



HHS Public Access

Author manuscript

J Affect Disord. Author manuscript; available in PMC 2024 March 01.

Published in final edited form as:

J Affect Disord. 2023 March 01; 324: 286–293. doi:10.1016/j.jad.2022.12.076.

Temporal multi-step predictive modeling of remission in major depressive disorder using early stage treatment data; STAR*D based machine learning approach

Haitham Salem, MD, PhD¹, Tung Huynh, MSc², Natasha Topolski, MD, PhD², Benson Mwangi, PhD², Madhukar H. Trivedi, MD³, Jair C. Soares, MD, PhD², A. John Rush, MD^{4,5}, Sudhakar Selvaraj, MD, PhD²

¹Department of Psychiatry and Human Behavior (DPHB), Warren Alpert School of Medicine, Brown University, Providence, RI

²Louis Faillace Department of Psychiatry and Behavioral Science, McGovern Medical School, University of Texas Health Science Center, Houston, TX

³Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA

⁴Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA

⁵Professor Emeritus, Duke-National University of Singapore, Singapore

Abstract

Background: Artificial intelligence is currently being used to facilitate early disease detection, better understand disease progression, optimize medication/treatment dosages, and uncover promising novel treatments and potential outcomes.

Methods: Utilizing the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) dataset, we built a machine learning model to predict depression remission rates using same clinical data as features for each of the first three antidepressant treatment steps in STAR*D. We only used early treatment data (baseline and first follow up) in each STAR*D step to temporally analyze predictive features of remission at the end of the step.

Corresponding author; Sudhakar Selvaraj, MD, PhD. Associate Professor of Psychiatry, Director, Depression Research Program, Louis Faillace Department of Psychiatry and Behavioral Science, McGovern Medical School, University of Texas Health Science Center, 1941 East Rd, BBSB, Suite 3152, Houston, TX, 77054, Sudhakar.Selvaraj@uth.tmc.edu.

⁹AUTHOR CONTRIBUTIONS:

HS, TH, BM, and SS conceptualized the study goals and designed and conducted regular discussions to verify the analytical methods and outcomes. SS was in charge of the project's overall supervision and direction and acquired access to the original STAR*D data. TH designed the computational framework. HS and NT drafted the manuscript and revisions. JCS, MHT, and JAR, critically revised the manuscript and the framework described within. All authors discussed the results and approved the final version of the manuscript for publication.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

CONFLICT ON INTEREST

Authors declare no conflict on interests associated with this work

Results: Our model showed significant prediction performance across the three treatment steps. At step 1, Model accuracy was 66%; sensitivity-65%, specificity-67%, positive predictive value (PPV)-65.5%, and negative predictive value (NPV)-66.6%. At step 2, model accuracy was 71.3%, sensitivity-74.3%, specificity-69%, PPV-64.5%, and NPV-77.9%. At step 3, accuracy reached 84.6%; sensitivity-69%, specificity-88.8%, PPV-67%, and NPV-91.1%. Across all three steps, the early Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR) scores were key elements in predicting the final treatment outcome. The model also identified key sociodemographic factors that predicted treatment remission at different steps.

Limitations: The retrospective design, lack of replication in an independent dataset, and the use of “a complete case analysis” model in our analysis.

Conclusions: This proof-of-concept study showed that using early treatment data, multi-step temporal prediction of depressive symptom remission results in clinically useful accuracy rates. Whether these predictive models are generalizable deserves further study.

Keywords

Depression; Machine learning; Remission; Predictive models; Decision Trees

3. INTRODUCTION

Major depressive disorder (MDD) is a debilitating mood dysregulation disorder characterized by various symptoms that encompass changes in mood, motivation, activity levels, sleep, cognitive symptoms, and recurrent thoughts of death (American Psychiatric Pub, 2013). The lifetime prevalence of MDD is 20%, with almost 70% of those affected seeking treatment at some point (Hasin et al., 2018). MDD is estimated to have an annual societal burden of up to \$188 billion in the United States. To put this into perspective, during the same year of this estimate, the societal cost of cancer was \$131 billion (WHO, 2020).

Treatment for MDD remains challenging. Less than 30% of patients with MDD will reach remission after their initial treatment. At least 12% of patients are characterized as treatment-resistant depression (TRD) (Dunlop et al., 2012; Mrazek et al., 2014). Treatment selection is still heavily reliant on trial and error. Many individuals are met with suboptimal outcomes causing them to live with a highly stigmatized illness. Patient response to available therapies differs widely among individuals, and the lack of response to one treatment does not mean the patient will not respond to another. In fact, up to half of the patients have significant improvement after changing their regimen (Clarke et al., 2016; Simon and Perlis, 2010). Nevertheless, many individuals are met with suboptimal outcomes causing them to live with a highly stigmatized illness.

Precision psychiatry, an emerging new interdisciplinary field of psychiatry and precision medicine, promises accurate medication choices, doses, and administration times to patients with psychiatric disorders (Lin et al., 2020). Unfortunately, there is currently little information available to assist clinicians in determining the optimal individualized treatment for an individual patient, forcing them to utilize a try and try again approach in which many patients drop out after a failed therapy. About half of patients would not have follow-up

appointments after starting an antidepressant, and only a quarter would pursue further treatment options (Clarke et al., 2016; Simon and Perlis, 2010). Other studies compared remission and dropout rates among patients with MDD randomly assigned to three different treatment regimens and found that despite remission rates not being significantly different among treatment regimens, dropout rates were significantly lower in randomly assigned. When patients are assigned to their preferred treatment regimen in randomized trials, dropout rates are lower than when they are not (Al-Harbi, 2012; Zilcha-Mano et al., 2016; Dunlop et al., 2017; Lin et al., 2020).

Artificial intelligence (AI) and machine learning (ML) techniques have provided a promising avenue to analyze large datasets that could inform treatment selection or predict longer-term outcomes from earlier information in a treatment trial. In general, the ML approach is essentially three steps; starting by building a model from the available large datasets, the model is then evaluated and tuned, and finally, the model is used for predictions on previously “unseen” observations (Bzdok and Meyer-Lindenberg, 2018). Therefore, due to their ability to handle high-dimensional and large datasets, ML techniques are well suited to promote the redefinition of clinical tools used to diagnose and treat psychiatric disorders (Luo, 2017; Lin et al., 2020).

Previously, researchers used multiple ML models to identify the factors most associated with treatment outcomes and build predictive models for clinical remission of depression with significant accuracy ranging from 59.7% to 64.6% (Chekroud et al., 2016). However, patients who do not respond to antidepressants in real-life clinical scenarios are given an alternate antidepressant. In such cases, it is unclear whether the predictive features are the same or change over time during the sequential antidepressant treatment trials. Despite many attempts at incorporating ML techniques in healthcare, their role in clinical decision-making predictions in temporally based situations is yet to be fully studied (Trivedi et al., 2006; Luo, 2017; Pham et al., 2017; Kaushik et al., 2020).

This retrospective data analysis aims to determine whether ML techniques can be used to build an efficient predictive model of depressive symptom remission after antidepressant therapy, as well as temporally analyze the remission features along with the first three STAR*D treatment steps, including patients that have failed initial therapy (step 1) and elected to undergo second-line and third-line therapies. We focused on predicting remission using only data from baseline and the first post-baseline visit (usually 2-3 weeks following baseline) in each of the first three treatment steps in STAR*D.

4. PATIENTS AND METHODS

Dataset

We utilized the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) clinical trial (NCT00021528) dataset. The STAR*D trial (Rush et al., 2004) enrolled outpatients aged 18-75 suffering from major depressive disorder (MDD) whose 17-item Hamilton Depression Rating Scale (HDRS) score was at least 14 who were receiving treatment at one of 41 psychiatry clinics in the United States between July 2001 and September 2006. Step 1 of the study included 4,039 patients who received selective serotonin

reuptake inhibitor (SSRI) citalopram. Step 2 included 1,438 participants who chose to continue when they did not reach remission, respond, or were intolerant of citalopram. Step 2 consisted of three combination options (either an antidepressant or cognitive behavioral therapy (CBT) added to citalopram), and four-switch options (to either a different antidepressant or CBT). The patients received antidepressants; sertraline (a “within class” SSRI switch), venlafaxine-XR (a serotonin-norepinephrine reuptake inhibitor), bupropion-SR (a norepinephrine and dopamine reuptake inhibitor), or augmentation of citalopram with bupropion-SR or buspirone or cognitive therapy. Subsequently, Step 3 included 377 participants who continued with the study after failing Step 2. At Step 3, lithium or triiodothyronine was added to their antidepressant, or patients were switched to another antidepressant (mirtazapine or nortriptyline) (McGrath et al., 2006). Each treatment step could include up to 14-weeks of the open-label trial (Rush et al 2004 see above) All data were acquired with permission from www.nda.nih.gov.

Study design

We defined depressive symptom remission as previously described (Rush et al 2004; Trivedi et al., 2006; Rush et al., 2008) using the 16-item Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR) as reported at the end of each step (between weeks 12-14). Remission was defined as a score of 5 or less out of a possible total score of 27 (Rush et al., 2003). Participants’ QIDS-SR assessments were completed at their last clinic visit, at either 12 or 14 weeks, whichever was the latest) in each study step. We used the score at 12-14 weeks timepoints (T_{12} or T_{14}) in each of the first three steps of the STAR*D trial as the output variable for our model predictive analysis and comparisons.

At the beginning of each treatment step, data (QIDS-SR, HRSD, QIDS-C (clinician-rated), and demographic parameters) were collected at the baseline (T_0), then every 2 weeks (or 3 weeks for some patients). For our analysis, in each step, we only used the data collected at the baseline (T_0) as well as data collected at the first follow up visit (T_2) to predict the outcome (remission) as determined by the QIDS-SR score by the end of the same step (T_{12} or T_{14}).

Inclusion and exclusion criteria

Across each of the 3 steps, we included only those with complete observations across the steps. We excluded any patients with incomplete observations as well as dropouts. We defined dropouts as those who completely terminated the study at any step before the step conclusion and did not move to the next step of treatment. However, those who did not continue one step for any reason (e.g: side effects, ...etc.) but elected to join the next treatment step, are excluded from the analysis of the step that they did not finish, but are included in the analysis of the new step they joined, as long as they finished all the new step follow up requirements and their data is complete.

Predictive Models Development (Figure 1):

Features selection and data processing—We used a hybrid model for feature selection to remove redundant features as well as features having more than 10% missing values to reduce computational complexity, improve model prediction accuracy, and increase

model interpretability. As a first step, we applied the Filter Method (Chi-squared score for categorical variables & Spearman Correlation testing) independent of any machine learning algorithm to determine the best subset of attributes (Bommert et al., 2021). Then, we used the Embedded methods (to incorporate the subset features within the construction of the Machine Learning algorithm), where ElasticNet regularization was used to avoid overfitting and a tree-based Random forest model for feature performance. We applied the elastic net to select the top predictive features at step 1 of STAR*D, then we checked the top predictive candidates in steps 2 and 3, and we used only those top features common across the three levels. We identified 25-30 main features in our predictive model.

Model testing and cross-validation—We constructed our model to predict participant remission rates for each of the first three treatment steps. We used five-fold cross-validation to evaluate the prediction results. In this method, the data was split into five groups in which the model was trained on four subsets and predicts the remaining subset. This process was repeated for all groups, and model performance was averaged across the test folds. We utilized the decision tree ensemble model known as Random Forest (RF). Here the model trains a set of weak learners, i.e., decision trees using simple heuristics, and randomly selects a subset of these learners with replacement (bootstrap) that best generalizes to an unseen sample. Hyper-parameters were tuned by using the Grid Search method. To boost the scalability of our model, especially at the third step of the STAR*D study, where the available amount of data was minimal for conventional ML techniques, we added the popular extreme gradient boosting (XGBoost) as our gradient boosting machine (GBM) technique. The popularity of XGBoost is attributed to its scalable system's ability to run more than ten times faster than existing popular solutions on a single machine and scales to billions of examples in distributed or memory-limited settings. XGBoost can achieve excellent prediction results with a small number of data (features) (Chen and Guestrin, 2016; Chang et al., 2019). For each step, we calculated the model's accuracy, sensitivity, specificity, and other predictive values to evaluate its performance. All analyses were done using the open-source freely available software Python 3.6 and its relevant open source packages.

5. RESULTS:

STAR*D remission rates

Figure 2 shows the distribution of the study participants across the three treatment steps. After excluding participants with incomplete data and dropouts, the STAR*D step 1 remission rate at the cutoff point (12-14 weeks) as identified by the QIDS-SR scores was 49.05% (729/1486). At step 2, the study remission rate was 42.5% (227/534). At step 3, the study remission rate was 22.12% (23/104).

Features predicting remission

Features that were associated with the prediction of remission (versus non-remission) were selected according to their absolute weights across each of the three treatment steps. For the first step, based on data from the baseline and first post-baseline visit (T_0 and T_2), the model identified 25 top features that showed the strongest predictive values for remission week 12

(T_{12}) or 14 (T_{14}) (Figure 3). Among those features, the highest predictive values were for QIDS-SR total baseline score, baseline HDRS total score, education years, and employment status. Also among the top features were; comorbidities (restlessness, anxiety, trauma reminders, avoidance) and the highest degree obtained. Notably, 11 of the 16 individual QIDS-SR items showed significant predictive values (concentration/decision making, early morning insomnia, sad mood, psychomotor agitation, self-image, interest (involvement), energy/fatiguability, sleep onset insomnia, mid-nocturnal insomnia, psychomotor slowing, and suicide ideations). QIDS-SR individual items for appetite and weight parameters did not yield enough significance to be included in the top predictive features at this step. For the HDRS scale, apart from the total score, only 2 individual parameters (late insomnia and suicide) were among the top features of remission as predicted by our model (Figure 3).

For the second step, the highest predictive values were presented by QIDS-SR total score, HDRS total score, age, employment status, and side effects from medicine. Other significant features were; comorbidities (restlessness, anxiety, avoidance, trauma reminders, fear of death). As with step 1, in step 2 analysis 11 individual parameters out of the usual 16 QIDS-SR items showed significant predictive values (self-image, sad mood, sleep onset insomnia, mid-nocturnal insomnia, concentration/decision making, psychomotor agitation, interest (involvement), energy/fatiguability, psychomotor slowing, suicide ideations, and early morning insomnia). It is also important to highlight that the patient's initial state as defined by their QIDS-C initial scores at baseline (T_0 of step 2) and state of improvement "QIDS-C percent improvement score" (T_2 of step 2) showed a significant predictive correlation of remission as per our model. For the HDRS scale, apart from the total score, only initial insomnia was among the top features of remission as predicted by our model (Figure 3).

For the third step, the highest predictive values were presented by QIDS-SR total score (consistent with previous levels), age, education years, the impact of family/friends, appetite/overeating disgust, multiple drug use, somatic symptoms (pain, aches), and developing side effects. Other significant predictors were; comorbidities (anxiety symptoms, trauma reminders, avoidance) and the highest degree obtained. Interestingly, apart from the total QIDS score, only 5 of the 16 QIDS items showed significant predictive values at this step (self-image, psychomotor agitation, mid-nocturnal insomnia, concentration/decision making, and energy/fatiguability). For the HDRS, only 2 individual scores showed a significant predictive value (Hypochondriasis and Initial insomnia). HDRS total score was not significant enough to be in the top features for predicting remission at this stage (Figure 3).

Across the three treatment steps, the QIDS-SR total score and 5 individual QIDS scores (self-image, mid-nocturnal insomnia, psychomotor agitation, concentration/decision-making, energy/fatiguability) showed the highest and most consistent predictive correlation with the remission status (figure 4).

By analyzing early treatment data from the baseline and subsequent first post-baseline treatment visit (after typically 2-3 weeks of treatment) in each step, our model showed a significant prediction performance of remission at 12-14 weeks post-baseline, for those

completing the full course of treatment), especially with the model trained with weights on each class to mitigate the effect of the imbalance of the distribution of classes. Table 1 shows; step 1 has an accuracy of 66%, with a sensitivity of 65%, specificity of 67%, positive predictive value (PPV) of 65.5%, and negative predictive value (NPV) of 66.6%. In step 2, model accuracy was 71.3%, with a sensitivity of 74.3%, specificity of 69%, positive predictive value (PPV) of 64.5%, and negative predictive value (NPV) of 77.9%. In step 3, accuracy reached 84.6%, with a sensitivity of 69%, specificity of 88.8%, positive predictive value (PPV) of 67%, and negative predictive value (NPV) of 91.1% (table 1).

6. DISCUSSION:

Most chronic mental illnesses, like MDD, require a multi-stage decision-making process, whereas, at each stage, treatment decision is adapted to the patient's response determined by multiple factors and observations that shape further modifications along the way. The use of sequential multiple assignment randomized trials (SMART) in medicine allows for a substantial opportunity to implement new data handling and processing techniques on multiple levels. SMART studies, like STAR*D, are typically designed to involve multiple stages of randomizations; each stage is designed to address scientific questions concerning the best intervention option to employ at that point in the intervention (Shortreed et al., 2011; Chow and Hampton, 2019; Nahum-Shani et al., 2020).

Incorporating AI-based predictions into human physicians' informed decisions is a promising method of personalized care delivery (Iniesta et al., 2016; Osuch et al., 2018). Various machine learning techniques have been widely implemented in medicine (Liu et al., 2018; Sidey-Gibbons and Sidey-Gibbons, 2019; Gao and Ding, 2020).

In our analysis, we aimed to develop a predictive model for remission that can predict outcomes of multiple treatment steps using early clinical data. Unlike the previous AI-based analyses of STAR*D study that were done solely for early steps (Trivedi et al., 2006; Rush et al., 2008; Sinyor et al., 2010; Chekroud et al., 2016), we aimed to analyze further levels of care and predict steps 2 and 3 of treatment of the STAR*D study.

Previously, Nie and colleagues used STAR*D study building on previous work to predict treatment outcomes at step 2 of the study using an RF algorithm. Interestingly, they compared two models; one used the full set of features (700), and the other model used only selected top 30 features. They reported an accuracy level of 78% using all features versus 77% using their top 30 features. In their study, the most important features were early or initial treatment response or symptom severity (Nie et al., 2018).

Kautzky and colleagues reported comparable results with their model. They used RF techniques to predict features affecting remission rates in treatment-resistant depression in 80 patients, reporting an accuracy of 85%, using response to a first antidepressant, age, severity, and suicidality as predicting features (Kautzky et al., 2017). Another interesting model by Salvo and colleagues who conducted a 12-monthly assessment on 112 depressed patients showed a remission prediction accuracy of 70% at six months and 72% at 12 months. The group used the baseline depression scores and other features such as; the

number of stressful events, perceived social support, melancholic features, the time prior to beginning treatment, and psychotherapy (Salvo et al., 2017). More recently, using Cluster analysis for individual features to represent all patterns instead of individual representation (Uher et al., 2011; Hatmann et al., 2018; Kelley et al., 2018), Paul et al used an RF algorithm to predict treatment response after 16 weeks for 834 depressed patients treated by both psychopharmacological and psychotherapeutic means, based on 50 clinical baseline parameters. They reported a significant improvement in prediction accuracy up to 96% at eight weeks when cluster-derived slopes were modeled instead of individual slopes (Paul et al., 2019).

In our model, we started by selecting the top performing features using their weights in their classes in all 3 stages, to mitigate the effect of the imbalance of the distribution of classes and to remove the redundancy as well as reduce computational complexity, hence improving the model prediction accuracy, and increasing its interpretability as compared to others' models. We then focused on those top features to conduct the analyses for the 3 stages of treatment. We showed a comparable accuracy to previously reported models with 66% accuracy in step 1, yet our model continued to do better than others in step 2 and then step 3 with an accuracy of 71.3% and 84.6%, respectively, indicating a potential benefit from the implementation of such models early in treatment trials to predict outcomes in long term multi-stage study designs.

In the third step of our analysis, we also used the XGBoost not only because it is computationally less expensive but also because it does not require a rich pool of data and is easier to interpret as compared to other models (e.g: recurrent neural network RNN based models) that we considered for our third level analysis. XGBoost is also a better modeling process compared to single machine learning models like Logistic Regression (LR), Support Vector Machine (SVM), and others. In XGBoost, the model's cross-validation is very robust as the XGBoost is a tree-based ensemble modeling wherein multiple models are built sequentially to reduce classification error on each iteration.

To our knowledge, this is the first time this model has been applied to very early stages of treatment data (baseline and first follow-up) of the STAR*D analysis. In this study, despite the significantly small sample size (n=23) in each group at step 3, our analysis showed consistency in the significance of QIDS-SR scores across the 3 steps with a high degree of accuracy in our model (84.6%), making them possible primary candidates for depression remission prediction and treatment adjustments. Moreover, due to the robust ability of the model to cross-check and tune data, the model was able to incorporate more of the clinically relevant time-related features (Multi-drug use, input from family/friends as well as weight/appetite concerns), showing them as significant predictors of remission at step 3.

When comparing the different models, in our opinion, the usual assumption of a linear relationship between the input variables and the outcomes might account for the relatively low accuracy of LR techniques in general, as the complex pathophysiological events in MDD may be correlated with each other in a nonlinear model. Instead, tree-based classifiers appeared to fit the characteristics of the dataset. We believe that the application

of regularization and high flexibility to allow for fine-tuning are important for better model performance (Hu et al., 2020; Sharma and Verbeke, 2020; Zhang et al., 2020).

In our analysis, QIDS item scores (self-image, mid-nocturnal insomnia, psychomotor agitation, concentration/decision making, energy/fatiguability) were the robust predictors across the three steps. However, it is paramount to note that the model also identified key sociodemographic factors that predicted treatment remissions at different levels, such as age, education, employment status, social support, comorbidities (anxiety, trauma), and somatic symptoms. Such factors play an important role as predictors of study attrition, especially with temporal treatment. Several studies consistently show that mental illnesses disproportionately affect people with social disadvantages due to treatment access, and economic reasons, among others. These findings further suggest that treatments need to be tailored, considering social and demographic variables, minding that different time points in MDD management may provide opportunities to engage and encourage populations at higher risk for attrition and treatment failure (Warden et al., 2007). Interestingly, at the 2nd and 3rd treatment steps, apart from the depression symptoms, more complex somatic symptoms and health variables come into play, indicating the significant impact of persistent long-term depression/stress.

Our study's strongest point is the ability to use very early clinical treatment data (baseline and first follow-up visit) to predict longer-term outcomes (12-14 weeks) with good accuracy levels consistent among each of the study's three steps. Although there are many studies using baseline clinical risk factors and/or early treatment weeks response to predict longer-term treatment outcomes within the same treatment regime in clinical samples, the studies using baseline clinical characteristics and initial or early partial response to predict the outcome of treatment are still limited (Hennings et al., 2009; Perils, 2013). It is crucial to identify reliable early predictors of treatment response that can be used to shorten or eliminate lengthy and ineffective trials, thus reducing the socioeconomic burden and moving a step forward toward more personalized treatment strategies (Leuchter et al., 2009; Taliaz et al., 2021). Due to the complex psychophysiological nature of depression, it is still challenging to identify a single marker of treatment outcome as currently there is no consensus on a biological or clinical marker for depression remission, in addition to the fact that the available clinical prognostic markers are not very effective in predicting remission rates (Dinga et al., 2018; Kraus et al. 2019; Trevizol et al., 2020). Until more broadly effective antidepressant treatments are available, AI techniques promise innovative predictive strategies that enable physicians to save critical time in choosing the best available treatment strategy for their patients is of paramount importance. However, it is worth mentioning that despite AI's promise and the substantial increase in its applications utilization in general medical practice witnessed in the past few years, its role in psychiatry, in particular, is far from optimized and remains to be explored and better elucidated (Janssen et al., 2018; Fakhoury, 2019; Graham et al., 2019; Shatte et al., 2019). In our opinion, future studies should apply AI techniques not only with clinical data but also neurobiological data (e.g: struct./fMRI, genetics, neurotropic factors) for increased accuracy and reliability.

Our study has several limitations. Our findings have not been replicated in an independent data set So generalizability could be limited (Hastie et al., 2009). An important next step

is to cross-validate our findings with independent data (Ahuja, 2019; Ermers et al., 2020). Also, due to the necessary tradeoff between higher statistical power through a large sample size and the use of powerful, specific single predictors, some useful clinical parameters were not included, which might have weakened our model's predictor ability and missed real-life important clinical parameters. Moreover, we used "a complete case analysis" model in our analysis, ignoring incomplete data due to inconsistencies or dropouts, which is also a source of bias in estimating the treatment effect. The retrospective design of this analysis is another limiting factor. One more important limitation contributing to the limited generalizability is that the STAR*D used measurement-based care that may not represent clinical practice in other places. That measurement-based care may have enabled the medication doses to levels above what is representative of the real world. Whether these findings apply to less stringently dosed depressed patients is open to question. Finally, we considered applying deep learning algorithms such as RNN, which is worth investigating for potential performance improvement and temporal sequences, but the available data was not big enough for utilizing such techniques. We believe it is worth exploring in future analysis.

Our proof-of-concept study showed that multi-step temporal prediction of depressive symptom remission using early treatment data (baseline and first follow-up visit) results in clinically useful accuracy rates. Still, further studies are needed to better assess our model's generalizability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

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 Digital Object Identifier (DOI) 10.15154/1522956]. 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.

SS has received grants/research support from NIMH R21 (1R21MH119441 – 01A1), (1R21MH129888-01A1), NICHD 1R21HD106779-01A1 and SAMHSA (FG000470-01). The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or SAMHSA.

This paper uses data collected in the STAR-D trial. STAR-D was supported by NIMH grants. The STAR*D trial ([ClinicalTrials.gov](https://clinicaltrials.gov), number [NCT00021528](https://clinicaltrials.gov/ct2/show/study/NCT00021528)) is the largest prospective, randomized controlled study of outpatients with major depressive disorder. Patients were recruited from primary and psychiatric care settings in the USA from June, 2001, to April, 2004. Study protocols have been described in detail previously [20-22].

8. FUNDING AND DISCLOSURES

TH is currently employed by Bayer Pharmaceuticals. JCS has been an advisor or received research grants from Compass Pathways, Livanova, Boehringer Ingelheim, Relmada, J&J, and Alkermes. AJR has received consulting fees from Compass Inc., Curbstone Consultant LLC, Emmes Corp., Evecxia Therapeutics, Inc., Holmusk Technologies, Inc., Johnson and Johnson (Janssen), Liva-Nova, MindStreet, Inc., Neurocrine Biosciences Inc., Otsuka-US; speaking fees from Liva-Nova, Johnson and Johnson (Janssen); and royalties from Wolters Kluwer Health, Guilford Press and the University of Texas Southwestern Medical Center, Dallas, TX (for the Inventory of Depressive Symptoms and its derivatives). He is also named co-inventor on two patents: U.S. Patent No. 7,795,033: Methods to Predict the Outcome of Treatment with Antidepressant Medication, Inventors: McMahon FJ, Laje G, Manji H, Rush AJ, Paddock S, Wilson AS; and U.S. Patent No. 7,906,283: Methods to Identify Patients at Risk of Developing Adverse Events During Treatment with Antidepressant Medication, Inventors: McMahon FJ,

Laje G, Manji H, Rush AJ, Paddock S. MHT has received research support from NIMH, NIDA, J&J, Janssen Research and Development LLC; has served as a consultant for Alkermes Inc., Allergan, Arcadia Pharmaceuticals Inc., AstraZeneca, Lundbeck, Medscape, MSI Methylation Sciences Inc., Merck, Otsuka America Pharmaceuticals Inc., and Takeda Pharmaceuticals Inc. SS has received honoraria from British Medical Journal Publishing Group and consultant or advisor for Worldwide clinical trials/Inversago and Vicore pharma. SS has received research support from Flow Neuroscience and is a study or sub-investigator for Compass Pathways, Relmada, LivaNova, and Janssen. HS, NT, and BM declare no competing conflict of interests associated with this work. No funding source had any role in the study design, data collection, analysis, interpretation, writing, or submission.

10. REFERENCES:

1. Ahuja A,S The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*. (2019) Oct 4;7:e7702. doi: 10.7717/peerj.7702. [PubMed: 31592346]
2. Al-Harbi KS. Treatment-resistant depression: therapeutic trends, challenges, and future directions. *Patient Prefer Adherence*. (2012) 6:369–88. doi: 10.2147/PPA.S29716. [PubMed: 22654508]
3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. (American Psychiatric Pub, 2013).
4. Amos T,B, Tandon N, Lefebvre P, et al. Direct and indirect cost burden and change of employment status in treatment-resistant depression: a matched-cohort study using a US commercial claims database. *J Clin Psychiatry*. (2018) 79(2):17m11725. doi: 10.4088/JCP.17m11725.
5. Bommert A, Welchowski T, Schmid M, et al. Benchmark of filter methods for feature selection in high-dimensional gene expression survival data, *Briefings in Bioinformatics*, (2021), bbab354, 10.1093/bib/bbab354
6. Bzdok D, Meyer-Lindenberg A Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. (2018) Mar;3(3):223–230. doi: 10.1016/j.bpsc.2017.11.007. [PubMed: 29486863]
7. Chang W, Liu Y, Xiao Y, et al. A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. *Diagnostics (Basel)*. (2019) 9(4): 178. Published 2019 Nov 7. doi:10.3390/diagnostics9040178 [PubMed: 31703364]
8. Chekroud A,M, Zotti R,J, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*. (2016) Mar;3(3):243–50. doi: 10.1016/S2215-0366(15)00471-X. [PubMed: 26803397]
9. Chen T, Guestrin C XGBoost: A Scalable Tree Boosting System, the 22nd ACM SIGKDD International Conference. (2016) Aug:785–794 DOI:10.1145/2939672.2939785
10. Chow J,C, Hampton L,H Sequential Multiple-Assignment Randomized Trials: Developing and Evaluating Adaptive Interventions in Special Education. *Remedial and Special Education*. (2019) 40(5):267–276. doi:10.1177/0741932518759422
11. Clarke J,L, Skoufalos A, Medalia A, et al. Improving Health Outcomes for Patients with Depression: A Population Health Imperative. Report on an Expert Panel Meeting. *Population health management*. (2016) Sep;19 Suppl 2(Suppl 2):S1–S12. doi: 10.1089/pop.2016.0114.
12. Dinga R, Marquand A,F, Veltman D,J, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry*. (2018) 8, 241. 10.1038/s41398-018-0289-1 [PubMed: 30397196]
13. Dunlop B,W, Binder E,B, Cubells J,F, et al. Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. *Trials*. (2012) Jul 9;13:106. doi: 10.1186/1745-6215-13-106. [PubMed: 22776534]
14. Dunlop B,W, Kelley M,E, Aponte-Rivera V, et al. Effects of Patient Preferences on Outcomes in the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) Study. *Am. J. Psychiatry* (2017) 174, 546–556. doi: 10.1176/appi.ajp.2016.16050517. Epub 2017 Mar 24. [PubMed: 28335624]
15. Ermers N,J, Hagoort K, Scheepers F,E The Predictive Validity of Machine Learning Models in the Classification and Treatment of Major Depressive Disorder: State of the Art and Future Directions. *Frontiers in psychiatry*. (2020) 11, 472. 10.3389/fpsy.2020.00472 [PubMed: 32523557]
16. Fakhoury M Artificial Intelligence in Psychiatry. *Adv Exp Med Biol*. (2019) 1192:119–125. doi: 10.1007/978-981-32-97121-0_6. [PubMed: 31705492]

17. Gao L, Ding Y Disease prediction via Bayesian hyperparameter optimization and ensemble learning. *BMC Res Notes*. (2020) 13(1):205. Published 2020 Apr 10. doi:10.1186/s13104-020-05050-0 [PubMed: 32276658]
18. Graham S, Depp C, Lee E,E, et al. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Current psychiatry reports*. (2019) 21(11), 116. 10.1007/s11920-019-1094-0 [PubMed: 31701320]
19. Greenberg P,E, Fournier A,A, Sisitsky T, et al. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psych*. (2015) 76(2): 155–162. doi: 10.4088/JCP.14m09298.
20. Hartmann A, von Wietersheim J, Weiss H, et al. . Patterns of symptom change in major depression: classification and clustering of long term courses. *Psychiatry Res*. (2018) 267, 480–489. 10.1016/j.psychres.2018.03.086 [PubMed: 29980128]
21. Hasin D,S, Sarvet A,L, Meyers J,L, et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry*. (2018) 75, 336–346. [PubMed: 29450462]
22. Hastie T, Tibshirani R, Friedman J *The Elements of Statistical Learning The Elements of Statistical Learning*Data Mining, Inference, and Prediction. 2nd edn, (Springer New York, New York, NY, 2009).
23. Hennings J,M, Owashi T, Binder E,B, et al. Clinical characteristics and treatment outcome in a representative sample of depressed inpatients - findings from the Munich Antidepressant Response Signature (MARS) project. *J Psychiatr Res*. (2009) Jan;43(3):215–29. doi: 10.1016/j.jpsychires.2008.05.002. Epub 2008 Jun 30. [PubMed: 18586274]
24. Hu C,A, Chen C,M, Fang Y,C, et al. Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. *BMJ Open*. (2020) Feb 25;10(2):e033898. doi: 10.1136/bmjopen-2019-033898.
25. Iniesta R, Stahl D, McGuffin P Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol Med*. (2016) Sep; 46(12):2455–65. [PubMed: 27406289]
26. Janssen R,J, Mourão-Miranda J, Schnack H,G Making Individual Prognoses in Psychiatry Using Neuroimaging and Machine Learning. *Biol Psychiatry Cogn Neurosci Neuroimaging*. (2018) Sep; 3(9):798–808.) [PubMed: 29789268]
27. Kaushik S, Choudhury A, Sheron P,K, et al. AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures. *Front Big Data*. (2020) Mar 19;3:4. doi: 10.3389/fdata.2020.00004. [PubMed: 33693379]
28. Kautzky A, Baldinger-Melich P, Kranz G,S, et al. A New Prediction Model for Evaluating Treatment-Resistant Depression. *J Clin Psychiatry*. (2017) Feb;78(2):215–222. doi: 10.4088/JCP.15m10381. [PubMed: 28068461]
29. Kelley M,E, Dunlop B,W, Nemeroff C,B, et al. Response rate profiles for major depressive disorder: characterizing early response and longitudinal nonresponse. *Depress Anxiety*. (2018) Oct;35(10):992–1000. doi: 10.1002/da.22832. Epub 2018 Sep 7. [PubMed: 30260539]
30. Kraus C, Kadriu B, Lanzenberger R, et al. Prognosis and improved outcomes in major depression: a review. *Transl Psychiatry*. (2019) Apr 3;9(1): 127. doi: 10.1038/s41398-019-0460-3. [PubMed: 30944309]
31. Kuk A,Y, Li J, Rush A,J Recursive subsetting to identify patients in the STAR*D: a method to enhance the accuracy of early prediction of treatment outcome and to inform personalized care. *J Clin Psychiatry*. (2010) Nov;71(11): 1502–8. doi: 10.4088/JCP.10m06168blu. [PubMed: 21114950]
32. Kuk A, Li J, Rush A,J Variable and threshold selection to control predictive accuracy in logistic regression. *Applied Statistics*. (2014) 63(4):657–72. 10.1111/rssc.12058
33. Leuchter A,F, Cook I,A, Hunter A,M, et al. A new paradigm for the prediction of antidepressant treatment response. *Dialogues Clin Neurosci*. (2009) 11(4):435–446. doi:10.31887/DCNS.2009.11.4/afleuchter [PubMed: 20135901]
34. Lin E, Lin C,H, Lane H,Y Precision Psychiatry Applications with Pharmacogenomics: Artificial Intelligence and Machine Learning Approaches. *Int J Mol Sci*. (2020) Feb 1;21(3):969. doi: 10.3390/ijms21030969. [PubMed: 32024055]

35. Liu L, Yu Y, Fei Z, et al. An interpretable boosting model to predict side effects of analgesics for osteoarthritis. *BMC Syst Biol.* (2018) 12, 105. 10.1186/s12918-018-0624-4 [PubMed: 30463545]
36. Luo Y Recurrent neural networks for classifying relations in clinical notes. *J Biomed Inform.* (2017) Aug;72:85–95. doi: 10.1016/j.jbi.2017.07.006. Epub 2017 Jul 8. [PubMed: 28694119]
37. McGrath PJ, Stewart JW, Fava M, et al. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *American Journal of Psychiatry.* (2006) 163:1531–1541 [PubMed: 16946177]
38. Mrazek D,A, Hornberger J,C, Altar CA, et al. Review of the Clinical, Economic, and Societal Burden of Treatment-Resistant Depression: 1996–2013. *Psychiatr Serv.* (2014) 1:65(8):977–987. doi: 10.1176/appi.ps.201300059. [PubMed: 24789696]
39. Nahum-Shani I, Almirall D, Yap JRT, et al. SMART longitudinal analysis: A tutorial for using repeated outcome measures from SMART studies to compare adaptive interventions. *Psychol Methods.* (2020) Feb;25(1):1–29. doi: 10.1037/met0000219. Epub 2019 Jul 18. [PubMed: 31318231]
40. Nie Z, Vairavan S, Narayan V,A, et al. Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study. *PLoS One.* (2018) Jun 7;13(6):e0197268. doi: 10.1371/journal.pone.0197268. [PubMed: 29879133]
41. Osuch E, Gao S, Wammes M, et al. Complexity in mood disorder diagnosis: fMRI connectivity networks predicted medication-class of response in complex patients. *Acta Psychiatr Scand.* (2018) Nov;138(5):472–482. doi: 10.1111/acps.12945. Epub 2018 Aug 6. [PubMed: 30084192]
42. Paul R, Andlauer T,F,M., Czamara D, et al. Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models. *Transl Psychiatry.* (2019) 9, 187. 10.1038/s41398-019-0524-4 [PubMed: 31383853]
43. Perils R, A H clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry.* (2013) Jul 1;74(1):7–14. doi: 10.1016/j.biopsych.2012.12.007. Epub 2013 Feb 4. [PubMed: 23380715]
44. Pham T, Tran T, Phung D, et al. Predicting healthcare trajectories from medical records: A deep learning approach. *J Biomed Inform.* (2017) May;69:218–229. doi: 10.1016/j.jbi.2017.04.001. Epub 2017 Apr 12. [PubMed: 28410981]
45. Rush A,J, Trivedi M,H, Ibrahim H,M, et al. , 2003. 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) Sep 1;54(5):573–83. doi: 10.1016/s0006-3223(02)01866-8. Erratum in: *Biol Psychiatry.* 2003 Sep 1;54(5):585. [PubMed: 12946886]
46. 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) Feb;25(1):119–42. doi: 10.1016/s0197-2456(03)00112-0. [PubMed: 15061154]
47. Rush A,J, Wisniewski S,R, Warden D, et al. Selecting among second-step antidepressant medication monotherapies: predictive value of clinical, demographic, or first-step treatment features. *Arch Gen Psychiatry.* (2008) Aug;65(8):870–80. doi: 10.1001/archpsyc.65.8.870. [PubMed: 18678792]
48. Salvo L, Saldivia S, Parra C, et al. Predictores de remisión del trastorno depresivo mayor en tratamiento en el nivel secundario de atención [Predictors of remission from major depressive disorder in secondary care]. *Rev Med Chil.* (2017) Dec;145(12):1514–1524. Spanish. doi: 10.4067/s0034-98872017001201514. [PubMed: 29652947]
49. Sharma A, Verbeke W,J MI Improving Diagnosis of Depression with XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). *Front Big Data.* (2020) Apr 30;3:15. doi: 10.3389/fdata.2020.00015. [PubMed: 33693389]
50. Shatte A,B,R, Hutchinson D,M, Teague S,J Machine learning in mental health: a scoping review of methods and applications. *Psychol Med.* (2019) Jul; 49(9):1426–1448. [PubMed: 30744717]
51. Shortreed S,M, Laber E, Lizotte D,J, et al. Informing sequential clinical decision-making through reinforcement learning: an empirical study. *Mach Learn.* (2011) 84(1-2):109–136. doi:10.1007/s10994-010-5229-0 [PubMed: 21799585]

52. Sidey-Gibbons J,A,M, Sidey-Gibbons C,J Machine learning in medicine: a practical introduction. *BMC Med Res Methodol.* (2019) Mar 19; 19(1):64. doi: 10.1186/s12874-019-0681-4. [PubMed: 30890124]
53. Simon GE, Perlis RH. Personalized medicine for depression: can we match patients with treatments? *Am. J. Psychiatry* (2010) Dec;167(12):1445–55. doi: 10.1176/appi.ajp.2010.09111680. Epub 2010 Sep 15. [PubMed: 20843873]
54. Sinyor M, Schaffer A, Levitt A The sequenced treatment alternatives to relieve depression (STAR*D) trial: a review. *Can J Psychiatry.* (2010) Mar;55(3):126–35. doi: 10.1177/070674371005500303. [PubMed: 20370962]
55. Szukis H, Joshi K, Huang A, et al. Economic burden of treatment-resistant depression among veterans in the United States, *Current Medical Research and Opinion.* (2021) 37:8, 1393–1401, DOI: 10.1080/03007995.2021.1918073.
56. Taliaz D, Spinrad A, Barzilay R, et al. Optimizing prediction of response to antidepressant medications using machine learning and integrated genetic, clinical, and demographic data. *Transl Psychiatry.* (2021) 11, 381. 10.1038/s41398-021-01488-3 [PubMed: 34238923]
57. Trevizol A,P, Downar J, Vila-Rodriguez F, et al. Predictors of remission after repetitive transcranial magnetic stimulation for the treatment of major depressive disorder: An analysis from the randomized non-inferiority THREE-D trial. *EClinicalMedicine.* (2020) 22, 100349. 10.1016/j.eclinm.2020.100349. [PubMed: 32382720]
58. Trivedi M,H, Rush A,J, Wisniewski S,R, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am. J. Psychiatry* (2006) Jan;163(1):28–40. doi: 10.1176/appi.ajp.163.1.28. [PubMed: 16390886]
59. Uher R, Mors O, Rietschel M, et al. Early and delayed onset of response to antidepressants in individual trajectories of change during treatment of major depression: a secondary analysis of data from the Genome-Based Therapeutic Drugs for Depression (GENDEP) study. *J Clin Psychiatry.* (2011) Nov;72(11):1478–84. doi: 10.4088/JCP.10m06419. [PubMed: 22127194]
60. Warden D, Trivedi M,H, Wisniewski S,R, et al. Predictors of attrition during initial (citalopram) treatment for depression: a STAR*D report. *Am J Psychiatry.* (2007) Aug;164(8):1189–97. doi: 10.1176/appi.ajp.2007.06071225. [PubMed: 17671281]
61. World Health Organization. Depression. 2020; January.
62. Zhang X, Yan C, Gao C, et al. Predicting Missing Values in Medical Data via XGBoost Regression. *J Healthc Inform Res.* (2020) Dec;4(4):383–394. doi: 10.1007/s41666-020-00077-1. Epub 2020;Aug 3. [PubMed: 33283143]
63. Zilcha-Mano S, Keefe J,R, Chui H, et al. Reducing Dropout in Treatment for Depression: Translating Dropout Predictors Into Individualized Treatment Recommendations. *J. Clin. Psychiatry* (2016);Dec;77(12):e1584–e1590. doi: 10.4088/JCP.15m10081. [PubMed: 28086005]

Highlights

- Artificial intelligence is currently being used to facilitate early disease detection, better understand disease progression, optimize medication/treatment dosages, and uncover promising novel treatments and potential outcomes.
- STAR*D dataset, was used to build a machine learning model to predict depression remission for each of the first three antidepressant treatment steps in STAR*D.
- Early treatment data (baseline and first follow up) were used in each STAR*D step.
- Model showed significant prediction performance across the three treatment steps.
- Across all three steps, the early Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR) scores were key elements in predicting the final treatment outcome.

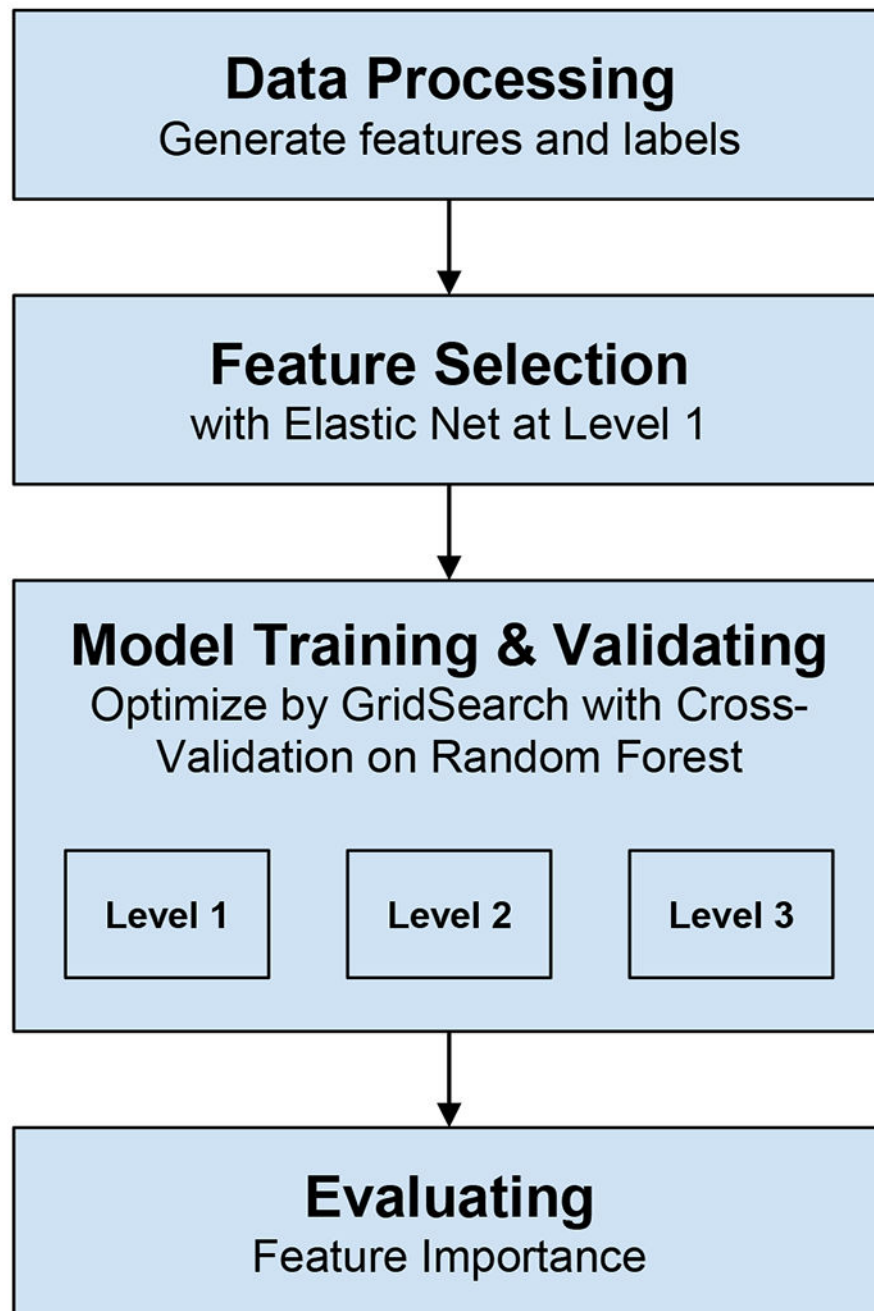


Figure 1. shows the our model development.

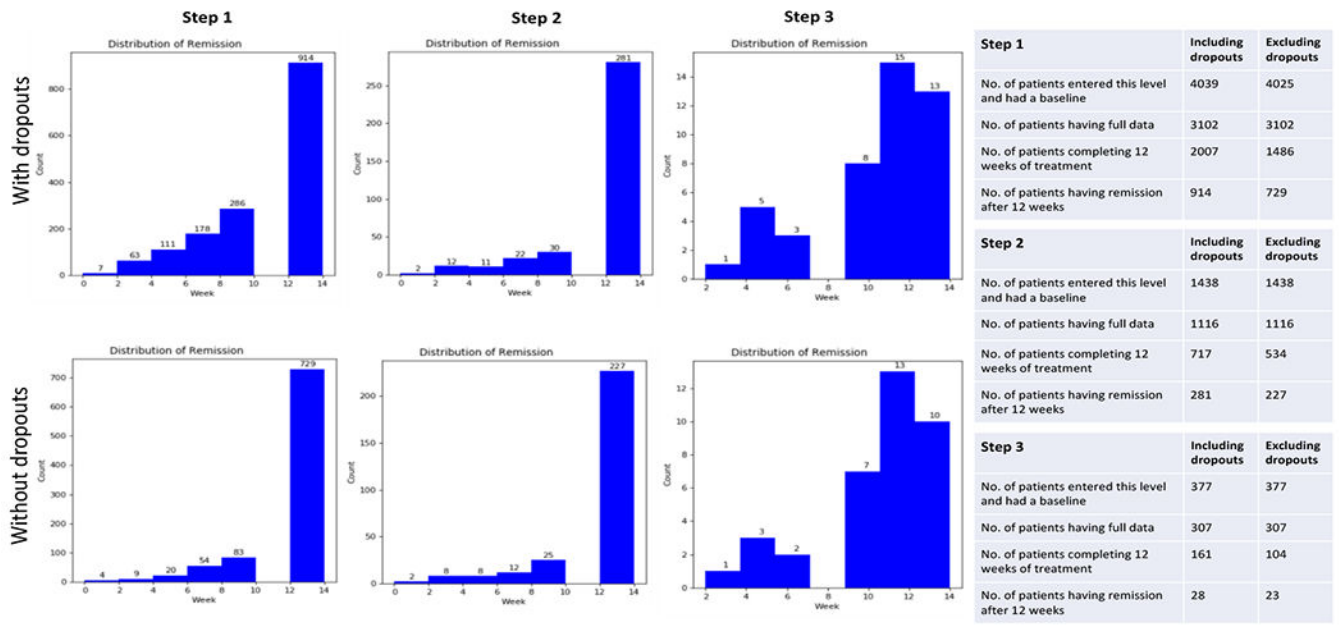


Figure 2:
Showing distribution for the study participants at each of the three steps.

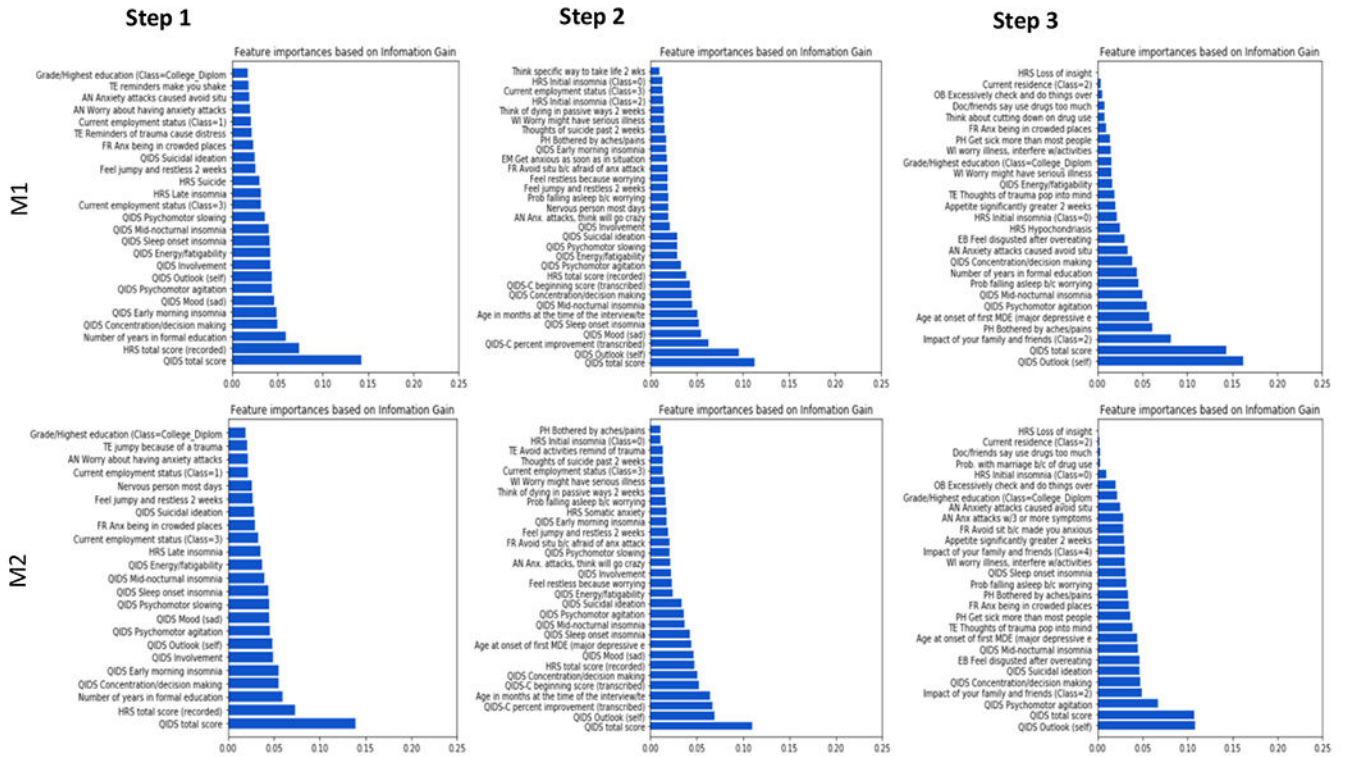
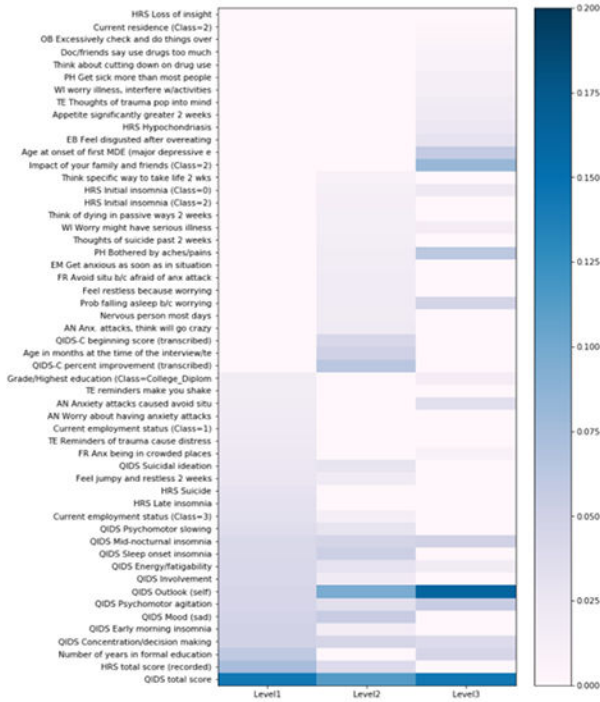
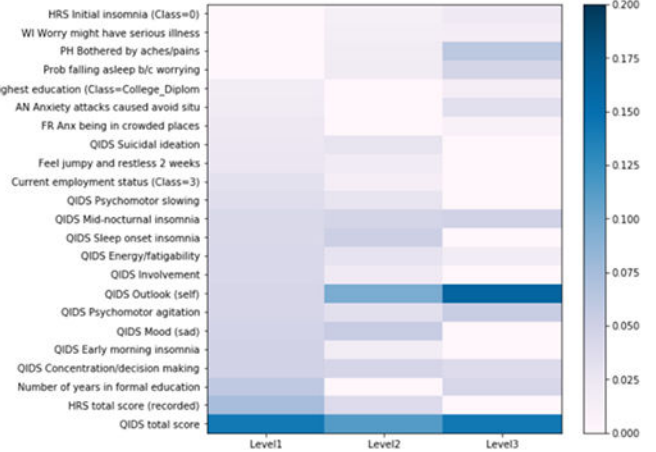


Figure 3: Important features and time points at which participants showed remission at each step during the STAR*D trial. Each step's cut-off time point is 12-14weeks. M1 = A model trained without weights on each class, M2 = A model trained with weights on each class to mitigate the effect of the imbalance of the distribution of classes.

All Features



Features contributing in at least 2 Steps



Features contributing in all of the 3 steps

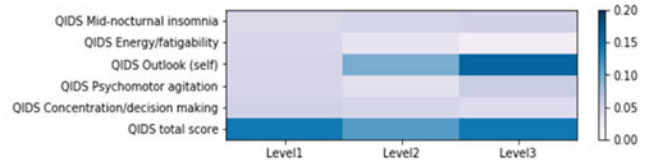


Figure 4:
Heatmaps contrasting the different features that predicted remission versus non-remission outcomes selected according to their absolute weights across the three study steps

Table 1:

Predictive model accuracy for the three different steps. M1: A model trained without weights on each class. M2: A model trained with weights on each class to mitigate the effect of the imbalance of the distribution of classes.

Remission Prediction	Step 1		Step 2		Step 3	
	M1	M2	M1	M2	M1	M2
Accuracy	0.680 (0.034)	0.66 (0.024)	0.709 (0.051)	0.713 (0.025)	0.837 (0.047)	0.846 (0.071)
AUC	0.735 (0.031)	0.731 (0.032)	0.770 (0.041)	0.775 (0.038)	0.866 (0.117)	0.842 (0.105)
Sensitivity	0.620 (0.037)	0.65 (0.032)	0.572 (0.060)	0.743 (0.06)	0.360 (0.185)	0.69 (0.12)
Specificity	0.737 (0.036)	0.67 (0.023)	0.815 (0.055)	0.69 (0.034)	0.976 (0.030)	0.888 (0.083)
PPV	0.695 (0.040)	0.655 (0.024)	0.706 (0.079)	0.649 (0.021)	0.667 (0.365)	0.67 (0.189)
NPV	0.664 (0.025)	0.666 (0.025)	0.727 (0.036)	0.779 (0.038)	0.660 (0.185)	0.911 (0.03)

AUC: Area under the Curve, PPV: Positive predictive value, NPV: Negative predictive value