European Neuropsychopharmacology 98 (2025) 26–34

Contents lists available at ScienceDirect

European Neuropsychopharmacology

journal homepage: www.sciencedirect.com/journal/european-neuropsychopharmacology

Clinical predictors of treatment resistant depression

Alessandro Serretti [a,b,*], Siegfried Kasper [c,d], Lucie Bartova [c,d], Joseph Zohar [e], Daniel Souery [f], Stuart Montgomery [g], Panagiotis Ferentinos [h], Dan Rujescu [c,d], Alexander Kautzky [c,d], Francesco Attanasio [j], Raffaella Zanardi [j,k], Chiara Fabbri [l], Bernhard T Baune [m,n,o], Raffaele Ferri [b], Julien Mendlewicz [i]

[a] Department of Medicine and surgery, Kore University of Enna, Italy
[b] Oasi Research Institute-IRCCS, Troina, Italy
[c] Department of Psychiatry and Psychotherapy, Medical University Vienna, Vienna, Austria
[d] Comprehensive Center for Clinical Neurosciences and Mental Health, Medical University of Vienna, Vienna, Austria
[e] Department of Psychiatry, Sheba Medical Center, Tel Hashomer, and Sackler School of Medicine, Tel Aviv University, Israel
[f] Epsylon caring for mental health Brussels and Laboratoire de Psychologie Médicale, ULB, Brussels, Belgium
[g] Imperial College School of Medicine, London, UK
[h] Department of Psychiatry, Athens University Medical School, Athens, Greece
[i] Université Libre de Bruxelles, Brussels, Belgium
[j] Vita-Salute San Raffaele University, Department of Clinical Neurosciences, Milan, Italy
[k] IRCCS San Raffaele Hospital, Mood Disorder Unit, Milan, Italy
[l] Department of Biomedical and NeuroMotor Sciences, University of Bologna, Bologna, Italy
[m] Department of Psychiatry and Psychotherapy, University of Münster, Münster, Germany
[n] Department of Psychiatry, Melbourne Medical School, University of Melbourne, Parkville, VIC, Australia
[o] The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia

ARTICLE INFO

Keywords:
Major depression
Antidepressants
Machine learning
Treatment resistant depression
Predictive models

ABSTRACT

Despite advances in the treatment of major depressive disorder (MDD) yet a substantial proportion of patients fail to achieve remission and instead develop treatment-resistant depression (TRD). Identifying robust clinical predictors of response is essential for early, personalized interventions.

We analyzed a large, multicenter sample (N = 2953) from the Group for the Study of Resistant Depression (GSRD) project, which included previously studied cohorts (TRD I–III) and a newly recruited cohort (TRD IV, N = 294). Patients were categorized as responders, non-responders, or TRD. Sociodemographic and clinical variables, including current and retrospective MADRS items, were used to train an XGBoost classifier. Primary outcomes were the multi-class metrics area under the curve (AUC), accuracy, and F1-scores.

Previously reported predictors were mainly confirmed in the new TRD IV sample. The XGBoost model showed a mean ROC AUC of 0.80 and an accuracy of 61 %, significantly above chance. Misclassification was more frequent among responders versus non-responders, while TRD was predicted most accurately (precision=0.73; recall=0.73). Measures of illness chronicity, such as duration of current episode, duration of disease lifetime, number of hospitalizations, and number of depressive episodes, as well as severity features, BMI and level of functioning were among the most important predictors. Secondary analyses using earlier cohorts to train and the new TRD IV sample to test confirmed stable performance metrics.

Our findings highlight the central role of chronicity indicators, severity measures and functioning in predicting antidepressant response and TRD. Future work should include prospective validation and integration of biomarker data to further enhance predictive power.

* Corresponding author.
E-mail address: alessandro.serretti@unikore.it (A. Serretti).

https://doi.org/10.1016/j.euroneuro.2025.06.011
Received 4 April 2025; Received in revised form 24 June 2025; Accepted 30 June 2025
Available online 11 July 2025

1. Introduction

Major depressive disorder (MDD) remains one of the most prevalent and debilitating psychiatric conditions (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018). Although pharmacological and psychotherapeutic interventions are available to treat MDD, a significant subset of patients fails to respond adequately to one or more treatment trials, defined as treatment-resistant depression (TRD) (McIntyre et al., 2023; Sforzini et al., 2022). Identifying reliable clinical predictors and risk factors is therefore critical, as it allows to make more informed decisions about treatment selection and optimization, potentially preventing the prolonged suffering and elevated healthcare costs associated with ineffective or partially effective interventions. Over the past two decades, several studies have focused on clinical features that may lead to TRD. Most consistent findings across many studies are symptom severity, anhedonia, early onset of depression, suicidal behaviors, comorbid psychiatric disorders (such as anxiety or substance use), and personality traits (e.g., neuroticism) (Bartova et al., 2019; De Carlo et al., 2016; Fava et al., 2008; Fekadu et al., 2009a; Serretti, 2023; Solmi et al., 2023; Souery et al., 2007; Uher et al., 2012; Zaninotto et al., 2014). Other less supported results were episode duration (Gilmer et al., 2008; Rush et al., 2012), atypical features (Stewart et al., 2010), family history of depression (Husain et al., 2009) and melancholic features (Szmulewicz et al., 2024).

Moreover, patient-specific factors such as chronic medical comorbidities and social determinants of health (e.g., low socioeconomic status or limited social support) have also been implicated in the increased likelihood of TRD (De Carlo et al., 2016; Mandelli et al., 2016; Solmi et al., 2023). While these correlational observations are informative, the complexity and interactive nature of multiple predictors can be challenging to capture using conventional statistical methods alone. Traditional regression-based approaches, for example, assume linearity and independent effects of predictors, which may not always hold true in complex psychiatric conditions like TRD, as previously discussed about the interplay between anhedonia and suicidal behaviors (Luca et al., 2024).

In recent years, machine learning (ML) techniques have been increasingly used in psychiatry due to their capacity to investigate complex, non-linear relationships within high-dimensional data (Chekroud et al., 2016). Several ML methods, including random forests, gradient boosting and support vector machines, have been applied to identify predictors of TRD, frequently demonstrating higher accuracy than conventional statistical analyses (Fabbri et al., 2020; Kautzky et al., 2017b; Perlis, 2013; Quinn et al., 2024).

However, ML studies have also reported heterogeneous results, with a variable weight of the various features included, some studies supported a preeminent role of socioeconomic status (Jain et al., 2013), early symptom improvement (Jakubovski and Bloch, 2014; Nie et al., 2018; Nunez et al., 2021), severe medical comorbidity (Falola et al., 2017), activity impairment (Jha et al., 2019), sociodemographic factors (Perlis, 2013; Salem et al., 2023), substance use (Tang et al., 2023) or a combination of a large number of features (Chekroud et al., 2016; Curtiss et al., 2024).

Among the methods used in ML analyses, gradient-boosted decision trees have attracted special attention due to their ability to handle large datasets, manage missing data effectively, and account for intricate interactions between variables (Chen and Guestrin, 2016). One of the most widely used implementations of gradient boosting is XGBoost (eXtreme Gradient Boosting). The flexible approach of XGBoost to missing data is quite relevant in this context, given the high rate of partially missing observations. The majority of previous studies limited the analysis to complete data thus dramatically reducing the sample size. It has been argued that this may limit the informativeness of results and reduce power (Nijman et al., 2022). XGBoost may handle missing data within the ML analysis (Aydin and Ozturk, 2021; Chen and Guestrin, 2016) without the need of previous raw data imputation, that may introduce bias (Chen and Xu, 2023; Latief et al., 2020; Rusdah and Murfi, 2020; Turska et al., 2021), and also relatively large missing data has been suggested to be acceptable via default direction in tree split (Bishara et al., 2022). Therefore XGBoost has shown to outperform traditional regression models in clinical data analysis (Beaulieu-Jones et al., 2023; Feng et al., 2023; Liu et al., 2024; Mortazavi et al., 2019; Wang et al., 2023). Despite the potential of XGBoost for psychiatric research, few studies have focused on leveraging this method for the prediction of TRD in large, real-world clinical cohorts.

In our previous two works investigating machine learning predictors of TRD, we included smaller samples compared to the present study, namely 480 and 558 patients (Kautzky et al., 2017a, 2017b), because we used methods that need complete data for the analysis. In the present paper we did not apply this restriction and we expanded sample size by adding our previously recruited samples (D. Souery et al., 2007; D. 2011), and a newly recruited sample not previously investigated.

2. Methods

2.1. Study sample

The sample was collected within the "Group for the Study of Resistant Depression" (GSRD) multicentric research project (Bartova et al., 2019). The sample included all the previously reported samples (named TRD I to III) (Bartova et al., 2019) and a new sample collected according to the same protocol of the TRD III study (named TRD IV). Ethical Committees of all the participating centers approved the study protocol that was performed according to the Helsinki Declaration. The enrolled subjects signed a written informed consent prior inclusion. Study details have been previously reported (Bartova et al., 2019; Dold et al., 2016), the study included adults suffering from a single or recurrent major depressive episode (MDE) diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders IV-TR and treated with at least one antidepressant agent administered at a therapeutic dosage for ≥4 weeks. Patients with current primary psychiatric disorders other than MDD, any substance dependence or severe abuse (except from nicotine and caffeine intake) in the previous 6 months, clinically relevant personality disorder and/or other relevant conditions interfering with the clinical evaluation were excluded. Demographic and clinical data were systematically evaluated including clinical interview and a range of tools: the Montgomery-Åsberg Depression Rating Scale (Montgomery and Asberg, 1979) was used to assess both the current (MADRS-c) and baseline, retrospectively assessed, (MADRS-r) severity of symptoms, Sheehan Disability Scale (SDS) (Sheehan et al., 1996) and the Mini-international neuropsychiatric interview (MINI) (Sheehan et al., 1998).

Both TRD III and the new sample TRD IV were cross sectional studies which included also a retrospective MADRS assessment of the current episode initial symptomatology (Birk et al., 2020), treatment response was therefore characterized by a MADRS-c total score of <22 and *a* ≥ 50 % reduction of the MADRS total score after an adequate AD trial. Non-response was defined as a total score of ≥22 at the MADRS-c or *a* < 50 % MADRS total score reduction after one AD trial of adequate daily dosing and duration. Treatment resistant depression (TRD) was defined by a non-response to two or more consecutive AD trials of adequate daily dosing and duration (Bartova et al., 2019; Dold et al., 2016). In the early waves of recruitment (named TRD I, retrospective, and TRD II, prospective) HAMD response (≥50 % reduction) was used (D. Souery et al., 2007; D. 2011).

2.2. Statistical analysis

Univariate methods were used to analyze the whole range of sociodemographic and clinical variables with treatment outcome (responders, non-responders, TRD). Pairwise correlations were used to evaluate reciprocal correlations across predictors. We then employed

XGBoost to predict antidepressant treatment outcomes from clinical and sociodemographic variables. In order to select the most informative baseline predicting features, we included all clinical and sociodemographic variables except for the variables related to the post baseline treatment and MADRS single items (Nunez et al., 2021). The list of the whole set of variables and the final set selected for XGBoost is reported in supplementary Table 1 and Fig. 3 respectively. The dataset was divided into training (80 %) and test (20 %) sets, maintaining class distribution using stratified sampling (train_test_split(X, y, test_size=0.2, stratify=*y*, random_state=42)). A secondary analysis was performed considering TRD I-II and TRD III samples as training and TRD IV sample as test in order to confirm the generalizability of results in the new sample. Even if the study protocol was similar, there were some differences in available variables in older samples (TRDI-II, which included also a small sample of bipolar disorder patients) compared to the most recent ones (TRD III and TRD IV, that share the same protocol) (Bishara et al., 2022), therefore another secondary analysis was performed using TRD III as training and TRD IV as testing sample in order to minimize the possible missing variables bias.

All continuous features were standardized, and categorical variables were transformed using one-hot encoding. Data were converted in DMatrix (xgb.DMatrix). Missing values were handled in the analysis and not imputed (Aydin and Ozturk, 2021; Chen and Guestrin, 2016). Samples size and complete/missing data are reported in supplementary Tables 1–4. The model was configured for multi-class classification using the soft probability objective function (`multi:softprob`). Hyperparameters were optimized to balance predictive performance and overfitting risk using Bayesian optimization, that was repeated over 50 trials using Optuna (version 4.2.1), also controlling for potentially large AUC Training-Test difference. The final parameters included (with starting ranges before bayesian optimization in brackets): learning rate: 0.13 (0.05–0.3), maximum tree depth: 4 (3–6), subsample ratio: 0.87 (0.7–0.9), colsample_bytree: 0.65 (0.6–1), lambda: 4.14 (0.5–10), alpha: 3.38 (0.5–10), number of boosting rounds: 150, evaluation metric: Multi-class log loss (`mlogloss`), early stopping rounds: 20 (based on validation set performance). Model performance was assessed using area under the curve (AUC) computed using the one-vs-rest (OvR) approach. The OvR method computes three separate AUCs: Class 1 vs. (Class 2 + Class 3), Class 2 vs. (Class 1 + Class 3) and Class 3 vs. (Class 1 + Class 2). The final AUC is the average of these three AUC scores. Confusion matrix and classification report were produced, summarizing precision, recall, and F1-score for each class. F1 = 2 * ((Precision * Recall) / (Precision + Recall)) Where: Precision = (True Positives) / (True Positives + False Positives) and Recall (Sensitivity) = (True Positives) / (True Positives + False Negatives). SHapley Additive exPlanations (SHAP) was also reported, which quantified the contribution of each feature to individual predictions. To assess whether the classification performance of the XGBoost model exceeded chance level, we conducted a permutation test (10,000 iterations) by randomly shuffling predicted labels and recalculating classification accuracy at each iteration. The observed accuracy was then compared to this null distribution to compute a non-parametric p-value. Little's MCAR test was applied to the predictors using the na.test function from the R package misty, to evaluate whether missingness was consistent with the missing completely at random assumption. Univariate analyses were not corrected for multiple testing because TRD I-II and TRD III were previously reported and TRD IV was analyzed under a confirmatory perspective. All analyses were conducted using Python 3.11 for MacOS, with XGBoost (`v2.1.4`), pandas, numpy, xgboost, matplotlib, SHAP, seaborn and SciKit-Learn packages (Virtanen et al., 2020).

3. Results

The whole sample included 2953 patients (supplementary Table 1), comprising TRD I-II patients (N = 1354, supplementary Table 2) (D. Souery et al., 2011; D. 2007), TRD III patients (N = 1305, supplementary Table 3) (Bartova et al., 2019; Dold et al., 2016) and the newly recruited sample TRD IV (N = 294, Table 1 and supplementary Table 4 for more details).

The new TRD IV sample, reported here in detail for the first time, included only responders (N = 134) and TRD patients (N = 160). Previously reported factors associated with TRD were mainly confirmed: inpatient status, antipsychotic use and more in general combination therapy, more severe baseline MADRS, greater functional impairment across all three domains, longer episode duration, higher number of previous depressive episodes and previous hospitalizations, higher suicidal risk and higher medication induced side effects (Table 1 and supplementary Tables 2–4).

As a first step before the XGBoost analysis, a correlation matrix was produced (Fig. 1 and supplementary Table 5 for more details). As can be visually noted in Fig. 1, no abnormal collinearity was detected and therefore the final set of variables selected as previously described were included in the model.

The XGBoost model demonstrated good discriminative ability with a ROC AUC of 0.8009, indicating that the model correctly differentiates between outcome classes in approximately 80.09 % of cases (Fig. 2 and supplementary Table 6). The overall classification accuracy was 61 %, significantly higher than the expected chance level of ∼33.3 % (random guessing) or 34.5 % (predicting the majority class) (permutation, p < 0.0001). The model performed best for Resistant patients (Class 2), achieving a precision of 0.73, recall of 0.73, and an F1-score of 0.73. In contrast, responders (Class 0) and non-responders (Class 1) exhibited lower but still satisfactory F1-scores (0.58 and 0.55, respectively). However the macro and weighted F1-scores were both 0.61, suggesting a relatively balanced model performance across all classes.

The confusion matrix (Fig. 2) indeed shows that the model correctly classified 107 responder, 113 non-responder, and 142 resistant cases. Misclassification was more common between responders and non-responders, while resistant cases were predicted with higher accuracy (only 19 misclassified as responders and 33 as non-responders).

The feature importance analysis (Fig. 3) showed that longer duration of current episode, longer duration of disease, higher age, onset and number of hospitalizations were the most influential predictors in the XGBoost model. Other relevant features included higher number of previous antidepressants and depressive episodes, BMI, and higher baseline MADRS severity. In contrast, variables such as substance abuse, OCD, and somatic disorders had minimal contributions to the model's predictions, in agreement with our previous analyses (Bartova et al., 2019).

SHAP analysis identified nonlinear interactions only regarding the impact of MADRS Retrospective on predictions that varied depending on MDD/BD status and sex (Supplementary figure 1).

A secondary analysis was performed with TRD I-II and TRD III samples as training and TRD IV as testing (supplementary Table 7). Results showed a small reduction on AUC (0.72) but with increased accuracy (0.70) and stable macro and weighted F1-scores. Prediction of TRD showed a very high precision (0.94) though with a less satisfactory recall (0.37). We then repeated the analysis including only the TRD III sample as training and TRD IV as testing because of lack of relevant missingness, given that the Little's MCAR test suggested a non random missing (p < 0.001), due to the TRD I-II sample rate of missing variables because of the different protocols. Results showed a similar AUC (0.74) and accuracy (0.69) and stable macro and weighted F1-scores. Prediction of TRD showed in this case a lower precision (0.64) though with a much better recall (0.99). The ranking of features was overlapping. Therefore the TRD I-II variable missing rate did not influence the prediction model.

4. Discussion

In this large, multicenter study of nearly 3000 individuals with major depression (including a newly recruited cohort of 294 patients), we

Table 1
. Clinical and sociodemographic features of the TRD IV sample.

Variable (no/yes)	Responders ($N = 134$)	Resistant ($N = 160$)	Total Sample ($N = 294$)	Statistic	p-Value
Sex (M/F)	65 (48.51 %); 69 (51.49 %)	61 (38.12 %); 99 (61.88 %)	126 (42.86 %); 168 (57.14 %); $N = 294$	2.8	0.094
Married	78 (58.21 %); 56 (41.79 %)	77 (48.12 %); 83 (51.88 %)	155 (52.72 %); 139 (47.28 %); $N = 294$	2.58	0.108
Living with others	47 (35.07 %); 87 (64.93 %)	56 (35.00 %); 104 (65.00 %)	103 (35.03 %); 191 (64.97 %); $N = 294$	0	1
Children	56 (41.79 %); 78 (58.21 %)	68 (42.50 %); 92 (57.50 %)	124 (42.18 %); 170 (57.82 %); $N = 294$	0	0.997
Smoking	74 (55.22 %); 60 (44.78 %)	97 (60.62 %); 63 (39.38 %)	171 (58.16 %); 123 (41.84 %); $N = 294$	0.67	0.414
Family history MDD	87 (64.93 %); 47 (35.07 %)	108 (67.50 %); 52 (32.50 %)	195 (66.33 %); 99 (33.67 %); $N = 294$	0.12	0.733
Family history BD	128 (95.52 %); 6 (4.48 %)	153 (95.62 %); 7 (4.38 %)	281 (95.58 %); 13 (4.42 %); $N = 294$	0	1
Family history suicide	101 (75.37 %); 33 (24.63 %)	138 (86.25 %); 22 (13.75 %)	239 (81.29 %); 55 (18.71 %); $N = 294$	4.98	0.026
Thyroid Disorder	124 (92.54 %); 10 (7.46 %)	134 (83.75 %); 26 (16.25 %)	258 (87.76 %); 36 (12.24 %); $N = 294$	4.45	0.035
Diabetes	127 (94.78 %); 7 (5.22 %)	148 (92.50 %); 12 (7.50 %)	275 (93.54 %); 19 (6.46 %); $N = 294$	0.31	0.581
Psychotherapy	67 (56.78 %); 51 (43.22 %)	91 (63.19 %); 53 (36.81 %)	158 (53.74 %); 104 (35.37 %); $N = 262$	0.86	0.353
Melancholia	27 (20.30 %); 106 (79.70 %)	29 (18.35 %); 129 (81.65 %)	56 (19.05 %); 235 (79.93 %); $N = 291$	0.07	0.787
Psychotic Lifetime	128 (96.24 %); 5 (3.76 %)	145 (91.77 %); 13 (8.23 %)	273 (92.86 %); 18 (6.12 %); $N = 291$	1.77	0.183
Psychotic Current	128 (96.24 %); 5 (3.76 %)	145 (91.77 %); 13 (8.23 %)	273 (92.86 %); 18 (6.12 %); $N = 291$	1.77	0.183
Inpatient	60 (49.18 %); 62 (50.82 %)	21 (13.82 %); 131 (86.18 %)	81 (27.55 %); 193 (65.65 %); $N = 274$	38.97	<0.001
Panic Disorder	122 (91.04 %); 12 (8.96 %)	148 (93.08 %); 11 (6.92 %)	270 (91.84 %); 23 (7.82 %); $N = 293$	0.18	0.669
GAD	120 (89.55 %); 14 (10.45 %)	138 (86.79 %); 21 (13.21 %)	258 (87.76 %); 35 (11.90 %); $N = 293$	0.3	0.586
PTSD	132 (98.51 %); 2 (1.49 %)	156 (98.11 %); 3 (1.89 %)	288 (97.96 %); 5 (1.70 %); $N = 293$	0	1
Any Abuse	111 (82.84 %); 23 (17.16 %)	118 (73.75 %); 42 (26.25 %)	229 (77.89 %); 65 (22.11 %); $N = 294$	2.99	0.084
Mood Stabilizers	133 (99.25 %); 1 (0.75 %)	159 (99.38 %); 1 (0.62 %)	292 (99.32 %); 2 (0.68 %); $N = 294$	0	1
Antipsychotics	130 (97.01 %); 4 (2.99 %)	136 (85.00 %); 24 (15.00 %)	266 (90.48 %); 28 (9.52 %); $N = 294$	10.86	<0.001
Benzodiazepines	85 (63.43 %); 49 (36.57 %)	51 (31.87 %); 109 (68.12 %)	136 (46.26 %); 158 (53.74 %); $N = 294$	27.96	<0.001
Somatic disorders	82 (61.19 %); 52 (38.81 %)	82 (51.25 %); 78 (48.75 %)	164 (55.78 %); 130 (44.22 %); $N = 294$	2.53	0.111
SSRI	65 (48.51 %); 69 (51.49 %)	91 (57.23 %); 68 (42.77 %)	156 (53.06 %); 137 (46.60 %); $N = 293$	1.89	0.17
SNRI	103 (76.87 %); 31 (23.13 %)	119 (74.84 %); 40 (25.16 %)	222 (75.51 %); 71 (24.15 %); $N = 293$	0.07	0.79
NRI	132 (98.51 %); 2 (1.49 %)	151 (94.97 %); 8 (5.03 %)	283 (96.26 %); 10 (3.40 %); $N = 293$	1.79	0.181
TCA	133 (99.25 %); 1 (0.75 %)	148 (93.08 %); 11 (6.92 %)	281 (95.58 %); 12 (4.08 %); $N = 293$	5.57	0.018
NDRI	132 (98.51 %); 2 (1.49 %)	151 (94.97 %); 8 (5.03 %)	283 (96.26 %); 10 (3.40 %); $N = 293$	1.79	0.181
NASSA	112 (83.58 %); 22 (16.42 %)	140 (88.05 %); 19 (11.95 %)	252 (85.71 %); 41 (13.95 %); $N = 293$	0.86	0.353
SARI	130 (97.01 %); 4 (2.99 %)	149 (93.71 %); 10 (6.29 %)	279 (94.90 %); 14 (4.76 %); $N = 293$	1.09	0.296
Other antidepressants	129 (96.27 %); 5 (3.73 %)	157 (98.74 %); 2 (1.26 %)	286 (97.28 %); 7 (2.38 %); $N = 293$	0.99	0.319
Combination treatment	59 (44.03 %); 75 (55.97 %)	8 (5.03 %); 151 (94.97 %)	67 (22.79 %); 226 (76.87 %); $N = 293$	60.51	<0.001
Age	47.24 ± 14.33	53.40 ± 14.33	50.59 ± 14.66; $N = 294$	−3.66	<0.001
MADRS Retrospective	36.46 ± 5.44	40.16 ± 6.54	38.47 ± 6.34; $N = 294$	−5.28	<0.001
Education (lower to higher)	2.92 ± 0.98	2.99 ± 1.03	2.96 ± 1.01; $N = 292$	−0.58	0.56
Working status (lower to higher)	2.11 ± 1.31	2.42 ± 1.17	2.28 ± 1.25; $N = 294$	−2.09	0.038
Bmi	26.15 ± 6.08	25.99 ± 4.85	26.06 ± 5.45; $N = 293$	0.24	0.813
Sheehan Work	5.89 ± 3.48	7.65 ± 2.53	6.85 ± 3.12; $N = 289$	−4.81	<0.001
Sheehan Social	6.06 ± 3.07	7.88 ± 1.60	7.05 ± 2.55; $N = 293$	−6.18	<0.001
Sheehan Family	5.84 ± 3.04	7.92 ± 1.61	6.97 ± 2.59; $N = 293$	−7.1	<0.001
Apparent sadness MADRS retrospective	4.43 ± 0.79	5.05 ± 0.84	4.77 ± 0.87; $N = 294$	−6.54	<0.001
Reported sadness MADRS retrospective	4.36 ± 0.91	5.11 ± 0.86	4.77 ± 0.96; $N = 294$	−7.19	<0.001
Inner tension MADRS retrospective	3.87 ± 1.07	4.30 ± 1.17	4.11 ± 1.15; $N = 294$	−3.25	0.001
Reduced sleep MADRS retrospective	3.48 ± 1.48	3.69 ± 1.75	3.59 ± 1.63; $N = 294$	−1.11	0.267
Reduced appetite MADRS retrospective	2.54 ± 1.59	2.51 ± 1.75	2.53 ± 1.68; $N = 294$	0.17	0.869
Concentration difficulties MADRS retrospective	3.81 ± 1.11	4.25 ± 1.07	4.05 ± 1.11; $N = 294$	−3.46	<0.001
Lassitude MADRS retrospective	4.00 ± 0.99	4.36 ± 1.04	4.19 ± 1.03; $N = 294$	−2.99	0.003
Inability to feel MADRS retrospective	3.75 ± 1.14	4.24 ± 0.89	4.01 ± 1.04; $N = 294$	−4.03	<0.001
Pessimistic thoughts MADRS retrospective	3.61 ± 1.02	3.72 ± 1.09	3.67 ± 1.06; $N = 294$	−0.86	0.389
Suicidal thoughts MADRS retrospective	2.61 ± 1.77	2.94 ± 1.88	2.79 ± 1.84; $N = 294$	−1.52	0.129
Duration Current Episode	194.48 ± 161.45	285.33 ± 168.24	243.78 ± 171.26; $N = 293$	−4.69	<0.001
Number of Depressive Episodes	3.17 ± 2.54	4.38 ± 3.02	3.82 ± 2.87; $N = 274$	−3.58	<0.001
Age at Onset	34.19 ± 14.42	33.85 ± 14.42	34.01 ± 14.42; $N = 288$	0.2	0.844
Duration of Disease	13.16 ± 12.29	19.37 ± 13.27	16.52 ± 13.19; $N = 287$	−4.1	<0.001
Time of hospitalizations	7.67 ± 22.17	10.93 ± 19.29	9.31 ± 20.83; $N = 266$	−1.27	0.205
Age at first hospitalization	40.85 ± 13.11	42.12 ± 15.55	41.57 ± 14.56; $N = 195$	−0.62	0.539
Suicidal Risk (lower to higher)	1.78 ± 0.81	2.25 ± 0.87	2.03 ± 0.87; $N = 293$	−4.83	<0.001
psychiatric side effects	0.47 ± 0.23	0.57 ± 0.35	0.53 ± 0.31; $N = 294$	−3.04	0.003
neurologic side effects	0.04 ± 0.13	0.06 ± 0.16	0.05 ± 0.15; $N = 294$	−1.13	0.258
autonomic side effects	0.16 ± 0.25	0.17 ± 0.27	0.17 ± 0.26; $N = 294$	−0.53	0.6
other side effects	0.84 ± 0.54	0.62 ± 0.56	0.72 ± 0.56; $N = 294$	3.46	<0.001
Side Effects Total	1.03 ± 0.42	0.96 ± 0.45	0.99 ± 0.44; $N = 294$	1.37	0.171
Antidepressant dose equivalents Total	42.64 ± 22.79	49.38 ± 23.54	46.39 ± 23.45; $N = 248$	−2.27	0.024

applied XGBoost machine learning to investigate clinical and demographic predictors of antidepressant treatment outcome. Our results extend previous GSRD findings (Bartova et al., 2019; Kautzky et al., 2017b). Overall the evaluation metrics of the model are in line or superior when compared to previous studies, as well as many of the predicting variables (Chekroud et al., 2016; Curtiss et al., 2024; Nie et al., 2018; Nunez et al., 2021; Perlis, 2013).

A relevant result of our current analysis is the importance of more severe chronicity-related indices, duration of current episode, duration of disease, time of hospitalizations and number of depressive episodes, in identifying individuals at high risk of nonresponse or TRD. In our previous analysis, which was based on TRD III sample only, longer disease duration was among the strongest predictors for treatment resistance. Other studies have suggested that longer illness duration and more frequent or protracted depressive episodes lead to greater chronicity and worse long-term outcomes (Gilmer et al., 2008; Rush et al., 2012). Our findings further confirm this perspective and give further support, via a larger sample and a more flexible machine-learning approach, to the notion that accumulated burden of disease is a major driver of TRD risk.

Another important predictor that consistently emerged in our model is BMI, though not in univariate analyses, confirming findings from the previous literature that weight or metabolic status might modulate antidepressant response (Kraus et al., 2023). It is possible that BMI reflects complex interactions with medical comorbidities such as metabolic syndrome or type 2 diabetes, conditions that have been linked with poor response (Falola et al., 2017; Possidente et al., 2023). Although direct causal relationships remain unclear, these converging data underscore that evaluating physical health (e.g., BMI, metabolic status) is a worthwhile addition for risk stratification algorithms in patients with depression.

Severity indicators, such as higher baseline MADRS and suicidal risk, were also among the higher predictive factors and were replicated in the TRD IV sample, in line with evidence in our previous samples (Bartova et al., 2019) and in literature (De Carlo et al., 2016). However some prior clinical features with well-documented associations to TRD were less evident in the present model. For instance, substance abuse and somatic disorders (e.g., thyroid dysfunction, diabetes) contributed minimally to the predictions, despite some evidence in the literature on their role as risk factors (Chen et al., 2024; De Carlo et al., 2016; Serretti, 2024). It is possible that such effects become diluted when a large combination of variables are considered simultaneously in a machine-learning framework, particularly when robust correlates of

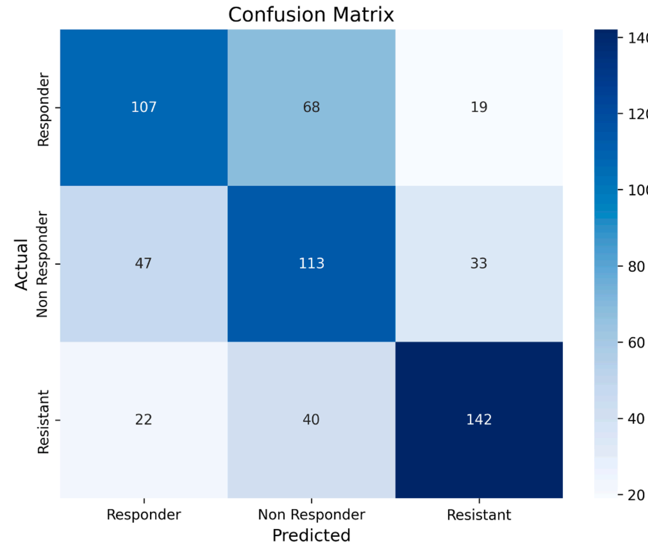

Fig. 2. XGBoost confusion matrix.

chronic courses are considered (e.g., episode duration, hospitalization), and the exclusion of severe abuse and dependence may reduce their impact. This may explain the discrepancy we observed in our results for some variables in univariate versus XGBoost results.

Consistent with prior GSRD studies (Panariello et al., 2023), more side effects reported during antidepressant therapy were associated with resistance though not included in the XGBoost model because it was a post baseline treatment. However, despite the fact that more complex treatment are prescribed to patients non responding thus leading to

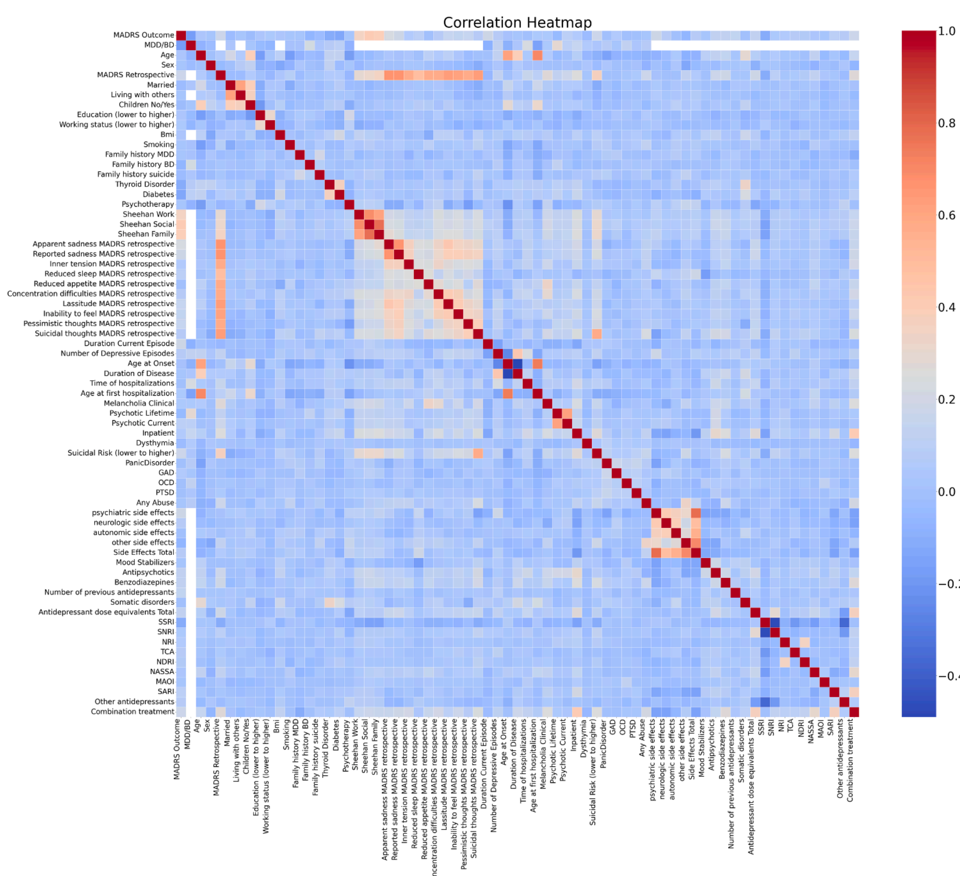

Fig. 1. Correlation heatmap of the complete set of variables in the study.

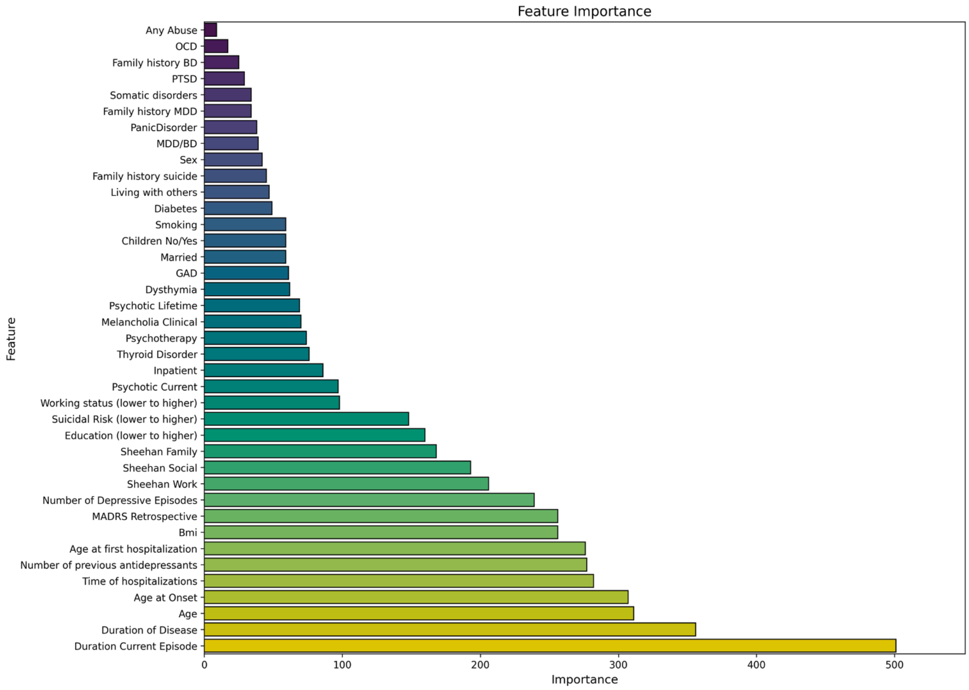

Fig. 3. XGBoost feature importance.

more side effects, it could also be hypothesized that patients who develop multiple or severe medication side effects are sometimes less able to sustain adequate treatment trials with compliance (Masand, 2003), thereby increasing their likelihood of incomplete response and resistance. This finding underscores the utility of systematically monitoring side effects throughout treatment. In some cases, optimizing side-effect profiles (e.g., switching to better-tolerated antidepressants, ensuring adequate management of sedation or weight gain) may improve treatment adherence, thereby reducing the likelihood of incomplete response and subsequent resistance. We also confirmed in the new TRD IV sample the predictive value of functional impairment via the SDS. In particular, Sheehan Work and Sheehan Family scores again emerged as relevant, in line with previous evidence (Curtiss et al., 2024).

When considering specific symptoms of depression, not included in the XGBoost model for collinearity with the total MADRS score but reported in the tables, insomnia was not identified as a major predictor of treatment response, even if in TRD III reduced sleep was associated with TRD and a non-significant trend was found in TRD IV (supplementary Tables 3 and 4). This is in line with its potential impact on antidepressant efficacy (Asarnow and Manber, 2019; Lustberg and Reynolds, 2000). In our sample reduced sleep was inversely correlated with chronicity measures, possibly mediated also by more sedative treatments as shown by the univariate negative correlations. Future analyses should include specific measures of sleep disturbances, such as the presence of insomnia at baseline or changes in sleep quality over the course of treatment, to better assess their potential predictive value. Additionally, as sleep disturbances may interact with other clinical features (e.g., anhedonia, suicidality, and anxiety), a more granular analysis of these relationships could further clarify insomnia's role in treatment outcomes. Anhedonia, defined using MADRS item 8, was the strongest single symptom predictor, together with suicidal thoughts and risk, confirming previous observations also in the new TRD IV sample (Luca et al., 2024; Serretti, 2023; Wong et al., 2024).

Anxiety disorders (GAD, panic disorder, OCD) and PTSD displayed comparatively modest contributions, except for a mild association with panic disorder in univariate tests. This partially diverges from previous studies (De Carlo et al., 2016; Fava et al., 2008) that identified anxiety comorbidities as robust correlates of worse response. However, our findings are in line with our previous analysis (Kautzky et al., 2017b) where these comorbidities, although directionally associated with TRD, were overshadowed by variables capturing severity, chronicity, and functional impairment once complex interactions were accounted for. The same may be suggested for melancholic features, as they showed mixed associations with TRD (Dold et al., 2021; Szmulewicz et al., 2024). Our results thus may suggest that single predictors such as anxiety have inconsistent effects: while they may increase risk when considered alone, they appear less consistently predictive than measures of long-standing depression burden and overall severity.

The prominence of duration of current episode and duration of disease underscores that chronic, entrenched depressive states are the most difficult to treat, possibly because of sustained neurobiological changes (Runia et al., 2022). These findings highlight the importance of early intervention to minimize recurrent episodes and accelerate remission (Fekadu et al., 2009b). Clinicians may therefore consider intensifying or augmenting treatments promptly in individuals with a long history of depression or extended ongoing episodes (Scala et al., 2023).

A clear strength of this study is the large sample size comprising multiple distinct cohorts (TRD I–II, TRD III, and the newly recruited TRD IV) assessed with a common protocol. The sample size allowed us to not use cross-validation, a procedure that may decrease power (Arlot and Celisse, 2010; Brykov et al., 2020; Just et al., 2020; Zhang et al., 2021). The addition of a new sample not previously reported further enhances generalizability relative to earlier GSRD work, which featured fewer patients or fewer variables. The use of XGBoost is another advantage: gradient-boosted decision trees accommodate complex, non-linear interactions among variables while directly handling missing data without imputation or sample size reduction.

Conversely, several limitations should be considered. First, although XGBoost handles missingness more flexibly than standard imputation, the reliance on multi-center retrospective data inevitably produces variability in how certain variables were recorded (e.g., in earlier waves, HAMD rather than MADRS was used). Second, the cross-sectional naturalistic design did not standardize antidepressant regimens beyond requiring an "adequate dose for $\geq$4 weeks," so variation in medication classes, augmentation strategies, and length of follow-up could influence outcome classification; the naturalistic nature of our study prevented us from using treatment as a predictor, having resistant patients usually a more complex treatment. In addition, the MADRS to assess symptoms at baseline was administered retrospectively. Therefore, we cannot exclude the possibility of recall bias, though retrospective depression assessment has been demonstrated to be substantially reliable (Birk et al., 2020). Third, while the machine-learning model showed good overall performance, the misclassification rate for responders and non-responders remains non-negligible, underscoring that further refinements, e.g., combining clinical variables with biomarkers or repeated measures of early symptom change, may be necessary to maximize predictive accuracy (Fabbri et al., 2020; Perlman et al., 2019). Fourth, the selection of variables to include in the model is crucial. We included all baseline variables, but not all potentially influential clinical variables were available (e.g. stressful life events, personality, neurocognitive evaluations). Finally, our external validation was limited to validation in distinct subsamples from the same GSRD cohorts. Replication in completely independent cohorts is a key next step to assess how well these results generalize to other populations. However, the satisfactory replication in the newly collected and independent TRD IV sample suggests a good potential for generalizability, also when the smaller, but more homogeneous, TRD III sample was used as training. This secondary analysis is supporting the lack of bias of including variables with relatively high missing values (supplementary tables). In any case the use of the same exact model for future replications is conceptually challenging considering the need of completely overlapping measures and scoring, with a negative impact on power (Nunez et al., 2021).

In conclusion, our findings indicate that clinical measures of disease chronicity, severity, functional impairment, and BMI are the most robust predictors of antidepressant (non)response and TRD across multiple cohorts. These results confirm and refine earlier observations and emphasize the need for vigilant, multimodal management, especially in individuals with long-standing depression and high functional disability. Future work should concentrate on (1) prospective (rather than purely retrospective) validation in large, well-characterized samples or large population studies (2) integration of biological or genetic markers into clinical models to address potential subtypes, and (3) more fine-grained measures of comorbidities and other potentially modulating factors not included in the present study. Such advances will help clinicians to identify TRD risk earlier and tailor interventions more precisely, ultimately improving long-term outcomes for patients.

CRediT authorship contribution statement

Alessandro Serretti: Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Siegfried Kasper:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Lucie Bartova:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Joseph Zohar:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Daniel Souery:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. **Stuart Montgomery:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Panagiotis Ferentinos:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Dan Rujescu:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Alexander Kautzky:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Francesco Attanasio:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Raffaella Zanardi:** Data curation, Investigation, Methodology, Project administration, Resources, Writing – review & editing. **Chiara Fabbri:** Methodology, Writing – review & editing. **Bernhard T Baune:** Methodology, Writing – review & editing. **Raffaele Ferri:** Methodology, Writing – review & editing. **Julien Mendlewicz:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

Dr. Rujescu served as consultant for Janssen, received honoraria from Boehringer-Ingelheim, Gerot Lannacher, Janssen and Pharmagenetix, received research/ travel support from Angelini, Boehringer-Ingelheim, Janssen and Schwabe, and served on advisory boards of AC Immune, Boehringer-Ingelheim, Roche and Rovi. Dr. Souery has received grant/research support from GlaxoSmithKline and Lundbeck; and he has served as a consultant or on advisory boards for AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Janssen, and Lundbeck. Dr. Mendlewicz is a member of the board of the Lundbeck International Neuroscience Foundation and of the advisory board of Servier. Dr. Zohar has received grant/research support from Lundbeck, Servier, and Pfizer; he has served as a consultant or on the advisory boards for Servier, Pfizer, Solvay, and Actelion; and he has served on speakers' bureaus for Lundbeck, GlaxoSmithKline, Jazz, and Solvay. Dr. Montgomery has served as a consultant or on advisory boards for AstraZeneca, Bionevia, Bristol-Myers Squibb, Forest, GlaxoSmithKline, Grunenthal, Intellect Pharma, Johnson & Johnson, Lilly, Lundbeck, Merck, Merz, M's Science, Neurim, Otsuka, Pierre Fabre, Pfizer, Pharmaneuroboost, Richter, Roche, Sanofi, Sepracor, Servier, Shire, Synosis, Takeda, Theracos, Targacept, Transcept, UBC, Xytis, and Wyeth. Dr. Serretti has served as a consultant or speaker for Abbott, Abbvie, Angelini, AstraZeneca, Clinical Data, Boehringer, Bristol-Myers Squibb, Eli Lilly, GlaxoSmithKline, Innovapharma, Italfarmaco, Janssen, Lundbeck, Naurex, Pfizer, Polifarma, Sanofi, and Servier and Taliaz. Dr. Kasper has received grant/ research support from Lundbeck; he has served as a consultant or on advisory boards for Angelini, Biogen, Boehringer-Ingelheim, Esai, Janssen, IQVIA, Lundbeck, Mylan, Recordati, Rovi, Sage and Schwabe; and he has served on speakers bureaus for Aspen Farmaceutica S.A., Angelini, Biogen, Janssen, Lundbeck, Neuraxpharma, Recordati, Sage, Sanofi, Schwabe, Servier and Sun Pharma. Dr. Baune received honoraria for serving as a consultant or on advisory boards unrelated to the present work for Angelini, AstraZeneca, Biogen, Boehringer Ingelheim, Bristol-Meyers Squibb, Janssen, LivaNova, Lundbeck, Medscape, Neurotorium, Novartis, Otsuka, Pfizer, Recordati, Roche, Rovi, Sanofi, Servier, Teva. Dr. Bartova has received travel grants and/or consultant/speaker honoraria from Market Access Transformation, Alpine Market Research, Medizin Medien Austria, Universimed, Vertretungsnetz, Diagnosia, Dialectica, EQT, IQVIA, AOP Orphan, Schwabe, Janssen (Johnson & Johnson), Angelini, Lundbeck, Novartis, Biogen, Takeda. The other authors declare no potential conflicts of interest.

Funding

The European Group for the Study of Resistant Depression (GSRD) obtained an unrestricted grant sponsored by Lundbeck A/S. The sponsor played no role in designing the study, data collection and analyses, interpretation of the data, writing of the manuscript, and in the decision to submit the research for publication.

Acknowledgments

Alessandro Serretti, Chiara Fabbri and Bernhard T Baune received support from Psych-STRATA, a project funded from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101057454. The European Group for the Study of Resistant Depression (GSRD) obtained an unrestricted grant sponsored by Lundbeck A/S. The sponsor played no role in designing the study, data collection and analyses, interpretation of the data, writing of the manuscript, and in the decision to submit the research for publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroneuro.2025.06.011.

References

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surv. 4, 40–79.

Asarnow, L.D., Manber, R., 2019. Cognitive behavioral therapy for insomnia in depression. Sleep Med. Clin. 14, 177–184.

Aydin, Z.E., Ozturk, Z.K., 2021. Performance analysis of XGBoost classifier with missing data, in: performance analysis of XGBoost classifier with missing data. Presented at the 1st int. Conf. Comput. Mach. Intell.

Bartova, L., Dold, M., Kautzky, A., Fabbri, C., Spies, M., Serretti, A., Souery, D., Mendlewicz, J., Zohar, J., Montgomery, S., Schosser, A., Kasper, S., 2019. Results of the European group for the study of resistant depression (GSRD) - basis for further research and clinical practice. World J. Biol. Psychiatry 20, 427–448.

Beaulieu-Jones, B.K., Villamar, M.F., Scordis, P., Bartmann, A.P., Ali, W., Wissel, B.D., Alsentzer, E., de Jong, J., Patra, A., Kohane, I., 2023. Predicting seizure recurrence after an initial seizure-like episode from routine clinical notes using large language models: a retrospective cohort study. Lancet Digit Health 5, e882–e894.

Birk, S.L., Olino, T.M., Klein, D.N., Seeley, J.R., 2020. Validity of retrospectively-reported depressive episodes. J. Affect. Disord. 277, 908–913.

Bishara, A., Chiu, C., Whitlock, E.L., Douglas, V.C., Lee, S., Butte, A.J., Leung, J.M., Donovan, A.L., 2022. Postoperative delirium prediction using machine learning models and preoperative electronic health record data. BMC Anesthesiol. 22, 8.

Brykov, M.N., Petryshynets, I., Pruncu, C.I., Efremenko, V.G., Pimenov, D.Y., Giasin, K., Sylenko, S.A., Wojciechowski, S., 2020. Machine learning modelling and feature engineering in seismology experiment. Sensors 20, 4228.

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R., 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 3, 243–250.

Chen, L.C., Chen, M.H., Bai, Y.M., Chen, T.J., Su, T.P., 2024. Resistance to antidepressant treatment among patients with major depressive disorder: a nationwide study. Int. Clin. Psychopharmacol. https://doi.org/10.1097/YIC.0000000000000574.

Chen, S., Xu, C., 2023. Handling high-dimensional data with missing values by modern machine learning techniques. J. Appl. Stat. 50, 786–804.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794.

Curtiss, J., Smoller, J.W., Pedrelli, P., 2024. Optimizing precision medicine for second-step depression treatment: a machine learning approach. Psychol. Med. 1–8.

De Carlo, V., Calati, R., Serretti, A., 2016. Socio-demographic and clinical predictors of non-response/non-remission in treatment resistant depressed patients: a systematic review. Psychiatry Res. 240, 421–430.

Dold, M., Bartova, L., Fugger, G., Kautzky, A., Mitschek, M.M.M., Fabbri, C., Montgomery, S., Zohar, J., Souery, D., Mendlewicz, J., Serretti, A., Kasper, S., 2021. Melancholic features in major depression - a European multicenter study. Prog. Neuropsychopharmacol. Biol. Psychiatry 110, 110285.

Dold, M., Kautzky, A., Bartova, L., Rabl, U., Souery, D., Mendlewicz, J., Porcelli, S., Serretti, A., Zohar, J., Montgomery, S., Kasper, S., 2016. Pharmacological treatment strategies in unipolar depression in European tertiary psychiatric treatment centers - A pharmacoepidemiological cross-sectional multicenter study. Eur. Neuropsychopharmacol. 26, 1960–1971.

Fabbri, C., Kasper, S., Kautzky, A., Zohar, J., Souery, D., Montgomery, S., Albani, D., Forloni, G., Ferentinos, P., Rujescu, D., Mendlewicz, J., Uher, R., Lewis, C.M., Serretti, A., 2020. A polygenic predictor of treatment-resistant depression using whole exome sequencing and genome-wide genotyping. Transl. Psychiatry 10, 50.

Falola, M.I., Limdi, N., Shelton, R.C., 2017. Clinical and genetic predictors of delayed remission after multiple levels of antidepressant treatment: toward early identification of depressed individuals for advanced care options. J. Clin. Psychiatry 78, e1291–e1298.

Fava, M., Rush, A.J., Alpert, J.E., Balasubramani, G.K., Wisniewski, S.R., Carmin, C.N., Biggs, M.M., Zisook, S., Leuchter, A., Howland, R., Warden, D., Trivedi, M.H., 2008. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. Am. J. Psychiatry.

Fekadu, A., Wooderson, S., Donaldson, C., Markopoulou, K., Masterson, B., Poon, L., Cleare, A.J., 2009a. A multidimensional tool to quantify treatment resistance in depression: the Maudsley staging method. J. Clin. Psychiatry 70, 177–184.

Fekadu, A., Wooderson, S.C., Markopoulo, K., Donaldson, C., Papadopoulos, A., Cleare, A.J., 2009b. What happens to patients with treatment-resistant depression? A systematic review of medium to long term outcome studies. J. Affect. Disord. 116, 4–11.

Feng, Y., Feng, Z., Wang, L., Lv, W., Liu, Z., Min, X., Li, J., Zhang, J., 2023. Comparison and analysis of multiple machine learning models for discriminating benign and malignant testicular lesions based on magnetic resonance imaging radiomics. Front. Med. 10, 1279622.

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392, 1789–1858.

Gilmer, W.S., Gollan, J.K., Wisniewski, S.R., Howland, R.H., Trivedi, M.H., Miyahara, S., Fleck, J., Thase, M.E., Alpert, J.E., Nierenberg, A.A., Warden, D., Fava, M., Rush, A. J., 2008. Does the duration of index episode affect the treatment outcome of major depressive disorder? A STAR*D report. J. Clin. Psychiatry 69, 1246–1256.

Husain, M.M., Rush, A.J., Wisniewski, S.R., McClintock, S.M., Fava, M., Nierenberg, A.A., Davis, L., Balasubramani, G.K., Young, E., Albala, A.A., Trivedi, M.H., 2009. Family history of depression and therapeutic outcome: findings from STAR*D. J. Clin. Psychiatry 70, 185–195.

Jain, F.A., Hunter, A.M., Brooks 3rd, J.O., Leuchter, A.F., 2013. Predictive socioeconomic and clinical profiles of antidepressant response and remission. Depress. Anxiety 30, 624–630.

Jakubovski, E., Bloch, M.H., 2014. Prognostic subgroups for citalopram response in the STAR*D trial. J. Clin. Psychiatry 75, 738–747.

Jha, M.K., South, C., Trivedi, J., Minhajuddin, A., Rush, A.J., Trivedi, M.H., 2019. Prediction of acute-phase treatment outcomes by adding a single-item measure of activity impairment to symptom measurement: development and validation of an interactive calculator from the STAR*D and CO-MED trials. Int. J. Neuropsychopharmacol. 22, 339–348.

Just, A.C., Arfer, K.B., Rush, J., Dorman, M., Shtein, A., Lyapustin, A., Kloog, I., 2020. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM2.5) using satellite data over large regions. Atmos. Environ. 239, 117649, 1994.

Kautzky, A., Baldinger-Melich, P., Kranz, G.S., Vanicek, T., Souery, D., Montgomery, S., Mendlewicz, J., Zohar, J., Serretti, A., Lanzenberger, R., Kasper, S., 2017a. A new prediction model for evaluating treatment-resistant depression. J. Clin. Psychiatry 78, 215–222.

Kautzky, A., Dold, M., Bartova, L., Spies, M., Vanicek, T., Souery, D., Montgomery, S., Mendlewicz, J., Zohar, J., Fabbri, C., Serretti, A., Lanzenberger, R., Kasper, S., 2017b. Refining prediction in treatment-resistant depression: results of machine learning analyses in the TRD III sample. J. Clin. Psychiatry 79. https://doi.org/10.4088/JCP.16m11385.

Kraus, C., Kautzky, A., Watzal, V., Gramser, A., Kadriu, B., Deng, Z.D., Bartova, L., Zarate, C.A., Lanzenberger, Jr, Souery, R., Montgomery, D., Mendlewicz, S., Zohar, J., Fanelli, J., Serretti, G., Kasper, S, A., 2023. Body mass index and clinical outcomes in individuals with major depressive disorder: findings from the GSRD European multicenter database. J. Affect. Disord. 335, 349–357.

Latief, M.A., Bustamam, A., Siswantining, T., 2020. Performance evaluation XGBoost in handling missing value on classification of hepatocellular carcinoma gene expression data. In: 2020 4th International Conference on Informatics and Computational Sciences. IEEE, pp. 1–6. ICICoS.

Liu, P., Li, X.J., Zhang, T., Huang, Y.H., 2024. Comparison between XGboost model and logistic regression model for predicting sepsis after extremely severe burns. J. Int. Med. Res. 52, 3000605241247696.

Luca, A., Luca, M., Kasper, S., Pecorino, B., Zohar, J., Souery, D., Montgomery, S., Ferentinos, P., Rujescu, D., Messina, A., Zanardi, R., Ferri, R., Tripodi, M., Baune, B. T., Fanelli, G., Fabbri, C., Mendlewicz, J., Serretti, A., 2024. Anhedonia is associated with a specific depression profile and poor antidepressant response. Int. J. Neuropsychopharmacol. 27. https://doi.org/10.1093/ijnp/pyae055.

Lustberg, L., Reynolds, C.F., 2000. Depression and insomnia: questions of cause and effect. Sleep Med. Rev. 4, 253–262.

Mandelli, L., Serretti, A., Souery, D., Mendlewicz, J., Kasper, S., Montgomery, S., Zohar, J., 2016. High occupational level is associated with poor response to treatment of depression. Eur. Neuropsychopharmacol. 26, 1320–1326.

Masand, P.S., 2003. Tolerability and adherence issues in antidepressant therapy. Clin. Ther. 25, 2289–2304.

McIntyre, R.S., Alsuwaidan, M., Baune, B.T., Berk, M., Demyttenaere, K., Goldberg, J.F., Gorwood, P., Ho, R., Kasper, S., Kennedy, S.H., Ly-Uson, J., Mansur, R.B., McAllister-Williams, R.H., Murrough, J.W., Nemeroff, C.B., Nierenberg, A.A., Rosenblat, J.D., Sanacora, G., Schatzberg, A.F., Shelton, R., Stahl, S.M., Trivedi, M.H., Vieta, E., Vinberg, M., Williams, N., Young, A.H., Maj, M., 2023. Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. World Psychiat 22, 394–412.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389.

Mortazavi, B.J., Bucholz, E.M., Desai, N.R., Huang, C., Curtis, J.P., Masoudi, F.A., Shaw, R.E., Negahban, S.N., Krumholz, H.M., 2019. Comparison of machine learning methods with national cardiovascular data registry models for prediction of risk of bleeding after percutaneous coronary intervention. JAMA Netw Open 2, e196835.

Nie, Z., Vairavan, S., Narayan, V.A., Ye, J., Li, Q.S., 2018. Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study. PLoS One 13, e0197268.

Nijman, S., Leeuwenberg, A.M., Beekers, I., Verkouter, I., Jacobs, J., Bots, M.L., Asselbergs, F.W., Moons, K., Debray, T., 2022. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. J. Clin. Epidemiol. 142, 218–229.

Nunez, J.J., Nguyen, T.T., Zhou, Y., Cao, B., Ng, R.T., Chen, J., Frey, B.N., Milev, R., Müller, D.J., Rotzinger, S., Soares, C.N., Uher, R., Kennedy, S.H., Lam, R.W., 2021. 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 16, e0253023.

Panariello, F., Kasper, S., Zohar, J., Souery, D., Montgomery, S., Ferentinos, P., Rujescu, D., Mendlewicz, J., De Ronchi, D., Serretti, A., Fabbri, C., 2023.

Characterisation of medication side effects in patients with mostly resistant depression in a real-world setting. World J. Biol. Psychiatry 24, 439–448.
Perlis, R.H., 2013. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. Biol. Psychiatry 74, 7–14.
Perlman, K., Benrimoh, D., Israel, S., Rollins, C., Brown, E., Tunteng, J.F., You, R., You, E., Tanguay-Sela, M., Snook, E., Miresco, M., Berlim, M.T., 2019. A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. J. Affect. Disord. 243, 503–515.
Possidente, C., Fanelli, G., Serretti, A., Fabbri, C., 2023. Clinical insights into the cross-link between mood disorders and type 2 diabetes: a review of longitudinal studies and mendelian randomisation analyses. Neurosci. Biobehav. Rev. 152, 105298.
Quinn, T.P., Hess, J.L., Marshe, V.S., Barnett, M.M., Hauschild, A.C., M]aciukiewicz, M., Elsheikh, S.S.M., Men, X., Schwarz, E., Trakadis, Y.J., Breen, M.S., Barnett, E.J., Zhang-James, Y., Ahsen, M.E., Cao, H., Chen, J., Hou, J., Salekin, A., Lin, P.I., Nicodemus, K.K., Meyer-Lindenberg, A., Bichindaritz, I., Faraone, S.V., Cairns, M.J., Pandey, G., Müller, D.J., Glatt, S.J., Machine Learning in Psychiatry (MLPsych) Consortium, 2024. A primer on the use of machine learning to distil knowledge from data in biological psychiatry. Mol. Psychiatry 29, 387–401.
Runia, N., Yücel, D.E., Lok, A., de Jong, K., Denys, D.A.J.P., van Wingen, G.A., Bergfeld, I.O., 2022. The neurobiology of treatment-resistant depression: a systematic review of neuroimaging studies. Neurosci. Biobehav. Rev. 132, 433–448.
Rusdah, D.A., Murfi, H., 2020. XGBoost in handling missing values for life insurance risk prediction. SN Appl. Sci. 2, 1336.
Rush, A.J., Wisniewski, S.R., Zisook, S., Fava, M., Sung, S.C., Haley, C.L., Chan, H.N., Gilmer, W.S., Warden, D., Nierenberg, A.A., Balasubramani, G.K., Gaynes, B.N., Trivedi, M.H., Hollon, S.D., 2012. Is prior course of illness relevant to acute or longer-term outcomes in depressed out-patients? A STAR*D report. Psychol. Med 42, 1131–1149.
Salem, H., Huynh, T., Topolski, N., Mwangi, B., Trivedi, M.H., Soares, J.C., Rush, A.J., Selvaraj, S., 2023. Temporal multi-step predictive modeling of remission in major depressive disorder using early stage treatment data; STAR*D based machine learning approach. J. Affect. Disord. 324, 286–293.
Scala, M., Fanelli, G., De Ronchi, D., Serretti, A., Fabbri, C., 2023. Clinical specificity profile for novel rapid acting antidepressant drugs. Int. Clin. Psychopharmacol. 38, 297–328.
Serretti, A., 2024. Mood disorders and somatic comorbidities. Int. Clin. Psychopharmacol. 39, 291–293.
Serretti, A., 2023. Anhedonia and depressive disorders. Clin. Psychopharmacol. Neurosci. 21, 401–409.
Sforzini, L., Worrell, C., Kose, M., Anderson, I.M., Aouizerate, B., Arolt, V., Bauer, M., Baune, B.T., Blier, P., Cleare, A.J., Cowen, P.J., Dinan, T.G., Fagiolini, A., Ferrier, I. N., Hegerl, U., Krystal, A.D., Leboyer, M., McAllister-Williams, R.H., McIntyre, R.S., Meyer-Lindenberg, A., Miller, A.H., Nemeroff, C.B., Normann, C., Nutt, D., Pallanti, S., Pani, L., Penninx, B.W.J.H., Schatzberg, A.F., Shelton, R.C., Yatham, L. N., Young, A.H., Zahn, R., Aislaitner, G., Butlen-Ducuing, F., Fletcher, C., Haberkamp, M., Laughren, T., Mäntylä, F.L., Schruers, K., Thomson, A., Arteaga-Henríquez, G., Benedetti, F., Cash-Gibson, L., Chae, W.R., De Smedt, H., Gold, S.M., Hoogendijk, W.J.G., Mondragón, V.J., Maron, E., Martynowicz, J., Melloni, E., Otte, C., Perez-Fuentes, G., Poletti, S., Schmidt, M.E., van de Ketterij, E., Woo, K., Flossbach, Y., Ramos-Quiroga, J.A., Savitz, A.J., Pariante, C.M., 2022. A Delphi-method-based consensus guideline for definition of treatment-resistant depression for clinical trials. Mol. Psychiatry 27, 1286–1299.
Sheehan, D.V., Harnett-Sheehan, K., Raj, B.A., 1996. The measurement of disability. Int. Clin. Psychopharmacol. 11, 89–95.
Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59, 34–57. Suppl 20, 22-33quiz.
Solmi, M., Cortese, S., Vita, G., De Prisco, M., Radua, J., Dragioti, E., Köhler-Forsberg, O., Madsen, N.M., Rohde, C., Eudave, L., Aymerich, C., Pedruzo, B., Rodriguez, V., Rosson, S., Sabé, M., Hojlund, M., Catalan, A., de Luca, B., Fornaro, M., Ostuzzi, G., Barbui, C., Salazar-de-Pablo, G., Fusar-Poli, P., Correll, C.U., 2023. An umbrella review of candidate predictors of response, remission, recovery, and relapse across mental disorders. Mol. Psychiatry 28, 3671–3687.
Souery, D., Oswald, P., Massat, I., Bailer, U., Bollen, J., Demyttenaere, K., Kasper, S., Lecrubier, Y., Montgomery, S., Serretti, A., Zohar, J., Mendlewicz, J., Group for the Study of Resistant Depression, 2007. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. J. Clin. Psychiatry 68, 1062–1070.
Souery, D., Serretti, A., Calati, R., Oswald, P., Massat, I., Konstantinidis, A., Linotte, S., Kasper, S., Montgomery, S., Zohar, J., Mendlewicz, J., 2011. Citalopram versus desipramine in treatment resistant depression: effect of continuation or switching strategies. A randomized open study. World J. Biol. Psychiatry 12, 364–375.
Stewart, J.W., McGrath, P.J., Fava, M., Wisniewski, S.R., Zisook, S., Cook, I., Nierenberg, A.A., Trivedi, M.H., Balasubramani, G.K., Warden, D., Lesser, I., John Rush, A., 2010. Do atypical features affect outcome in depressed outpatients treated with citalopram? Int. J. Neuropsychopharmacol. 13, 15–30.
Szmulewicz, A., Valerio, M.P., Lomastro, J., Martino, D.J., 2024. Melancholic features and treatment outcome to selective serotonin reuptake inhibitors in major depressive disorder: a re-analysis of the STAR*D trial. J. Affect. Disord. 347, 101–107.
Tang, V.M., Yu, D., Weissman, C.R., Jones, B.D.M., Wang, G., Sloan, M.E., Blumberger, D. M., Daskalakis, Z.J., Le Foll, B., Voineskos, D., 2023. Treatment outcomes in major depressive disorder in patients with comorbid alcohol use disorder: a STAR*D analysis. J. Affect. Disord. 339, 691–697.
Turska, E., Jurga, S., Piskorski, J., 2021. Mood disorder detection in adolescents by classification trees, random forests and XGBoost in presence of missing data. Entropy 23. https://doi.org/10.3390/e23091210.
Uher, R., Perlis, R.H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M.Z., Souery, D., Bajs, M., Maier, W., Aitchison, K.J., Farmer, A., McGuffin, P., 2012. Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. Psychol. Med. 42, 967–980.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272.
Wang, Y., Wei, B., Zhao, T., Shen, H., Liu, X., Wang, J., Wang, Q., Shen, R., Feng, D., 2023. Machine learning-based prediction models for parathyroid carcinoma using pre-surgery cognitive function and clinical features. Sci. Rep. 13, 19007.
Wong, S., Le, G.H., Phan, L., Rhee, T.G., Ho, R., Meshkat, S., Teopiz, K.M., Kwan, A.T.H., Mansur, R.B., Rosenblat, J.D., McIntyre, R.S., 2024. Effects of anhedonia on health-related quality of life and functional outcomes in major depressive disorder: a systematic review and meta-analysis. J. Affect. Disord. 356, 684–698.
Zaninotto, L., Souery, D., Calati, R., Scudellari, P., Janiri, L., Montgomery, S., Kasper, S., Zohar, J., Mendlewicz, J., Serretti, A., 2014. Mixed, melancholic, and anxious features in depression: a cross-sectional study of sociodemographic and clinical correlates. Ann. Clin. Psychiatry 26, 243–253.
Zhang, F., Petersen, M., Johnson, L., Hall, J., O'Bryant, S.E., 2021. Accelerating hyperparameter tuning in machine learning for Alzheimer's disease with high performance computing. Front. Artif. Intell. 4, 798962.