depression treatment selection.

Article Information

Published Online: November 15, 2023. https://doi.org/10.4088/JCP.23m14864

© 2023 Physicians Postgraduate Press, Inc.

Submitted: March 20, 2023; accepted September 12, 2023.

To Cite: Phaterpekar T, Nunez J-J, Morton E, et al. Machine learning prediction of quality of life improvement during antidepressant treatment of patients with major depressive disorder: a STAR*D and CAN-BIND-1 report. *J Clin Psychiatry.* 2024;85(1):23m14864

Author Affiliations: Department of Psychiatry, University of British Columbia, Vancouver, Canada (Phaterpekar, Nunez, Morton, Lam); Department of Psychiatry, University of Alberta, Edmonton, Canada (Liu, Cao); Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada (Frey); Departments of Psychiatry and Psychology, Queen's University, Kingston, Canada (Milev, Soares); Department of Psychiatry, University of Toronto, Toronto, Canada (Müller, Rotzinger, Kennedy); Centre for Addiction and Mental Health, Toronto, Canada (Müller); Department of Psychiatry, University of Calgary, Calgary, Canada (Taylor); Department of Psychiatry, Dalhousie University, Halifax, Canada (Uher).

Corresponding Author: Raymond W. Lam, MD, University of British Columbia, 2255 Wesbrook Mall, Vancouver, BC V6T 2A1 Canada (r.lam@ubc.ca).

Drs Phaterpekar and Nunez contributed equally to this work.

Relevant Financial Relationships: Dr Milev has received consulting and speaking honoraria from AbbVie, Allergan, Eisai, Janssen, KYE, Lallemand, Lundbeck, Otsuka, and Sunovion and research grants from the Canadian Biomarker Integration Network in Depression (CAN-BIND), the Canadian Institutes of Health Research (CIHR), Janssen, Lallemand, Lundbeck, Nubiyota, Ontario Brain Institute (OBI), and the Ontario Mental Health Foundation (OMHF). **Dr Rotzinger** holds a patent "Teneurin C-Terminal Associated Peptides (TCAP) and methods and uses thereof." **Dr Kennedy** has received research funding or honoraria from the following sources: Abbott, Alkermes, Allergan Abbvie, Brain Canada, CIHR, Janssen, Lundbeck, Lundbeck Institute, OBI, Ontario Research Fund, Otsuka, Pfizer, Servier, Sunovion, and Xian-Janssen and holds stock in Field Trip Health. **Dr Lam** has received honoraria for ad hoc speaking or advising/consulting, or received research funds, from Abbvie, Asia-Pacific Economic Cooperation, Bausch, BC Leading Edge Foundation, Brain Canada, CIHR, Canadian Network for Mood and Anxiety Treatments, CAN-BIND Solutions, Carnot, Grand Challenges Canada, Healthy Minds Canada, Janssen, Lundbeck, Medscape, Michael Smith Foundation for Health Research, Mitacs, Neurotorium, Ontario Brain Institute, Otsuka, Pfizer/Viatris, Shanghai Mental Health Center, Sunnybrook Health Sciences Centre, Unity Health, Vancouver Coastal Health Research Institute, and VGH-UBCH Foundation. **Dr Nunez** received unrestricted research funding from the BC Cancer Foundation, with funds originating from Pfizer. **Drs Phaterpekar, Morton, Liu, Cao, Frey, Müller, Soares, Taylor**, and **Uher** have no disclosures.

Funding/Support: CAN-BIND is an Integrated Discovery Program carried out in partnership with, and financial support from, the Ontario Brain Institute, an independent non-profit corporation, funded partially by the Ontario government. The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred. Additional funding is provided by the Canadian Institutes of Health Research (CIHR), Lundbeck, and Servier. Funding and/or in-kind support is also provided by the investigators' universities and academic institutions. All study medications are independently purchased at wholesale market values. Dr Nunez was partially funded by a UBC Institute of Mental Health Marshall Fellows Award.

Role of the Funders/Sponsors: The organizations that funded this work had no role in design, conduct, analysis, interpretation, writing, editing, or publication of this work.

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Disclaimer: The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred.

Acknowledgments: Data and/or research tools used in the preparation of this manuscript were obtained from the National Institute of Mental Health (NIMH) Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health. Dataset identifier(s): *[NIMH Data Archive Collection ID(s) or NIMH Data Archive Digital Object Identifier (DOI)]*. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDA. CAN-BIND is an Integrated Discovery Program carried out in partnership with, and financial support from, the Ontario Brain Institute, an independent non-profit corporation, funded partially by the Ontario government.

ORCID: Yang S. Liu: https://orcid.org/0000-0003-0406-8056**;** Benicio N. Frey: https://orcid.org/0000-0001-8267-943X**;** Bo Cao: https://orcid.org/0000-0001-9338-3271**;** John-Jose Nunez: https://orcid.org/0000-0002-1602-6382

Supplementary Material: Available at Psychiatrist.com.

References

1. Herrman H, Kieling C, McGorry P, et al. Reducing the global burden of depression: a Lancet-World Psychiatric Association Commission. *Lancet*. 2019;393(10189):e42–e43.
2. Malhi GS, Mann JJ. Depression. *Lancet*. 2018;392(10161):2299–2312.
3. Brenes GA. Anxiety, depression, and quality of life in primary care patients. *Prim Care Companion J Clin Psychiatry*. 2007;9(6):437–443.
4. Wells KB, Sherbourne CD. Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices. *Arch Gen Psychiatry*. 1999;56(10):897–904.
5. IsHak WW, Greenberg JM, Cohen RM. Predicting relapse in major depressive disorder using patient-reported outcomes of depressive symptom severity, functioning, and quality of life in the Individual Burden of Illness Index for Depression (IBI-D). *J Affect Disord*. 2013;151(1):59–65.
6. Schechter D, Endicott J, Nee J. Quality of life of 'normal' controls: association with lifetime history of mental illness. *Psychiatry Res*. 2007;152(1):45–54.
7. Morton E, Bhat V, Giacobbe P, et al; CAN-BIND Investigator Team. Impacts on quality of life with escitalopram monotherapy and aripiprazole augmentation in patients with major depressive disorder: A CAN-BIND report. *Pharmacopsychiatry*. 2021;54(5):225–231.
8. Morton E, Michalak EE, Levitt A, et al. Quality of life impacts of bright light treatment, fluoxetine, and the combination in patients with nonseasonal major depressive disorder: a randomized clinical trial. *Can J Psychiatry*. 2021;66(3):289–297.
9. Solomon DA, Leon AC, Endicott J, et al. Psychosocial impairment and recurrence of major depression. *Compr Psychiatry*. 2004;45(6):423–430.
10. IsHak WW, Greenberg JM, Balayan K, et al. Quality of life: the ultimate outcome measure of interventions in major depressive disorder. *Harv Rev Psychiatry*. 2011;19(5):229–239.
11. Machado-Vieira R, Baumann J, Wheeler-Castillo C, et al. The timing of antidepressant effects: a comparison of diverse pharmacological and somatic treatments. *Pharmaceuticals (Basel)*. 2010;3(1):19–41.
12. Ermers NJ, Hagoort K, Scheepers FE. The predictive validity of machine learning models in the classification and treatment of major depressive disorder: state of the art and future directions. *Front Psychiatry*. 2020;11:472.
13. Sajjadian M, Lam RW, Milev R, et al. Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. *Psychol Med*. 2021;51(16):2742–2751.
14. Cabitza F, Campagner A, Soares F, et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Comput Methods Programs Biomed*. 2021;208:106288.
15. Rush AJ, Fava M, Wisniewski SR, et al; STAR*D Investigators Group. Sequenced Treatment Alternatives to Relieve Depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25(1):119–142.
16. Kennedy SH, Lam RW, Rotzinger S, et al; CAN-BIND Investigator Team. Symptomatic and functional outcomes and early prediction of response to escitalopram monotherapy and sequential adjunctive aripiprazole therapy in patients with major depressive disorder: A CAN-BIND-1 report. *J Clin Psychiatry*. 2019;80(2):18m12202.
17. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573–583.
18. Lam RW, Milev R, Rotzinger S, et al; CAN-BIND Investigator Team. Discovering biomarkers for antidepressant response: protocol from the Canadian biomarker integration network in depression (CAN-BIND) and clinical characteristics of the first patient cohort. *BMC Psychiatry*. 2016;16(1):105.
19. Battle DE. Diagnostic and Statistical Manual of Mental Disorders (*DSM*). *CoDAS*. 2013;25(2):191–192.
20. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–389.
21. Endicott J, Nee J, Harrison W, et al. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. *Psychopharmacol Bull*. 1993;29(2):321–326.
22. IsHak WW, Mirocha J, James D, et al. Quality of life in major depressive disorder before/after multiple steps of treatment and one-year follow-up. *Acta Psychiatr Scand*. 2015;131(1):51–60.
23. Fayers PM, Machin D. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes*. John Wiley & Sons; 2013.
24. IsHak WW, Mirocha J, Pi S, et al. Patient-reported outcomes before and after treatment of major depressive disorder. *Dialogues Clin Neurosci*. 2014;16(2):171–183.
25. Nunez JJ, Nguyen TT, Zhou Y, et al. Replication of machine learning methods to predict treatment outcome with antidepressant medications in patients with major depressive disorder from STAR*D and CAN-BIND-1. *PLoS One*. 2021;16(6):e0253023.
26. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357–362.
27. McKinney W. pandas: a foundational Python library for data analysis and statistics. *Python High Performance Science Computer*. 2011;14(9):1-9.
28. Nie Z, Vairavan S, Narayan VA, et al. Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study. *PLoS One*. 2018;13(6):e0197268.
29. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Machine Learn Res*. 2011;12:2825–2830.
30. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry*. 2016;3(3):243–250.
31. Kautzky A, Baldinger-Melich P, Kranz GS, et al. A new prediction model for evaluating treatment-resistant depression. *J Clin Psychiatry*. 2017;78(2):215–222.
32. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: *Advances in Neural Information Processing Systems*. Vol 30. Curran Associates, Inc; 2017.
33. IsHak WW, Steiner AJ, Klimowicz A, et al. Major depression comorbid with medical conditions: analysis of quality of life, functioning, and depressive symptom severity. *Psychopharmacol Bull*. 2018;48(1):8–25.
34. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233–234.
35. McKnight PE, Kashdan TB. The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. *Clin Psychol Rev*. 2009;29(3):243–259.
36. Cava W, Bauer C, Moore JH, et al. Interpretation of machine learning predictions for patient outcomes in electronic health records. *AMIA Annu Symp Proc*. 2020;2019:572–581.
37. Tonekaboni S, Joshi S, McCradden MD, et al. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. In: Proceedings of the 4th Machine Learning for Healthcare Conference. PMLR; 2019:359-380.
38. Kudlow PA, McIntyre RS, Lam RW. Early switching strategies in antidepressant non-responders: current evidence and future research directions. *CNS Drugs*. 2014;28(7):601–609.
39. Wagner S, Engel A, Engelmann J, et al. Early improvement as a resilience signal predicting later remission to antidepressant treatment in patients with major depressive disorder: systematic review and meta-analysis. *J Psychiatr Res*. 2017;94:96–106.
40. Morton E, Bhat V, Giacobbe P, et al; CAN-BIND Investigator Team. Predictors of quality of life improvement with escitalopram and adjunctive aripiprazole in patients with major depressive disorder: A CAN-BIND study report. *CNS Drugs*. 2021;35(4):439–450.

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

The Journal of Clinical Psychiatry

Supplementary Material

Article Title: Machine Learning Prediction of Quality of Life Improvement During Antidepressant Treatment of Patients With Major Depressive Disorder: A STAR*D and CAN-BIND-1 Report

Author(s): Tejas Phaterpekar, MDS; John-Jose Nunez, MD, MSc; Emma Morton, PhD; Yang S. Liu, PhD; Bo Cao, PhD; Benicio N. Frey, MD, PhD; Roumen V. Milev, MD, PhD; Daniel J. Müller, MD, PhD; Susan Rotzinger, PhD; Claudio N. Soares, MD, PhD; Valerie H. Taylor, MD, PhD; Rudolf Uher, MD, PhD; Sidney H. Kennedy, MD; and Raymond W. Lam, MD

DOI Number: https://doi.org/10.4088/JCP.23m14864

LIST OF SUPPLEMENTARY MATERIAL FOR THE ARTICLE

© Physicians Postgraduate Press, Inc.

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

The Journal of Clinical Psychiatry

13. **Table 13** P-values for Random Forest balanced accuracy, comparing STAR*D internal validation vs CAN-BIND-1 external validation performance, across feature sets

14. **Appendix 1** Description of the overlapping features used in this work

15. **Appendix 2** Participant Selection

16. **Figure 1** A flowchart summary of model training and evaluation, where N represents sample size for each of the steps

17. **Figure 2** SHAP values taken from a randomly sampled Random Forest model, trained on the overlapping feature set (k = 100)

18. **Figure 3** SHAP values taken from a randomly sampled Random Forest model, trained on the full dataset (k= 480)

DISCLAIMER

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

© *Physicians Postgraduate Press, Inc.*

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Supplementary Table 1. Inclusion/exclusion criteria from the STAR*D and CAN-BIND-1 datasets.

Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Cohort	Canadian Biomarker Integration Network in Depression-1 (CAN-BIND-1) Cohort
Inclusion Criteria	
• 18-75 years of age	• 18-60 years of age
• Outpatients	• Outpatients
• HRSD17 score >=14	• MADRS score >= 24
• DSM-IV criteria for single or recurrent nonpsychotic MDD	• DSM-IV-TR criteria for MDE in MDD
• Not currently taking citalopram or have been taking for < 7 days	• No psychotropic medications for at least 5 half-lives before baseline
Exclusion Criteria	
• Pregnant individuals	• Pregnant or breastfeeding individuals
• Psychosis in current or prior MDD episodes	• Psychosis in current episode
• Diagnosis of bipolar disorder	• Diagnosis of bipolar I or II disorder
• History of schizophrenia, schizoaffective disorder, psychosis, anorexia, bulimia, or obsessive compulsive disorder	• Has a different psychiatric diagnosis as the primary diagnosis
• Individuals who required immediate hospitalization for substance detoxification or treatment	• Individuals who have substance abuse/dependence in the past 6 months
• Individuals who have additional medical conditions or use medication that contraindicate any level 1 or 2 treatments.	• Individuals who have uncontrolled medical conditions, or significant neurological disorders/head trauma
• Individuals currently requiring mood stabilizers or antipsychotic drugs	• Individuals with a history of antidepressant-induced hypomania or showing any other risk factors for hypomanic switch on antidepressants
• Individuals who have experienced lack of response or clear intolerability to an adequate trial of an SSRI in their current MDD episode	• Individuals who have failed or had intolerance to a trial of escitalopram or aripiprazole, OR have failed 4 or more pharmacologic interventions.
• Individuals who did not respond to 16 or more cognitive therapy sessions, or 7 or more electroconvulsive therapy sessions, during current episode	• Individuals who have started psychological treatment in the past 3 months leading up to baseline, who intend to continue this modality
	• Individuals with contraindications to magnetic resonance imaging
	• Individuals with high suicidal risk or any significant personality disorder which might interfere with the treatment protocol (decided by clinical judgement)

Abbreviations: HRSD, Hamilton Rating Scale for Depression; MADRS, Montgomery Asberg Depression Rating Scale; MDE, major depressive episode; MDD, major depressive disorder; SSRI, selective serotonin reuptake inhibitor.

Supplementary Table 2. A comparison of feature types included in the various feature sets studied.

Feature Set Name::	Overlapping	Q-LES-Q-SF only	QIDS only	Q-LES-Q-SF + QIDS	No QIDS+ Q-LES-Q-SF
Features:	Demo + WPAI + Psychiatric Hx + Q-LES-Q + QIDS	Q-LES-Q	QIDS	QLES-Q-SF + QIDS	Demo + WPAI + Psychiatric Hx

Abbreviations: Hx, history; Demo: demographics; WPAI, Work Productivity and Impairment Q-LES-Q; Quality of Life Enjoyment and Satisfaction Questionnaire Short Form; QIDS, 16-item Quick Inventory of Depressive Symptomatology.

Supplementary Table 3. STAR*D internal validation (k = 100) results with additional metrics[a].

	Mean Scores for Test and Train								
Model	TP	TN	FP	FN	Sensitivity	Specificity	PPV	NPV	F1 Score
Dummy Classification	4.95 (19.51)	86.06 (338.93)	19.94 (81.07)	20.05 (80.49)	0.2 (0.2)	0.81 (0.81)	0.2 (0.19)	0.81 (0.81)	0.2 (0.19)
Random Forest	14.45 (78.27)	90.24 (350.53)	15.76 (69.47)	10.55 (21.73)	0.58 (0.78)	0.85 (0.83)	0.48 (0.53)	0.9 (0.94)	0.52 (0.63)
Logistic Regression	19 (84)	85 (330)	21 (90)	6 (16)	0.76 (0.84)	0.8 (0.79)	0.48 (0.48)	0.93 (0.95)	0.58 (0.61)
Elastic Net	19.28 (84.81)	85 (328.12)	21 (91.88)	5.72 (15.19)	0.77 (0.85)	0.8 (0.78)	0.48 (0.48)	0.94 (0.96)	0.59(0.61)
KNN	11 (100)	90 (420)	16 (0)	14 (0)	0.44 (1)	0.85 (1)	0.41 (1)	0.87 (1)	0.42 (1)
SVC	17 (90)	80 (339)	26 (81)	8 (10)	0.68 (0.9)	0.75 (0.81)	0.4 (0.53)	0.91 (0.97)	0.5 (0.66)
Gradient Boosting Classifier	7.26 (97.23)	99.72 (420)	6.28 (0)	17.74 (2.77)	0.29 (0.97)	0.94 (1)	0.54 (1)	0.85 (0.99)	0.38 0.99)

[a] The 100 features (k) found in both CAN-BIND and STAR*D are used for these evaluations. Training and evaluation are repeated 100 times to obtain mean scores. All scores pertain to mean values across 100 independent runs of each model. Test scores are shown in nonbrackets, while training scores are shown with brackets. Abbreviations: KNN, K Nearest Neighbors; SVC Support Vector Classifier, GBDT, Gradient Boosting Decision Tree; TP, True Positive; TN, True Negative; FP, False Positive; FN, False Negative; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Supplementary Table 4. CAN-BIND-1 external validation results (k = 100) with additional metrics[a].

	Mean Scores for Test and Train								
model	TP	TN	FP	FN	Sensitivity	Specificity	PPV	NPV	F1 score
Dummy_Classification	7.5 (23.87)	108.77 (424.29)	26.23 (101.71)	31.5 (101.13)	0.19 (0.19)	0.81 (0.81)	0.22 (0.19)	0.78 (0.81)	0.21 (0.19)
Random_Forest	22.11 (97.97)	109.97 (422.48)	25.03 (103.52)	16.89 (27.03)	0.57 (0.78)	0.81 (0.80)	0.47 (0.49)	0.87 (0.94)	0.51 (0.60)
Logistic_Regression	20 (106)	95 (399)	40 (127)	19 (19)	0.51 (0.85)	0.7 (0.76)	0.33 (0.45)	0.83 (0.95)	0.4 (0.59)
Elastic Net	19.97 (105.4)	94.89 (397.95)	40.11 (128.05)	19.03 (19.6)	0.51 (0.84)	0.7 (0.76)	0.33 (0.45)	0.83 (0.95)	0.4 (0.59)
KNearest_Neighbors	7 (125)	111(526)	24 (0)	32 (0)	0.18 (1)	0.82 (1)	0.23 (1)	0.78 (1)	0.2 (1)
Support_Vector_Machine	22 (110)	91 (420)	44 (106)	17 (15)	0.56 (0.88)	0.67 (0.80)	0.33 (0.51)	0.84 (0.97)	0.42 (0.65)
Gradient Boosting Classifier	13.05 (114.9)	123.83 (526)	11.17 (0)	25.95 (10.1)	0.33 (0.92)	0.92 (1)	0.54 (1)	0.83 (0.98)	0.41 (0.96)

[a] The 100 features (k) found in both CAN-BIND and STAR*D are used for these evaluations. Training and evaluation are repeated 100 times to obtain mean scores. All scores pertain to mean values across 100 independent runs of each model. Test scores are shown in nonbrackets, while training scores are shown with brackets. Abbreviations: KNN, K Nearest Neighbors; SVM Support Vector Machine, GBDT, Gradient Boosting Decision Tree; TP, True Positive; TN, True Negative; FP, False Positive; FN, False Negative; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Supplementary Table 5. STAR*D internal validation (k = 100) comparison of Balanced Accuracy P-values.[a]

	Dummy Classifier	Logistic Regression	Random Forest	Elastic Net	KNN	SVM	GBDT
Dummy Classification	1	4.74E-138*	7.23E-111*	8.98E-138*	2.20E-84*	9.28E-117*	5.31E-61*
Logistic Regression	4.74E-138*	1	2.95E-105*	3.14E-09*	0*	0*	1.20E-155*
Random Forest	7.23E-111*	2.95E-105*	1	2.35E-99*	1.08E-109*	0.070874258	1.84E-96*
Elastic Net	8.98E-138*	3.14E-09*	2.35E-99*	1	6.32E-210*	3.66E-149*	1.16E-150*
KNN	2.20E-84*	0*	1.08E-109*	6.32E-210*	1	0*	5.84E-33*
SVM	9.28E-117*	0*	0.070874258	3.66E-149*	0*	1	7.78E-116*
GBDT	5.31E-61*	1.20E-155*	1.84E-96*	1.16E-150*	5.84E-33*	7.78E-116*	1

*P < 0.05, with Bonferroni Correction (n= 49).

[a] The 100 features (k) found in both CAN-BIND and STAR*D are used for these evaluations. Training and internal evaluation are repeated 100 times to obtain mean balanced accuracy, and two-tailed t-tests were performed comparing model performance. Abbreviations: KNN, K Nearest Neighbors; SVM Support Vector Machine, GBDT, Gradient Boosting Decision Tree.

Supplementary Table 6. CAN-BIND-1 external validation (k = 100) comparison of Balanced Accuracy P-values.[a]

	Dummy Classifier	Logistic Regression	Random Forest	Elastic Net	KNN	SVM	GBDT

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Dummy Classification	1	2.54E-72*	2.81E-112*	1.27E-71*	0.628499721	5.47E-79*	1.76E-78*
Logistic Regression	2.54E-72*	1	1.17E-159*	0.006315537	0*	0*	2.17E-27*
Random Forest	2.81E-112*	1.17E-159*	1	8.49E-157*	1.85E-230*	6.47E-148*	8.30E-94*
Elastic Net	1.27E-71*	0.006315537	8.49E-157*	1	1.55E-283*	9.68E-98*	2.69E-28*
KNN	0.628499721	0*	1.85E-230*	1.55E-283*	1	0*	2.15E-162*
SVM	5.47E-79*	0*	6.47E-148*	9.68E-98*	0*	1	1.84E-06*
GBDT	1.76E-78*	2.17E-27*	8.30E-94*	2.69E-28*	2.15E-162*	1.84E-06*	1

*P < 0.05, with Bonferroni Correction (n= 49).
[a] The 100 features (k) found in both CAN-BIND and STAR*D are used for these evaluations. Training and external evaluation are repeated 100 times to obtain mean balanced accuracy, and two-tailed t-tests compared model performance. Abbreviations: KNN, K Nearest Neighbors; SVM Support Vector Machine, GBDT, Gradient Boosting Decision Tree.

Supplementary Table 7. P-values for mean balanced accuracy, comparing STAR*D internal validation (k = 100) vs CAN-BIND-1 external (k = 100) validation performance.[a]

	Dummy Classification	Logistic Regression	Random Forest	Elastic Net	KNN	SVC	GBDT
P-value	0.291463749	0*	4.73E-30*	1.28E-225*	0*	0*	3.20E-05*

*P < 0.05, with Bonferroni Correction (n= 7).
[a] The 100 features (k) found in both CAN-BIND and STAR*D are used for these evaluations. Training and evaluation are repeated 100 times to obtain mean balanced accuracy. Two tailed t-tests were performed, comparing internal vs external model performance. Abbreviations: KNN, K Nearest Neighbors; SVM Support Vector Machine, GBDT, Gradient Boosting Decision Tree.

Supplementary Table 8: Balanced accuracy and AUC score of various machine learning models on the STAR*D holdout set including when using Elastic Net feature selection.[a]

Models	Mean Balanced Accuracy	Mean AUC Score
Logistic Regression (k = 480)	72%	0.80
Logistic Regression (ENet k = 61)	66%	0.76
Elastic Net Regression (k = 480)	68%	0.75
Random Forest (k = 480)	72%	0.78
Random Forest (ENet k = 61)	70%	0.79
SVC (k= 480)	71%	0.79
SVC (ENet k = 61)	68%	0.75
GBDT (k = 480)	62%	0.77
GBDT (ENet k = 61)	65%	0.76
KNN (k = 480)	54%	0.76
KNN (ENet k = 61)	51%	0.62

[a] Mean performance from 100 runs of each model is assessed using all 480 features and a reduced feature set of 61 features, via cross-validated elastic net (ENet). N represents the number of features that a model was trained and evaluated on. Abbreviations: SVC, Support Vector Classifier; GBDT, Gradient Boosting Decision Tree; KNN, K Nearest Neighbors.

Supplementary Table 9. Balanced accuracy and AUC scores (brackets) across different machine learning models after Elastic Net feature selection, assessed on holdout sets.[a]

	Mean balanced accuracies (mean AUC scores)	
Model	STAR*D Holdout Set (ENet k = 30)	CAN-BIND-1 External Validation (ENet k = 30)
Logistic Regression	72% (0.82)	60% (0.69)
Random Forest	71% (0.81)	65% (0.74)
SVC	72% (0.78)	59% (0.66)
GBDT	60% (0.73)	59% (0.67)
KNN	59% (0.71)	62% (0.68)

[a] The 30 features were derived via Elastic Net Selection from the 100 features (k) found in both CAN-BIND and STAR*D are used for these evaluations. Training and evaluation are repeated 100 times to obtain mean scores. Abbreviations: AUC, area under the curve; SVC, Support Vector Classifier; GBDT, Gradient Boosting Decision Tree; KNN, K Nearest Neighbors.

Supplementary Table 10. P-values for Random Forest mean balanced accuracy on STAR*D internal validation, across several feature set variations.[a]

	Full (k=480)	No QIDS-SR or Q-LES-Q-SF Full STAR*D (k=391)	Overlapping (k = 100)	QIDS-SR only (k=47)	Q-LES-Q-SF only (k=16)	QIDS-SR + Q-LES-Q-SF (k=63)	No QIDS-SR or Q-LES-Q-SF Over-lapping (k=37)
Full (k =480)[b]	1	9.77E-97	0.008102091	2.74E-39	1.50E-82	0.011822109	2.24E-76
No QIDS-SR or Q-LES-Q-SF Full STAR*D (k =391)[c]	9.77E-97	1	8.71E-82	5.85E-47	7.13E-37	6.44E-92	0.000167881
Overlapping (k = 100)[d]	0.008102091	8.71E-82	1	2.13E-26	1.19E-59	0.567348003	4.80E-64
QIDS-SR only (k = 47)[e]	2.74E-39	5.85E-47	2.13E-26	1	1.44E-12	8.10E-33	5.75E-30
Q-LES-Q-SF only (k =16)[f]	1.50E-82	7.13E-37	1.19E-59	1.44E-12	1	7.44E-75	4.65E-17
QIDS-SR + Q-LES-Q-SF (k = 63)[g]	0.011822109	6.44E-92	0.567348003	8.10E-33	7.44E-75	1	1.25E-71
No QIDS-SR or Q-LES-Q-SF Overlapping (k = 37)[h]	2.24E-76	0.000167881	4.80E-64	5.75E-30	4.65E-17	1.25E-71	1

[a] Training and evaluation are repeated 100 times to obtain mean scores. [b] Full 481 features. [c] Full feature set, excluding QIDS-SR or Q-LES-Q related features. [d] Overlapping 100 features. [e] Only features related to QIDS. [f] Only features related to QLESQ-SF. [g] Both QIDS-related and QLESQ-SF-related features. [h] All overlapping features, excluding those related to QIDS-SR or QLESQ-SF. Abbreviations: Q-LES-Q-SF, Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form; QIDS-SR Quick Inventory of Depressive Symptomatology - Self Report.

Supplementary Table 11. A comparison of Random Forest balanced accuracies and AUC scores for top performing models on STAR*D holdout set, using several feature variations[a].

	Mean balanced accuracies (mean AUC scores)			
	Elastic Net	Logistic Regression	Random Forest	Support Vector Classifier
Full (k= 480)[b]	0.68 (0.75)	0.72 (0.80)	0.72 (0.78)	0.71 (0.79)
QIDS-SR (k = 72)[c]	0.69 (0.77)	0.63 (0.76)	0.69 (0.76)	0.73 (0.77)
QLESQ-SF (k = 17)[d]	0.73 (0.77)	0.68 (0.75)	0.66 (0.78)	0.71 (0.77)
QIDS-SR + QLESQ-SF (k = 89)[e]	0.73 (0.82)	0.73 (0.82)	0.69 (0.78)	0.73 (0.81)
Exclude QIDS-SR + QLESQ-SF (k = 391)[f]	0.66 (0.68)	0.57 (0.70)	0.64 (0.72)	0.66 (0.74)

[a] Training and evaluation are repeated 100 times to obtain mean scores. [b] Full 481 features. [c] Only features related to QIDS. [d] Only features related to QLESQ-SF. [e] Both QIDS and QLESQ-SF features. [f] All features, excluding those related to QIDS-SR or QLESQ-SF. Abbreviations: AUC, area under the curve; k, number of features, Q-LES-Q-SF, Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form; QIDS-SR Quick Inventory of Depressive Symptomatology - Self Report.

Supplementary Table 12. P-values for Random Forest mean balanced accuracy on CAN-BIND-1 (k=100) external validation, across several feature set variations.[a]

	Overlapping (k=100)	QIDS-SR only (k=47)	Q-LES-Q-SF only (k=16)	QIDS-SR + Q-LES-Q-SF (k=63)	No QIDS-SR or Q-LES-Q-SF Overlapping (k=37)
Overlapping (k = 100)[b]	1	1.99E-51	3.24E-171	3.82E-11	2.09E-142
QIDS-SR only (k = 47)[c]	1.99E-51	1	7.85E-150	3.40E-31	1.81E-124

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Q-LES-Q-SF only (k =16)[d]	3.24E-171	7.85E-150	1	1.27E-165	2.29E-13
QIDS-SR + Q-LES-Q-SF (k = 63)e	3.82E-11	3.40E-31	1.27E-165	1	3.32E-137
No QIDS-SR or Q-LES-Q-SF Overlapping (k = 37)[f]	2.09E-142	1.81E-124	2.29E-13	3.32E-137	1

[a] Training and evaluation are repeated 100 times to obtain mean scores. [b] Overlapping features compatible between STAR*D and CAN-BIND-1. [c] Only features related to QIDS. [d] Only features related to QLESQ-SF. [e] Both QIDS-related and QLESQ-SF-related features. [f] All features excluding those related to QIDS-SR or QLESQ-SF. Abbreviations: Q-LES-Q-SF, Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form; QIDS-SR Quick Inventory of Depressive Symptomatology - Self Report.

Supplementary Table 13. P-values for Random Forest balanced accuracy, comparing STAR*D internal validation vs CAN-BIND-1 external validation performance, across feature sets.[a]

Feature Set:	Overlapping (k = 100)	QIDS-SR (k = 47)	Q-LES-Q-SF (k=16)	QIDS-SR + Q-LES-Q-SF (k=63)	No QIDS-SR or Q-LES-Q-SF (k=37)
P-Value:	4.73E-30*	1.69E-27*	3.40E-157*	1.11E-60*	9.45E-97*

*P < 0.05, with Bonferroni Correction (n= 5).

[a] Training and evaluation are repeated 100 times to obtain mean scores. [b] Overlapping features. [c] Only features related to QIDS. [d] Only features related to QLESQ-SF. [e] Both QIDS-related and QLESQ-SF-related features. [f] All features excluding those related to QIDS-SR or QLESQ-SF. Abbreviations: Q-LES-Q-SF, Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form; QIDS-SR Quick Inventory of Depressive Symptomatology - Self Report.

Appendix 1: Description of the overlapping features used in this work.

One hundred features overlapped appropriately, between the STAR*D and CAN-BIND-1 datasets. Each STAR*D variable included in the overlapped dataset that was used for both internal and external validation, is described below along with its CAN-BIND-1 equivalent scale. Overlapping Features (k = 100):

Demographics:	Educat - # of years in formal education; Empl – Current employment status; Totincom – Monthly household income; Marital – current marital status; Interview age - Age in months at the time of the interview/test/sampling/imaging; Gender - Female, Male.
Work Productivity and Impairment (WPAI) (CAN-BIND-1 equivalent –Lam Employment Absence and Productivity Scale (LEAPS))	Totalhrs – hours missed + hours worked; Wpai02 – work hours missed due to health problems.
Psychiatric Hx (CAN-BIND-1 equivalent – Mini International Neuropsychiatric Interview (MINI))	Dep – family history of depression; Dage – age of onset of first major depressive episode (MDE); Episode_date – onset of current MDE; Epino – number of MDEs; Pd_ag – Panic w/ agoraphobia; Pd_noag – panic w/ agoraphobia; Alcoh – alcohol abuse; Anorexia – anorexia present; Antis – antisocial personality; Bulimia – current diagnosis of bulimia; Gad_phx – generalized anxiety; Ocd_phx - obsessive-compulsive; Psd – post traumatic stress; Amphet – amphetamine abuse; Soc_phob – social phobia; Imput_Any anxiety – phx01_psd + pd_noag + pd_ag+ soc_phob + gad_phx + specphob; PHYHIS_MDD_PREV:: - number of MDE's >= 2 (similar to epino).
QIDS-SR (W0sr = Week 0 Self-rated, W2sr = Week2 self-rated)	W0sr_qstot – QIDS Total Score; W0sr_vcntr – Concentration/Decision Making; W0sr_vvwsf – self outlook; W0sr_vsuic – suicidal ideation; W0sr_vintr – Involvement in interests/activities; W0sr_vengy – Energy/Fatigability; W0sr_vslow – psychomotor slowing; W0sr_vagit – psychomotor agitation; W0sr_vsoin – sleep onset insomnia; W0sr_vmnin -mid-nocturnal insomnia; W0sr_vemin – early morning insomnia; W0sr_vhysm - hypersomnia; W0sr_vmdsd – Sad Mood; W0sr_vapdc – decreased appetite; W0sr_vapin – increased appetite; W0sr_vwtdc – decreased weight in past 2 weeks; W0sr_vwtin – increased weight in past 2 weeks.
QIDS Atypical Baseline QIDS Atypical Week 2	Imput_QIDS_SR_appetite_domain_week0 – including both increased and decreased appetite; Imput_QIDS_SR_appetite_domain_week2 – including both increased and decreased appetite; Imput_QIDS_SR_ insomnia_week0 – including sleep-onset, mid-nocturnal, and early-morning subtypes; Imput_QIDS_SR_ insomnia_week2 – including sleep-onset, mid-nocturnal, and early-morning subtypes; Imput_QIDS_SR_ overeating_week0 - including increased appetite and increased weight; Imput_QIDS_SR_ overeating_week2- including increased appetite and increased weight; Imput_QIDS_SR_ psychomotor_week0 - including psychomotor slowing and agitation; Imput_QIDS_SR_ psychomotor_week2 - including psychomotor slowing and agitation; Imput_QIDS_SR_ sleep_week0 - including all 3 subtypes of insomnia + hypersomnia; Imput_QIDS_SR_ sleep_week2 - including all 3 subtypes of insomnia + hypersomnia; Imput_QIDS_perc_change – percent change in total score from week 0 to week 2.
QLESQ Total	Total QLESQ baseline score; Qlesq01 – Overall physical Health; Qlesq02 - Mood; Qlesq03 – Work Performance; Qlesq04 – Household Activities; Qlesq05 – Social Relationships; Qlesq06 – Familial relationships; Qlesq07 – Leisure Activities; Qlesq08 – Ability to function; Qlesq09 – Sexual drive; Qlesq10 – economic status; Qlesq11 – Housing situation; Qlesq12 – Physical Mobility; Qlesq13 - Vision; Qlesq14 -Overall well-being; Qlesq15 - Treatment; Qlesq16 – overall contentment.
Work and Social Adjustment Scale (CAN-BIND-1 Equivalent – Sheehan Disability Scale, SDS))	Wsas01 – Work is impaired; Wsas02 – Home management impairment; Wsas03 – Social activities impairment;

Some models are trained with all 480 available STAR*D features which are well defined elsewhere, and are accessible on the NDA data archive.

Appendix 2: Participant Selection

Aside from clinical inclusion/exclusion criteria included in STAR*D and CAN-BIND-1 studies, several data processing steps were undertaken that affected subject selection, resulting in a total of 651 examples for STAR*D and 178 for CAN-BIND-1. Due to difference in the formatting of the STAR*D and CAN-BIND-1 datasets, different processing steps were applied and are summarized below, along with rationale as necessary.The STAR*D selection steps and rationale are described below and can found in classes.py and stard_preprocessing_manager.py. They are listed in the order they were performed in the script, and the number of participants excluded via each step is shown.

Starting number of unique participants, after initial dataset cleaning: 3818

1. Remove participants who went into early follow-up or Level 2, likely due to adverse side effects of the medication (3739) (-79)
2. Only include participants who have at least one Q-LES-Q-SF value between 4-9 weeks of the study, using last-outcome-carried-forward (LOCF) (3657) (-82)
3. Drop rows with missing Q-LES-Q-SF values. (3592) (-65)
4. Remove duplicate rows. (3592) (-0)
5. Remove rows involving Level 3 or Level 4, as the scope is limited to the end point of Level 1 in the STAR*D trial. If a Q-LES-Q-SF value located between 4-9 weeks but was associated with a first occurrence of Level 2 for that patient, it was kept. The rationale is that the patient has just stopped their Level 1 drug. (3592) (-0)
6. Remove participants where there is one or less QoL data point. (-2879)
7. Remove participant if their baseline Q-LES-Q-SF value is beyond the 4th week of the study (-7)
8. Remove participant if their final Q-LES-Q-SF is before the 4th week of the study (- 38)
9. Remove patients who start with a Q-LES-Q-SF baseline score that is already above the established threshold for a nonimpaired QoL (≥67). (--17)

Final number of unique participants: 651

The CAN-BIND-1 selection steps and rationale are described below and can found in canbind_ygen.py:

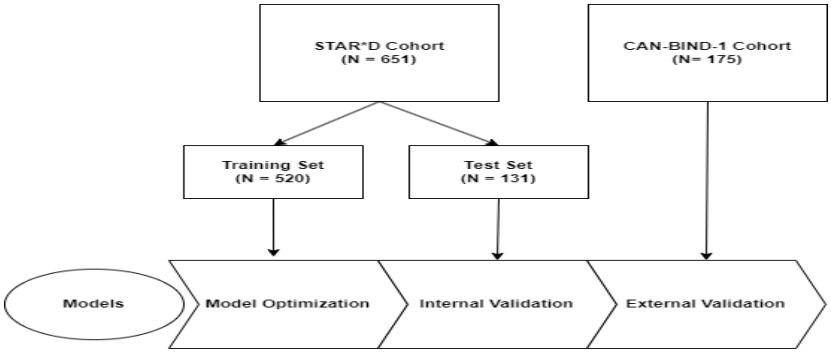
Starting number of unique participants, after initial dataset cleaning: 323

1. Only keep data from the treatment group, excluding the control group (211) (-112)
2. Remove rows with missing values for Q-LES-Q-SF (-0)
3. Only retain participants that have both a baseline + Week 8 Q-LES-Q-SF score (176) (-35)
4. Exclude participants if missing ≥ 5 questions from the 14 items Q-LES-Q-SF scale (-0)
5. Remove patients who start with a Q-LES-Q-SF baseline score that is already above the established threshold for a nonimpaired QoL (67). (-1)

Final number of unique participants: 175

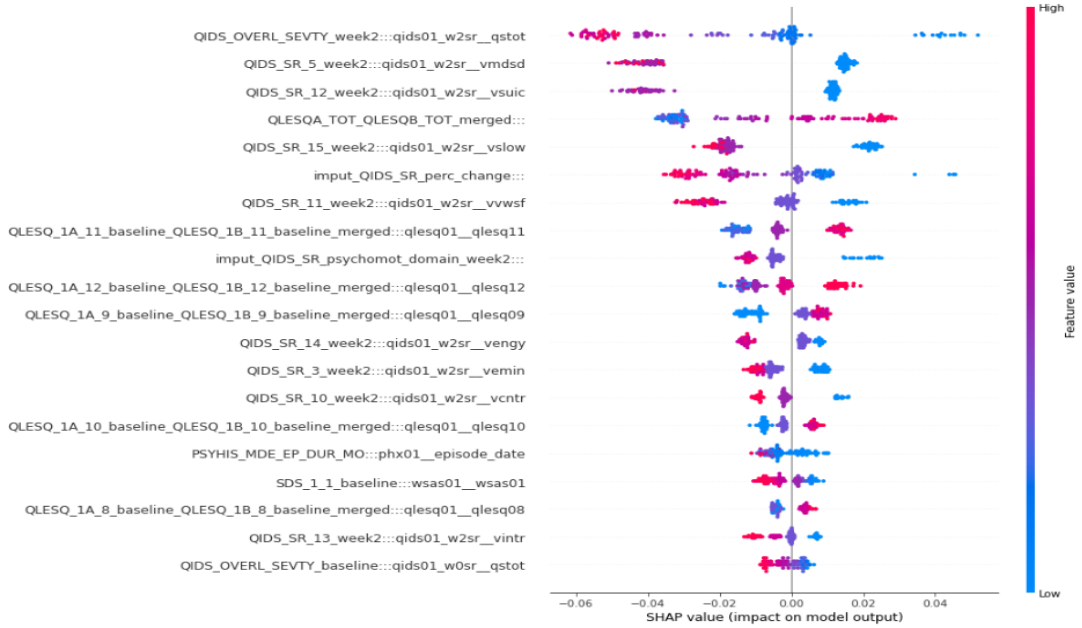
Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.

Supplementary Figure 1. A flowchart summary of model training and evaluation, where N represents sample size for each of the steps.



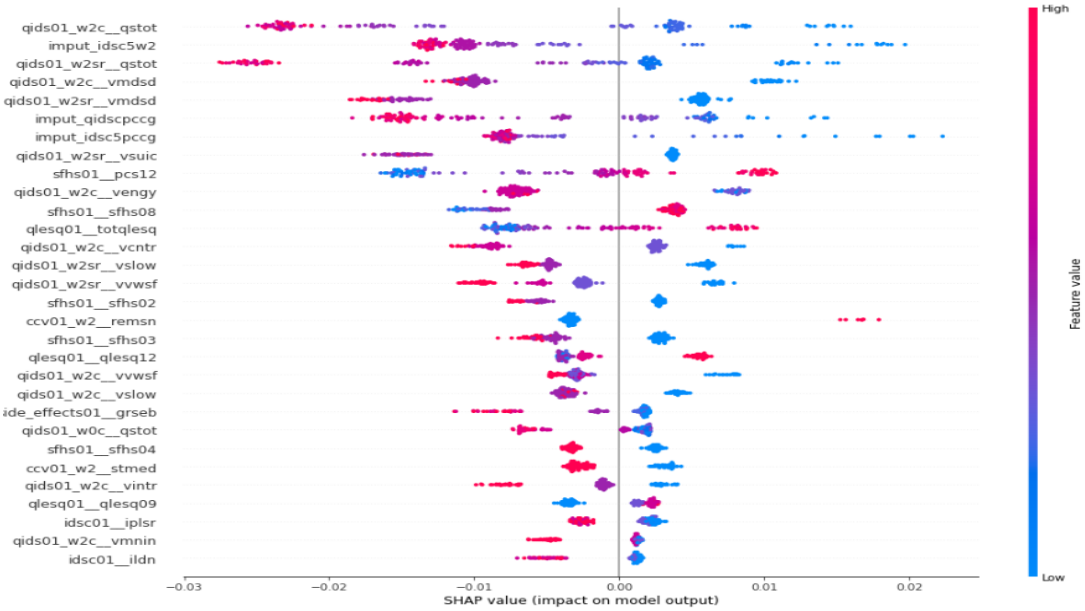
Abbreviations: CANBIND-1, Canadian Biomarker Integration Network in Depression; STAR*D, Sequenced Treatment Alternatives to Relieve Depression.

Supplementary Figure 2. SHAP values taken from a randomly sampled Random Forest model, trained on the overlapping feature set (k = 100).



Abbreviations: SHAP, SHapley Additive exPlanations; CANBIND-1, Canadian Biomarker Integration Network in Depression; STAR*D, Sequenced Treatment Alternatives to Relieve Depression; Q-LES-Q-SF, Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form; QIDS-SR Quick Inventory of Depressive Symptomatology - Self Report.

Supplementary Figure 3. SHAP values taken from a randomly sampled Random Forest model, trained on the full dataset (k= 480).



Abbreviations: SHAP SHapley Additive exPlanations; CANBIND-1, Canadian Biomarker Integration Network in Depression; STAR*D, Sequenced Treatment Alternatives to Relieve Depression; Q-LES-Q-SF, Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form; QIDS-SR Quick Inventory of Depressive Symptomatology - Self Report.

Posting of this PDF is not permitted. | For reprints or permissions, contact permissions@psychiatrist.com. | © 2023 Physicians Postgraduate Press, Inc.