Psychiatry Research 334 (2024) 115817

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Explainable multimodal prediction of treatment-resistance in patients with depression leveraging brain morphometry and natural language processing

Dong Yun Lee [a,b,1], Narae Kim [a,c,1], ChulHyoung Park [a,b], Sujin Gan [c], Sang Joon Son [d], Rae Woong Park [a,c,2,*], Bumhee Park [a,e,2,*]

[a] Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea
[b] Department of Medical Sciences, Graduate School of Ajou University, Suwon, South Korea
[c] Department of Biomedical Sciences, Graduate School of Ajou University, Suwon, South Korea
[d] Department of Psychiatry, Ajou University School of Medicine, Suwon, South Korea
[e] Office of Biostatistics, Medical Research Collaborating Center, Ajou Research Institute for Innovative Medicine, Ajou University Medical Center, Suwon, South Korea

ARTICLE INFO

Keywords:
Treatment-resistant depression
Machine learning
Brain imaging
Natural language processing

ABSTRACT

Although 20 % of patients with depression receiving treatment do not achieve remission, predicting treatment-resistant depression (TRD) remains challenging. In this study, we aimed to develop an explainable multimodal prediction model for TRD using structured electronic medical record data, brain morphometry, and natural language processing. In total, 247 patients with a new depressive episode were included. TRD-predictive models were developed based on the combination of following parameters: selected tabular dataset features, independent components-map weightings from brain T1-weighted magnetic resonance imaging (MRI), and topic probabilities from clinical notes. All models applied the extreme gradient boosting (XGBoost) algorithm via five-fold cross-validation. The model using all data sources showed the highest area under the receiver operating characteristic of 0.794, followed by models that used combined brain MRI and structured data, brain MRI and clinical notes, clinical notes and structured data, brain MRI only, structured data only, and clinical notes only (0.770, 0.762, 0.728, 0.703, 0.684, and 0.569, respectively). Classifications of TRD were driven by several predictors, such as previous exposure to antidepressants and antihypertensive medications, sensorimotor network, default mode network, and somatic symptoms. Our findings suggest that a combination of clinical data with neuroimaging and natural language processing variables improves the prediction of TRD.

1. Introduction

Depression is a severely debilitating disorder that is significantly associated with mortality, personal costs, and family burden (Kessler, 2012; Rush et al., 2006). Furthermore, management of depression is particularly challenging for patients showing an inadequate response to treatment (Pigott, 2015). Approximately 60 % of people with depression experience slight relief after being treated with antidepressants (Perlis, 2014). Given these limitations of pharmacological therapy, predicting the course of depression, such as treatment-resistant depression (TRD) (Dold and Kasper, 2017), is essential. Patients who fail to respond to multiple antidepressant treatments are classified as having TRD (Cepeda et al., 2018). Identifying TRD may be key to developing personalized depression treatment that can improve the efficiency of treatment planning and resource allocation as well as outcomes (Simon and Perlis, 2010).

Important risk factors for TRD previously identified include symptom severity, suicidal risk, psychotic symptoms, and comorbid anxiety disorder (Kautzky et al., 2019). Moreover, specific symptoms such as anhedonia, represent important prognostic indicators for TRD (McMakin et al., 2012). Previous studies have shown that both diagnosis and detailed symptoms need to be considered while assessing treatment resistance in clinical settings. Medical records accomplished by psychiatrists during treatment provide a good source of data for reviewing specific symptoms. However, the qualitative nature of these records has made quantitative analysis difficult. One method that can overcome

* Corresponding authors at: Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea.
E-mail addresses: veritas@ajou.ac.kr (R.W. Park), bhpark@ajou.ac.kr (B. Park).
[1] These authors contributed equally as co-first author.
[2] These authors contributed equally as co-corresponding author.

https://doi.org/10.1016/j.psychres.2024.115817
Received 7 August 2023; Received in revised form 19 February 2024; Accepted 23 February 2024
Available online 25 February 2024

these limitations is the recently developed automated natural language processing (NLP). NLP entails the analysis of words derived from qualitative records to generate quantitative data (Chowdhary, 2020). In support of this, NLP-derived predictors from narrative notes, such as depression, euthymic affect, and energy, have improved the predictive power for TRD (Perlis et al., 2012). Similarly, patient's detailed psychopathological symptoms can be statistically derived from their medical records to determine whether they are likely to progress to TRD. Furthermore, magnetic resonance imaging (MRI) of the brain is actively used in psychiatric clinical settings to differentiate organic causes. However, psychiatric disorders, such as depression, often involve no visible brain lesions, making the confirmation of significant changes through qualitative image analysis difficult. To overcome this problem, methods are being developed for predicting TRD by measuring structural regional brain volumes using quantitative image analysis techniques (Johnston et al., 2015; Phillips et al., 2015). Previous studies on the structural pattern of the brain have demonstrated that MRI data can potentially be used as predictive biomarkers for TRD (Johnston et al., 2015; Phillips et al., 2015).

In recent years, researchers have attempted to use machine learning techniques to identify and predict factors that contribute to treatment resistance in depression (Pigoni et al., 2019; Sajjadian et al., 2022). Because TRD is a complex and heterogeneous state with multiple causative mechanisms (McIntyre et al., 2014), machine learning techniques are viable approaches for addressing such a complex problem (Lichtenberg and Belmaker, 2010). Machine learning predictions have been developed using multiple modalities but predominantly clinical features (Pigoni et al., 2019). Some researchers have improved performance by incorporating variables from multiple modalities (Sajjadian et al., 2022; Bailey et al., 2019). The combination of clinical, molecular, and imaging variables has been reported to improve TRD prediction in comparison with single-modality measurements (Sajjadian et al., 2022). A previous study reported that analyzing electroencephalographic (EEG) variables along with mood measures resulted in a more tailored treatment response in patients with depression than considering EEG values alone (Bailey et al., 2019). Another study showed that the combined analysis of clinical features and genetic markers improved prediction accuracy compared with the analysis of clinical features alone (Taliaz et al., 2021). Prediction performance can be improved using data from multiple modalities because each modality offers unique information on different aspects of the patient (Brodersen et al., 2014). The predictive performance of NLP and neuroimaging have been separately investigated to predict psychiatric disorders; however, the complexity of language and the multidimensional nature of brain imaging still present challenges (Hauser et al., 2022). A recent review of AI models of mental illness found that while combinations of text and clinical data or brain imaging variables and clinical data existed, the combination of clinical text, brain imaging data, and clinical data was rarely applied (Torner-o-Costa et al., 2023). In this regard, machine learning methods that integrate text, brain imaging data, and clinical data have the potential to improve early prediction for effective interventions for people with TRD.

In the present study, we aimed to predict TRD at the time of depression diagnosis by leveraging structured electronic medical records data, brain morphometry, and NLP. To improve the clinical usefulness (Chandler et al., 2020), we also aimed to develop an explainable model by reducing the variables for each modality.

2. Methods

2.1. Data sources and study population

Clinical data were extracted from the electronic health records (EHRs) of the patients at Department of Psychiatry and Mental Health Center at the Ajou University School of Medicine (AUSOM) in South Korea between 2010 and 2022 (Lee et al., 2022b). The clinical data included socio-demographics, diagnoses, observations, provider visits, procedures performed, medications filled, clinical notes, and brain imaging data, including MRI data. The databases were formatted according to the Observational Medical Outcomes Partnership–Common Data Model version 5.3.1, maintained by the Observational Health Data Sciences and Informatics (OHDSI) and de-identified (Makadia and Ryan, 2014). This study was approved by the Institutional Review Board of the Ajou University Hospital (AJOUIRB-DB-2022–335). Informed consent was not required owing to the use of de-identified data.

In the present study, patients with a new depressive episode were included. The index date was defined as the date when the patient was first diagnosed with depressive disorder. To avoid any bias from left-censored data and to verify their first diagnosis of depressive disorder, we excluded patients who had been enrolled in the database for <1 year before the index date. Patients who were diagnosed with bipolar disorder, schizophrenia, and/or dementia were also excluded. To include patients who were treated for depression, those who had antidepressant prescriptions within 1 month after the index date were included. In addition, only patients who had undergone brain MRI from 1 year before the index date to 1 month after were included.

The outcome for the predictive models was TRD after the index date. The STAR*D naturalistic trial defined TRD as the failure of two antidepressant treatment trials at adequate dose and duration (Berman et al., 1997). In case of the absence of information on how patients are responding to treatments in observational databases, failure is considered to occur when a new antidepressant is added (Fife et al., 2017). To summarize, if a person had a history of using two types of antidepressants and a new type of antidepressant was started, we considered that a TRD occurred at that time. If TRD occurred after the index date, the observation was stopped on the day TRD was diagnosed. Thus, the predictive models were developed using the outcome. An outcome within 1 week after the index date was considered to have not occurred because we assumed that TRD within 1 week after a depression diagnosis was probably caused by a previous history or transient conditions (Moran et al., 2019). Further details regarding the cohort definitions and code lists are described in the Supplemental Materials.

2.2. Clinical data and variable extraction

In the clinical data, except for brain MRI data and clinical notes, the predictive variables for model training were extracted and dichotomized for existence within short-term (30 days) and long-term (365 days) intervals before the index. Subsequently, a tabular dataset was generated from the clinical data. The variables included patient demographics (sex and age), condition (medical diagnosis, grouped using a SNOMED–CT hierarchy), drug (based on the active ingredients), procedure (e.g., psychotherapy and electroconvulsive therapy.), measurement (e.g., assessment scale and laboratory test.), and observation (e.g., smoking status and alcohol intake). Predictors not recorded in our EHR system were considered to be non-occurring. Through this process, 7351 candidate variables were generated. Considering the lack of computational resources and improvement in model interpretability, we conducted a feature selection procedure with the least absolute shrinkage using a selection operator method (LASSO) and selected predictors for model development.

2.3. MRI acquisition and source-based morphometry analysis

For considering regional volume information in our model, structural T1-weighted MRI data were collected from all participants using 1.5T or 3T scanners at Ajou University Hospital (Signa HDx 1.5T and Discovery MR750w 3T, respectively; GE Healthcare, Milwaukee, MI, USA). The T1-weighted image acquisitions used a spin-echo sequence with the following scan parameters for each scanner: 1) 1.5T: TR = 500 ms, TE = 16 ms, FA (flip angle) = 72°, FOV = 125.12 × 75.07 mm^2, acquisition matrix = 320 × 192, and slice thickness = 5 mm) 3T: TR = 466.67 ms, TE = 10 ms, FA (flip angle) = 75°, FOV = 100.01 × 87.10

mm^2, acquisition matrix = 256 × 224, and slice thickness = 6 mm. All MRI images were visually inspected by neuroradiologists, and radiology reports were reviewed by psychiatrists. No observable scanning artifacts or gross brain abnormalities were identified in any participant included in the following analyses.

All structural MRI scans were used to estimate spatially-independent morphologies as common patterns for gray-matter concentration in all subjects through the source-based morphometry (SBM) approach (Kašpárek et al., 2010). This approach enables effective separation of several noise effects from true independent sources, reduces multiple comparisons of a tremendous number of voxels, and helps in reducing predictors' dimensions (Xu et al., 2009). In particular, we performed cross-sectional independent component analysis (ICA) on data preprocessed using voxel-based morphometry (VBM), which is conventionally used to measure voxel-level volumetric maps (Zhang et al., 2018). The SPM12 VBM-Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) procedure was used to process the VBM procedure (Ashburner, 2007). This procedure included gray-matter segmentation of T1-weighted images based on a standard tissue probability map, creation of a study-specific template, spatial normalization of individual images to the DARTEL template, modulation to adjust for volume signal changes, and spatial smoothing of the gray-matter partitions using a Gaussian kernel of 6 mm full-width at half maximum (Son et al., 2022). After estimating individual VBM maps, we applied a FastICA + ICASSO framework on the VBM-preprocessed data. We first reduced the original data to the optimal number of principal components determined using Laplace principal component analysis (Beckmann and Smith, 2004). After running FastICA on the reduced data 100 times using random initial values (Himberg et al., 2004), ICASSO was used to compute hierarchical clustering according to the dissimilarities among independent components (ICs) in each run (Murley et al., 2020). Among the ICASSO results, meaningful IC maps with high reliability of >0.8 were visually identified. All group-level IC maps were z-scored and thresholded with $z > 3$ for visualization.

2.4. Clinical notes and NLP

Among the clinical notes, we used the initial data recorded by psychiatrists at the time of the depression diagnosis. Previous NLP studies of psychiatric disorders have reported that linguistic differences affect feature extraction (Kishimoto et al., 2022). Therefore, we chose to extract only the English part in this study. In the initial record, the main symptoms were written in the chief complaint section. Psychiatrists wrote the chief complaints in English based on what is included in the DSM-5. Given these circumstances, we extracted chief complaints by regular expression. After a regular expression, NLP algorithms were used to extract topics as predictive variables from each patient's chief complaints. In particular, we converted the patient's chief complaints into a bag-of-words model of the corpus after stemming, normalization, and stop-word removal. Latent Dirichlet allocation (LDA), an unsupervised learning method, was used to cluster the topics from each patient (Blei, 2012). With an LDA-based topic model, topic probabilities were calculated for each chief complaint. For instance, if five topics were created by the LDA from chief complaints, the probability of being assigned to five topics for each chief complaint was generated (Lee et al., 2022a).

2.5. Model development

We developed prediction models for TRD based on the combination of the following types of data: selected features from the tabular dataset, weights of IC maps from brain MRI scans, and topic probabilities from clinical notes. As shown in Fig. 1, in total, seven models were developed and their performances were compared. All predictive models were developed using the extreme gradient boosting (XGBoost) algorithm, which can produce accurate predictions in the medical field (Yadaw et al., 2020).

2.6. Statistical analysis

Baseline characteristics are presented as counts with proportions for categorical variables and as means with standard deviations for continuous variables. Chi-squared test was used to compare categorical variables between populations. In all analyses, p-values <0.05 were considered to be statistically significant.

The enrolled patients were randomly divided into training and validation datasets at a ratio of 7:3. The hyperparameters of each model were optimized on the basis of a grid search using the area under the receiver operating characteristic curve (AUROC). For model development, five-fold cross-validation of the training dataset was performed. AUROC, area under the precision and recall curve (AUPRC), accuracy,

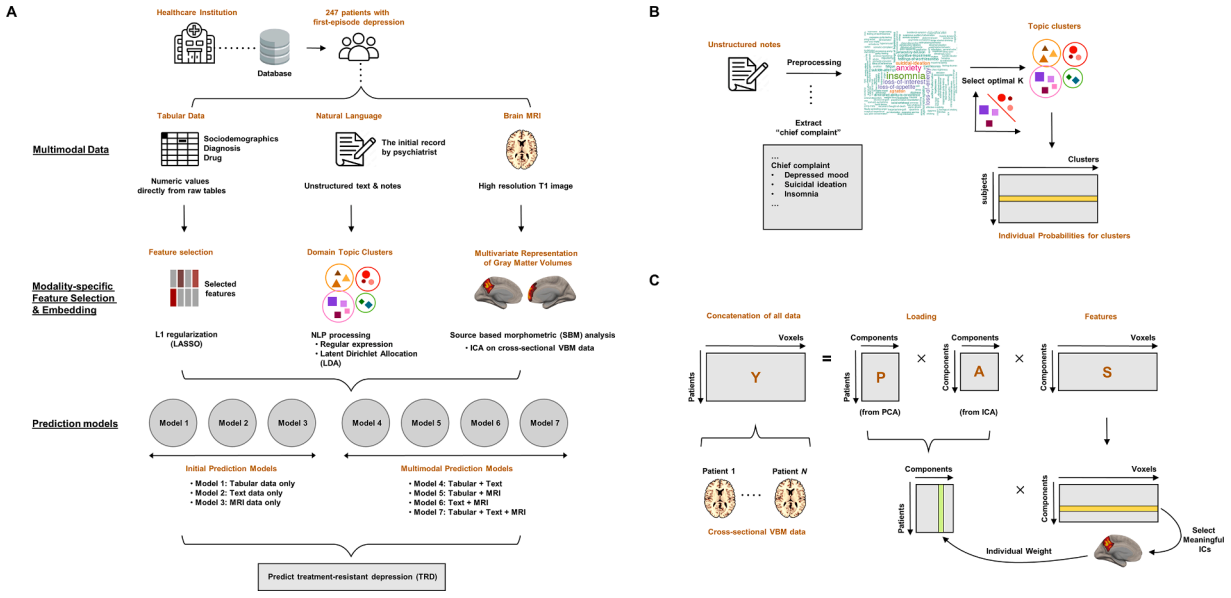

Fig. 1. Analysis workflow of treatment-resistant depression prediction. (A) Selected features from structured clinical data, NLP-derived features, and brain morphometric features were used to develop prediction models according to combinations of data. (B) NLP, including preprocessing and LDA-based topic model. (C) Source-based morphometric analysis. NLP: natural language processing; LDA: Latent Dirichlet allocation.

and F1 score were calculated to evaluate the performance of the prediction models using the validation dataset. The method reported by DeLong et al. (1988). was used to compare AUC values. Furthermore, the maximal Youden index was used to select the optimal cutoff value in the prediction model (Fluss et al., 2005).

Shapley additive explanation (SHAP) values were used to present the feature importance of the prediction model. The effect of each feature on TRD was presented as a SHAP value representing the importance of a variable by deriving a marginal distribution and weighted average with all but the variable of interest fixed (Aas et al., 2021). The SHAP summary plot sorts features in descending order on the basis of the effects on TRD. Each dot on each variable line represents one patient, and the horizontal location indicates the level of association between the feature and outcome. The right side shows the SHAP values >0, and variable-specific SHAP values >0 indicate an increased risk of outcome.

All analyses except brain MRI scans were performed using R software version 3.6 (R Foundation for Statistical Computing, Vienna, Austria), OHDSI's Health Analytics Data to Evidence Suite packages, and open-source statistical R packages. For brain MRI scans, statistical analyses were performed using MATLAB (MathWorks, Sherborn, MA, USA) ("MATLAB - MathWorks - MATLAB &Simulink," n.d.)-based custom software.

3. Results

3.1. Demographic and clinical characteristics

Table 1 shows the demographic and clinical characteristics of the study population. In total, 247 patients were selected based on the inclusion and exclusion criteria (Supplementary Fig. 1). Among the 247 patients with depression in the AUSOM database, 71 (28.7 %) experienced TRD after the depression diagnosis. The mean interval for TRD was 547 days. No significant differences in age, sex, medical history, and psychiatric history were observed between the groups. Female patients were the most predominant in the study population (without TRD: 74.4 %; with TRD: 69.0 %). Hypertension and anxiety disorder were frequent diagnoses (hypertension, 17.0 % and 28.2 %; anxiety disorder, 6.8 % and 5.6 %, respectively). Of the total number of patients, 95 (38.5 %) had undergone MRI before the diagnosis of depression. For these patients, the department and reason for the MRI are presented in Supplementary Table 2.

3.2. Model specification

As part of the variable selection in the tabular dataset, 4 predictors were selected through LASSO among a total of 7351 candidate predictors and were used in the prediction models. The characteristics selected were antihypertensive drugs within 1 year before diagnosis, an antidepressant within 1 year before diagnosis, an antidepressant at diagnosis, and diuretics within 1 year before diagnosis.

After SBM analysis, 26 innovation configuration (IC) maps that were localized at hypothetical large-scale functional brain networks related to depression (the default mode network [DMN], salience network [SN], frontoparietal network [FPN], superior parietal network [SPN], auditory network [AN], sensorimotor network [SMN], visual network [VN], cerebellum network [CN], limbic network [LN], and thalamus and basal ganglia network [THL/BG]) were selected (Lu et al., 2023; Singh et al., 2013). In particular, there were four maps in DMN, one in SN, three in FPN, one in SPN, two in AN, four in SMN, three in VN, three in CN, three in LN, and two in THL/BG. The specific brain regions of the 26 IC maps are shown in Supplementary Fig. 2. The reason why structural IC maps are defined based on functional network is that the area derived from SBM analysis is highly related and overlapped with the one of functional alterations (Wang et al., 2023). The areas that are coactivated frequently may be changed together in volume as they exhibit same neuro plasticity-related experiences (Alexander-Bloch et al., 2013; Evans, 2013).

Through NLP, we selected five topics as the most reliable hyperparameters for LDA performance (Supplementary Fig. 3) based on the perplexity scores. Each topic had the probability of being assigned to the topic as the variable value. Table 2 presents the topics used for the prediction models. The topics clustered were somatic symptoms, cognitive symptoms, suicidality, anxiety symptoms, and psychotic symptoms.

Overall, 35 features from tabular data, MRI data, and text data were used to develop the prediction model.

3.3. Prediction of TRD

The performance of the baseline model using only one type of the data from tabular data, text data, and MRI data gave AUROCs of 0.684, 0.569, and 0.703, respectively. In terms of AUROC, all models with two data types performed better than models with only one data type (Tabular + text: 0.728; Tabular + MRI: 0.770; and text + MRI: 0.762). Combining all types of data gave the highest predictive performance (accuracy: 0.796, F1 score: 0.667, AUPRC: 0.508, and AUROC: 0.794). Fig. 2 shows the ROC curves of the prediction models obtained using XGBoost. All evaluation metrics for model performance are presented in Table 3. All AUROC comparisons using the DeLong test are shown in Supplementary Table 1. With the exception of MRI, there was a significant difference in AUROC between models with all data types and models with only one data type (Tabular vs All: $p = 0.032$; Text vs All: $p = 0.005$; and MRI vs All: $p = 0.169$). Comparison of models with two data types and models with only one data type reveal that the model with the additional MRI or tabular data had a significantly higher AUROC than the model with text data alone (Text vs Text + Tabular: $p =$

Table 1
Demographic and clinical characteristics of patients with depression.

Characteristics	Without TRD (n = 176)	With TRD (n = 71)	P-value
Socio-demographics, n (%)			
Female	131 (74.4)	49 (69.0)	0.48
Race, Korean	176 (100.0)	71 (100.0)	1.00
Age, Mean (SD)	54.1 (17)	58.8 (18)	0.07
Medical history, n (%)			
Hypertension	30 (17.0)	20 (28.2)	0.07
Hyperlipidemia	4 (2.3)	3 (4.2)	0.68
Diabetes	8 (4.5)	7 (9.9)	0.20
Chronic liver disease	0 (0.0)	2 (2.8)	0.15
Renal impairment	3 (1.7)	1 (1.4)	1.00
Psychiatric history, n (%)			
Anxiety disorder	12 (6.8)	4 (5.6)	0.96
Sleep disorder	8 (4.5)	7 (9.9)	0.20
Substance use disorder	8 (4.5)	3 (4.2)	1.00
Personality disorder	4 (2.3)	0 (0.0)	0.47

Table 2
Topics clustered by LDA algorithm.

Topic	Main features of the topic	Topic examples
Topic 1	Somatic symptoms	headache, dizziness, fatigue, somatic concern, allergy, loss of energy, abdominal discomfort
Topic 2	Cognitive symptoms	diminished ability to concentrate, cognitive impairment, worthlessness, excessive worry, guilty feeling, hopelessness
Topic 3	Suicidality	suicide attempt, impulsivity, irritability, mood instability, CO intoxication, suicidal ideation, anger
Topic 4	Anxiety symptoms	anxiety, insomnia, vomiting, panic attack, chest discomfort, palpitation, sweating, anxiety attack
Topic 5	Psychotic symptoms	auditory hallucination, psychomotor retardation, visual hallucination, persecutory ideation, persecutory delusion, inattention

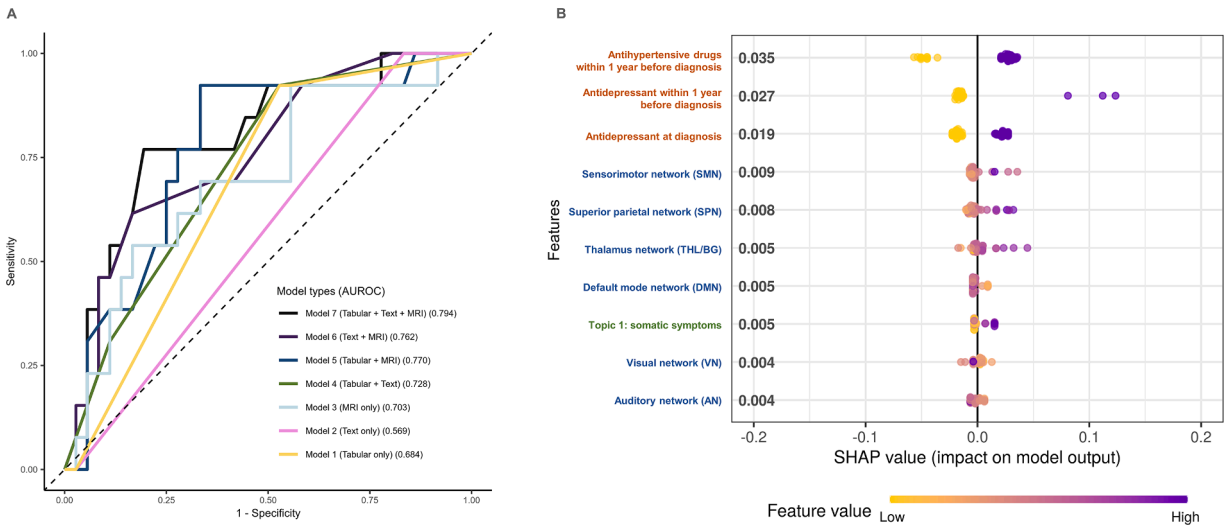

Fig. 2. Performance and feature importance of models predicting treatment-resistant depression. (A) ROC curve for the models according to data combinations. (B) SHAP beeswarm plot of the prediction model using all types of data. SHAP: Shapley Additive explanations; ROC: Receiver operating characteristic.

Table 3
Performance results of prediction models.

Performance metrics	Model 1 (Tabular)	Model 2 (Text)	Model 3 (MRI)	Model 4 (Tabular+Text)	Model 5 (Tabular+MRI)	Model 6 (Text +MRI)	Model 7 (All)
ACC	0.592	0.388	0.755	0.592	0.735	0.776	0.796
AUPRC	0.351	0.281	0.421	0.436	0.462	0.497	0.508
AUROC	0.684	0.569	0.703	0.728	0.770	0.762	0.794
F1 score	0.545	0.464	0.538	0.545	0.649	0.593	0.667

Notes: ACC, accuracy; AUPRC, area under the precision recall curve; AUROC, area under the receiver operating characteristics curve.

0.017; Text vs Text + MRI: $p = 0.012$). There was no significant difference between models with two data types and models with all data types (Tabular + Text vs All: $p = 0.187$; Text + MRI vs All: $p = 0.287$; and Text + MRI vs All: $p = 0.327$).

3.4. Features with the greatest contributions to prediction

Within the model using all data types, a graphical explanation of the top 10 features based on the average impact on the magnitude of the model output is shown in Fig. 2. The Y-axis represent the top 10 features in the prediction model, ranked in descending order. The top predictors were tabular variables, followed by brain MRI variables. The X-axis in the SHAP beeswarm plot shows the SHAP values. The SHAP value of each dot reflects the effect of a feature in the SHAP plot. For example, high median values of antihypertensive drugs within 1 year before diagnosis, antidepressants within 1 year before diagnosis, antidepressants at diagnosis, SMN, SPN, THL/BG, and topic 1 (somatic symptoms) were powerful predictors of TRD. Low median values of DMN, VN, and AN were more strongly predictive of TRD (Fig. 2B). Fig. 3 shows the brain MRI features included in the top 10 features.

3.5. Prediction of TRD with an adjusted number of MRI variables

Given that the number of MRI variables is larger than that of other modalities, we compared the performance of the predictions using only the top five selected MRI variables (SMN, SPN, THL/BG, DMN, and VN) from the total model including all variables. The performance results showed a consistent trend when all MRI variables were used. The model using all data sources showed the highest AUROC of 0.780, followed by models that used combined brain MRI and structured data, brain MRI and clinical notes, and brain MRI only (0.759, 0.731, and 0.728, respectively).

4. Discussion

In this study, we developed and tested an integrative model of multimodal data to predict TRD in patients with depression. To the best of our knowledge, this is the first study to develop a TRD prediction model combining structured clinical data, brain morphometric features, and NLP-derived symptoms. Furthermore, we extracted 4 out of 7351 clinical variables, 26 variables from brain MRI images, and 5 variables from free text to develop an explainable prediction model. Combining all three types of features was found to be higher across multiple performance metrics than using only one type of feature or combining two types of features. We found that combining all three types of features resulted in significantly higher AUROC scores than using only one type of feature except a model with only brain MRI. Compared with the accuracy reported in a previous study predicting treatment outcome in depression (0.66) (Sajjadian et al., 2022), our study showed better accuracy with a relatively larger patient population.

We found a gradient of increasing accuracy with the inclusion of additional measurement modalities. Although the model with all data types did not have a significantly higher AUROC than the model with two data types or the model with brain MRI only, AUROC, AUPRC, accuracy, and F1 score were the highest in the model with all data types. This finding supports previous findings on the added value of neuroimaging and NLP-derived features and extends them to the combination of the three types of data (Perlis et al., 2012; Williams et al., 2015). These results suggest that three modalities reflect different aspects of TRD etiopathology and thus are likely to improve classification performance (Gallo et al., 2023). Importantly, because depression has a high degree of heterogeneity in symptoms and prognosis across patients, models using multiple dimensions may enable a more personalized approach for depression prediction.

Previous research has shown that the improvement in predictive

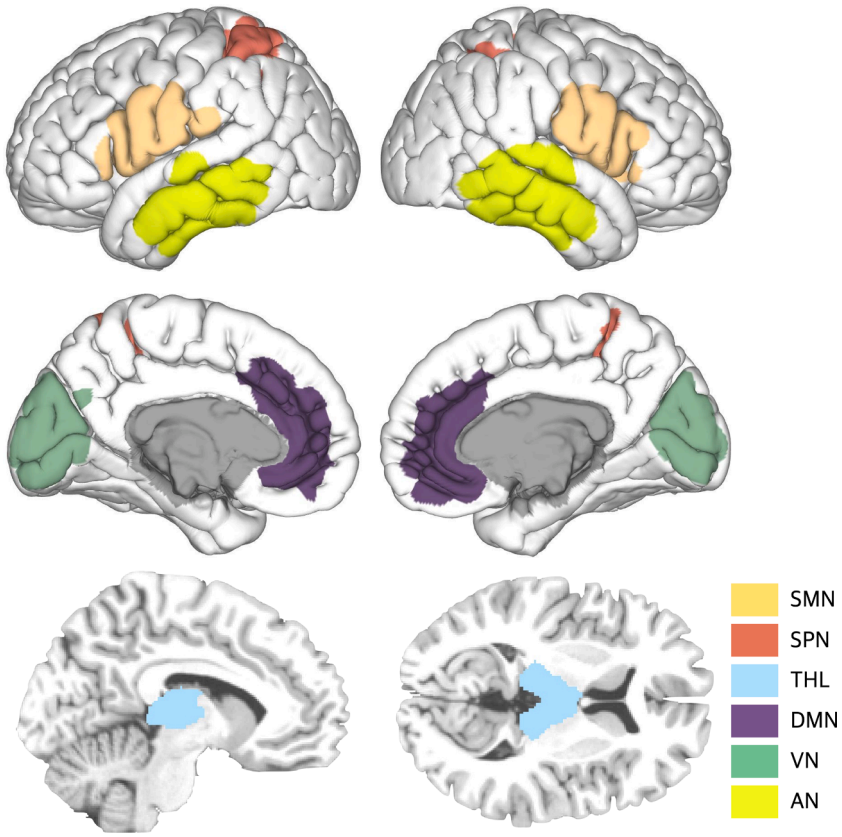

Fig. 3. Structural brain networks among the top 10 features of the prediction model using all types of data. SMN: sensorimotor network; SPN: superior parietal network; THL: thalamus network; DMN: default mode network; VN: visual network; AN: auditory network.

performance is attributed to multiple data modalities, not just to a greater number of variables (Sajjadian et al., 2022). In particular, when the number of predictors exceeds the number of participants, the efficiency of developing predictive models decreases. In this respect, feature selection was done by modality in this study. Using LASSO, we were able to reduce over 7000 variables of structured data to 4 variables. We used the ICA method, which can extract robust and reproducible brain components, to extract 26 features from brain MRI images for brain imaging (Cerliani et al., 2015). For free text, we used the LDA method, which can represent topic clustering of similar words, to extract five topics (Park et al., 2021). LASSO, ICA, and LDA enable the use of a reasonably small number of variables in the model, making the model sufficiently explainable (Vucenovic et al., 2020). In particular, the ICA and LDA methods are relatively easy to explain by reflecting brain network or semantic characteristics. Furthermore, consistent trends in performance were observed in analyses using a reduced number of MRI variables to determine the impact of differences in the number of variables, indicating that the number of modalities rather than the number of variables may have an impact on model performance. In addition, since this study used an algorithm that used machine learning rather than deep learning to develop prediction models, understandable explanations for prediction were obtained (Wang et al., 2020). Considering that the interpretability of prediction models is vital in health care (Chen et al., 2020), the multidimensional explainable model developed in this study can be useful in clinical settings.

The most discriminative brain morphometric pattern was the sensorimotor network. Our results align with those of Qin et al. (2015) who observed that sensorimotor networks were one of the most discriminative functional connections for predicting clinical responses in patients with major depression. A previous study found a greater amplitude of low-frequency fluctuation in the sensorimotor network in patients with TRD (Zhang et al., 2019). Moreover, a previous study on structural covariance networks have also reported significant associations between the sensorimotor network and depression (Yang et al., 2021). Stimulation of sensory modalities or motor systems affects the same circuits that affect mood regulation and can lead to mood disorders. In addition, depression interacts with sensorimotor processing, which exacerbates depressive symptoms (Canbeyli, 2010). We also observed that the SPN acted as a discriminative region. Similar to the sensorimotor network, the superior parietal region is involved in interpreting sensory information (Radua et al., 2010). People with depression who undergo changes in the parietal lobe experience difficulties in perceiving emotions (Culham and Kanwisher, 2001). Previous morphometric correlation analysis revealed increased caudate–cortical connectivity in the bilateral superior parietal gyrus in patients with major depressive disorder (MDD). This region was also positively correlated with Beck Depression Inventory scores in patients with MDD (Yang et al., 2015). Furthermore, patients with MDD have shown increased functional connectivity between the bilateral amygdala and superior parietal gyrus (Jiang et al., 2021). The thalamus is also considered an important predictor. In a study that compared individuals with depression and healthy controls, hyperconnectivity of the thalamus was identified as the most influential predictor (Gallo et al., 2023). Metabolic abnormalities in the thalamus of patients with depression were also described elsewhere (Dougherty et al., 2003). Unlike the abovementioned brain networks, DMN, VN, and AN were negatively correlated with TRD. Depression is characterized by abnormal brain networks, particularly DMN, which regulates awareness of internal states and is related to affective and cognitive symptoms in depression (Kaiser et al., 2015). Although hyperconnectivity of the DMN is observed in depression, hypoconnectivity of the DMN has been reported in patients unresponsive to antidepressants (Kaiser et al., 2015; Korgaonkar et al., 2020). In the same study, hypoconnectivity was also observed in the VN and AN of patients who responded poorly to

antidepressants. In a previous study, functional connectivity in the AN and VN of patients with depression was lower than that in healthy controls (Lu et al., 2020). An abnormal auditory processing-related AN and impaired facial emotion processing-related VN have also been reported (Guo et al., 2013; Tollkötter et al., 2006). Conversely, other networks (SN, FPN, CN, and LN), which were not considered significant but were included as variables in the model, are also known to be associated with TRD. SN is considered an indicator of resting state treatment response in depression (Horn et al., 2010). Disrupted FPN is also associated with the severity of suicide risk in patients with depression (Dai et al., 2022). CN is positively correlated with Hamilton Depression Rating Scale score (Yin et al., 2015). In addition, the course of depression is associated with altered LN activity (Lemke et al., 2022).

The most significant variable among the NLP-derived symptoms was somatic symptoms. In a previous study, the majority of patients with TRD presented with somatic symptoms. The presence of somatic symptoms was related to a poorer prognosis of depression (Papakostas et al., 2003). Other NLP-derived symptoms have been reported as potential risk factors for TRD. Evidence suggests that TRD is associated with the "melancholia" subtype of depression (Malhi et al., 2005). Symptoms of the melancholia subtype include suicidal thoughts and difficulty concentrating, which are consistent with our findings on cognitive symptoms and suicidality (Parker et al., 2010). In terms of psychiatric comorbidity, TRD has been associated with a higher prevalence of comorbid anxiety disorders and psychotic features, which align with our findings on psychotic and anxiety symptoms (Souery et al., 2007; Dold et al., 2019). However, when predicting TRD using only NLP-derived variables, the model performed poorly. Previous attempts to classify patients using NLP showed insufficient performance for clinical application (Rumshisky et al., 2016). Because of the overlap in symptoms among patients (American Psychiatric Association, 2013), it may not be accurate to differentiate them on the basis of symptoms alone. Nevertheless, the inclusion of topics from NLP allows a more accurate discrimination of TRD.

For tabular data, exposure to antidepressants on the day of diagnosis and before diagnosis, and to antihypertensive medications before diagnosis were positively related to predicting prediction. Prescription of antidepressants on the day of diagnosis may indicate a more severe level of depression. Previous studies have reported that depression severity is a strong risk factor for TRD (Gronemann et al., 2020; Kautzky et al., 2019). The severity of depression increases the likelihood of applying different treatment options, which thus fulfills the definition of TRD. Certain situations may require the prescription of antidepressants to people who are not depressed. For example, for anxiety disorders or somatic symptom disorders, antidepressants are often the primary medication of choice (Bandelow et al., 2017; Kurlansik et al., 2016). In fact, 6.5 % of our patients had a history of anxiety disorders, and somatic symptoms were extracted as NLP features in our patients. The presence of anxiety or somatic symptoms in depressive disorders negatively impacts depression prognosis, suggesting that a history of antidepressant use for such symptoms could be predictive of TRD (De Carlo et al., 2016; Bekhuis et al., 2016). Regarding the use of antihypertensive drugs and diuretics, Gronemann et al. (2020) showed that patients prescribed with β-blockers had increased rates of TRD. According to meta-analytic evidence, diuretics are significantly associated with an increased risk of depression and anxiety disorders (Zhang et al., 2022). However, other reports refute the association between the use of angiotensin antagonists, β-blockers, and diuretics and the risk of depression (Li et al., 2021). While the inherent risk of these medications is worth consideration, the influence of cardiovascular disease on depression should be accounted for since it is the underlying condition often necessitating antihypertensive medication. Shang et al. (2019) identified cardiovascular disease as a risk factor for incident depression. This association might be partly explained by the perceived loss of health, functional capability, and independence among patients with cardiovascular disease, potentially contributing to an increased incidence of depression (Hare et al., 2014). Furthermore, a bidirectional association between TRD and common medical conditions has been reported, with hypertension being associated with TRD in both men and women (Madsen et al., 2021). Given the known relationship between age and cardiovascular disease and considering the most patients in the present study were in their 50 s, it is possible that the use of antihypertensive drugs and diuretics were selected as predictors.

This study had several limitations. First, the results should be interpreted in light of sample size limitations. Although the study sample was larger than that in previously reported multimodal studies (Iniesta et al., 2018; Sajjadian et al., 2022), it was not large enough to optimally support the learning of complex prediction models. Considering the size, we used five-fold cross-validation for the model development in the training dataset while maintaining strict separation of the training and test sets. Second, this study did not perform external validation of the model. It is difficult to access clinical data, notes, and brain imaging records at the same time. For this reason, there has been a lack of external validation in multimodal studies (Sajjadian et al., 2022). In addition, psychiatric records are often kept separately or not shared, making it difficult to find hospital records that could be used for external validation (Appelbaum, 2002). Therefore, future research is needed to investigate and validate our results. Third, this study used data from Koreans only, so the results cannot be generalized. However, the characteristics of the patients in our study were similar to those enrolled in previous TRD studies. For example, the average age of our patients was 50 years, similar to that in a previous study (Kautzky et al., 2019), and the median time to TRD was 547 days, similar to the 420 days found in another previous study (Fife et al., 2017). Fourth, we did not use data from other imaging modalities, such as resting-state functional MRI and DTI, to supplement the prediction models. However, the structural covariance network using ICA appears to be consistent with the findings of a previous functional network study (Watanabe et al., 2020). Fifth, this study used MRI data obtained using 1.5T and 3T scanners. Although using one scanner consistently is preferable, a previous study reported similar brain volume patterns between 1.5T and 3T (Chow et al., 2015). Sixth, the inclusion of MRIs within 1 year prior to diagnosis of depression in the inclusion criteria is not standard practice for patients with depression. Because MRIs are not routinely performed in patients with depression, we had to include cases that were performed as part of a routine checkup or by another department before diagnosis of depression. Furthermore, we excluded patients who had undergone MRI for serious conditions by manually checking imaging data, radiology reports, and patients' medical histories, given that serious conditions could have required a brain MRI. However, while most MRI procedures were indicated for somatic symptoms or routine check-ups, a subset of patients presented with suicidal or psychotic symptoms. Due to the small sample size of our study, our findings have limited generalizability, and further research is needed to validate our results. Seventh, the change in antidepressant medication may have been due to pharmacological side effects rather than treatment resistance. However, directly capturing patients' response to treatments is difficult in observational databases. For this reason, previous studies using observational data have defined antidepressant trial failure as occurring when a new antidepressant is added, as we did in our study. Nevertheless, given these possibilities, our results should be interpreted with caution. Lastly, the average age of our study sample may not represent the entire population of patients with depression. Because of the relatively higher age group of our patients, variables such as antihypertensive medications may have been included in the analysis. Further research is needed, including external validation and studies involving a larger and more diverse patient population. Despite the abovementioned limitations, this remains the first multimodal TRD prediction study to combine structural clinical data, brain morphometric data, and NLP-derived data, resulting in high classification accuracy.

5. Conclusion

In conclusion, our study showed that the combination of clinical data with neuroimaging and NLP variables improved the prediction of TRD in comparison with single-modality measurement. Similar to clinical situations where assessments are made using a variety of data, our results suggest that using a combination of data may be a more useful option for prediction of TRD. Further studies using a multidimensional approach to achieve meaningful prediction of TRD are needed by other groups.

Ethics declaration

This study was approved by the Institutional Review Board of the Ajou University Hospital (AJOUIRB-DB-2022-335). Informed consent was not required owing to the use of de-identified data.

CRediT authorship contribution statement

Dong Yun Lee: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Narae Kim:** Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. **ChulHyoung Park:** Data curation, Investigation. **Sujin Gan:** Data curation, Investigation. **Sang Joon Son:** Investigation, Writing – review & editing. **Rae Woong Park:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Bumhee Park:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

All authors declare no competing interests.

Source of funding

This research was supported by a grant of the MD-Phd/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea. This work was supported by the M.D. research fund of Ajou University School of Medicine

Also, this work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (Ministry of Science and ICT) (NRF-2022R1A2C1005967, NRF-2019R1A5A2026045).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2024.115817.

References

Aas, K., Jullum, M., Løland, A., 2021. Explaining individual predictions when features are dependent: more accurate approximations to Shapley values. Artif. Intell. 298, 103502.
American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental disorders: DSM-5. American Psychiatric Association, Washington, DC.
Appelbaum, P.S., 2002. Privacy in psychiatric treatment: threats and responses. Am. J. Psychiatry 159 (11), 1809–1818.
Alexander-Bloch, A., Giedd, J.N., Bullmore, E., 2013. Imaging structural co-variance between human brain regions. Nat. Rev. Neurosci. 14 (5), 322–336.
Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113.
Bailey, N.W., Hoy, K.E., Rogasch, N.C., Thomson, R.H., McQueen, S., Elliot, D., Sullivan, C.M., Fulcher, B.D., Daskalakis, Z.J., Fitzgerald, P.B., 2019. Differentiating responders and non-responders to rTMS treatment for depression after one week using resting EEG connectivity measures. J. Affect. Disord. 242, 68–79.
Bandelow, B., Michaelis, S., Wedekind, D., 2017. Treatment of anxiety disorders. Dialogues Clin. Neurosci. 19 (2), 93–107.
Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23 (2), 137–152.
Bekhuis, E., Boschloo, L., Rosmalen, J.G., de Boer, M.K., Schoevers, R.A., 2016. The impact of somatic symptoms on the course of major depressive disorder. J. Affect. Disord. 205, 112–118.
Berman, R.M., Narasimhan, M., Charney, D.S., 1997. Treatment-refractory depression: definitions and characteristics. Depress. Anxiety 5 (4), 154–164.
Blei, D.M., 2012. Probabilistic topic models. Commun. ACM 55 (4), 77–84.
Brodersen, K.H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W.D., Buhmann, J.M., Stephan, K.E., 2014. Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin. 4, 98–111.
Canbeyli, R., 2010. Sensorimotor modulation of mood and depression: an integrative review. Behav. Brain Res. 207 (2), 249–264.
Cepeda, M.S., Reps, J., Fife, D., Blacketer, C., Stang, P., Ryan, P., 2018. Finding treatment-resistant depression in real-world data: how a data-driven approach compares with expert-based heuristics. Depress. Anxiety 35 (3), 220–228.
Cerliani, L., Mennes, M., Thomas, R.M., Di Martino, A., Thioux, M., Keysers, C., 2015. Increased functional connectivity between subcortical and cortical resting-state networks in autism spectrum disorder. JAMA Psychiatry 72 (8), 767–777.
Chandler, C., Foltz, P.W., Elvevåg, B., 2020. Using machine learning in psychiatry: the need to establish a framework that nurtures trustworthiness. Schizophr. Bull. 46 (1), 11–14.
Chen, P., Dong, W., Wang, J., Lu, X., Kaymak, U., Huang, Z., 2020. Interpretable clinical prediction via attention-based neural network. BMC Med. Inform. Decis. Mak. 20 (3), 1–9.
Chow, N., Hwang, K.S., Hurtz, S., Green, A.E., Somme, J.H., Thompson, P.M., Elashoff, D. A., Jack, C., Weiner, M., Apostolova, L.G., 2015. Comparing 3T and 1.5 T MRI for mapping hippocampal atrophy in the Alzheimer's Disease neuroimaging initiative. Am. J. Neuroradiol. 36 (4), 653–660.
Chowdhary, K., 2020. Fundamentals of Artificial Intelligence. Springer.
Culham, J.C., Kanwisher, N.G., 2001. Neuroimaging of cognitive functions in human parietal cortex. Curr. Opin. Neurobiol. 11 (2), 157–163.
Dai, Z., Shao, J., Zhou, H., Chen, Z., Zhang, S., Wang, H., Jiang, H., Yao, Z., Lu, Q., 2022. Disrupted fronto-parietal network and default-mode network gamma interactions distinguishing suicidal ideation and suicide attempt in depression. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 113, 110475.
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.
DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 837–845.
Dold, M., Kasper, S., 2017. Evidence-based pharmacotherapy of treatment-resistant unipolar depression. Int. J. Psychiatry Clin. Pract. 21 (1), 13–23.
Dold, M., Bartova, L., Kautzky, A., Porcelli, S., Montgomery, S., Zohar, J., Mendlewicz, J., Souery, D., Serretti, A., Kasper, S., 2019. Psychotic features in patients with major depressive disorder: a report from the European group for the study of resistant depression. J. Clin. Psychiatry 80 (1), 16309.
Dougherty, D.D., Weiss, A.P., Cosgrove, G.R., Alpert, N.M., Cassem, E.H., Nierenberg, A. A., Price, B.H., Mayberg, H.S., Fischman, A.J., Rauch, S.L., 2003. Cerebral metabolic correlates as potential predictors of response to anterior cingulotomy for treatment of major depression. J. Neurosurg. 99 (6), 1010–1017.
Evans, A.C., 2013. Networks of anatomical covariance. Neuroimage 80, 489–504.
Fife, D., Feng, Y., Wang, M.Y.H., Chang, C.J., Liu, C.Y., Juang, H.T., Furnback, W., Singh, J., Wang, B., 2017. Epidemiology of pharmaceutically treated depression and treatment resistant depression in Taiwan. Psychiatry Res. 252, 277–283.
Fluss, R., Faraggi, D., Reiser, B., 2005. Estimation of the Youden Index and its associated cutoff point. Biom. J. J. Math. Methods Biosci. 47 (4), 458–472.
Gallo, S., El-Gazzar, A., Zhutovsky, P., Thomas, R.M., Javaheripour, N., Li, M., Bartova, L., Bathula, D., Dannlowski, U., Davey, C., 2023. Functional connectivity signatures of major depressive disorder: machine learning analysis of two multicenter neuroimaging studies. Mol. Psychiatry 28, 3013–3022.
Gronemann, F.H., Jorgensen, M.B., Nordentoft, M., Andersen, P.K., Osler, M., 2020. Socio-demographic and clinical risk factors of treatment-resistant depression: a Danish population-based cohort study. J. Affect. Disord. 261, 221–229.
Guo, W., Liu, F., Xue, Z., Gao, K., Liu, Z., Xiao, C., Chen, H., Zhao, J., 2013. Abnormal resting-state cerebellar–cerebral functional connectivity in treatment-resistant depression and treatment sensitive depression. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 44, 51–57.
Hauser, T.U., Skvortsova, V., De Choudhury, M., Koutsouleris, N., 2022. The promise of a model-based psychiatry: building computational models of mental ill health. Lancet Digit. Health 4 (11), e816–e828.
Hare, D.L., Toukhsati, S.R., Johansson, P., Jaarsma, T., 2014. Depression and cardiovascular disease: a clinical review. Eur. Heart J. 35 (21), 1365–1372.
Himberg, J., Hyvärinen, A., Esposito, F., 2004. Validating the independent components of neuroimaging time series via clustering and visualization. Neuroimage 22 (3), 1214–1222.
Iniesta, R., Hodgson, K., Stahl, D., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M.Z., 2018. Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. Sci. Rep. 8 (1), 1–9.
Horn, D.I., Yu, C., Steiner, J., Buchmann, J., Kaufmann, J., Osoba, A., Eckert, U., Zierhut, K.C., Schiltz, K., He, H., Biswal, B., 2010. Glutamatergic and resting-state functional connectivity correlates of severity in major depression–the role of pregenual anterior cingulate cortex and anterior insula. Front. Syst. Neurosci. 4, 33.

Jiang, X., Wang, X., Jia, L., Sun, T., Kang, J., Zhou, Y., Wei, S., Wu, F., Kong, L., Wang, F., 2021. Structural and functional alterations in untreated patients with major depressive disorder and bipolar disorder experiencing first depressive episode: a magnetic resonance imaging study combined with follow-up. J. Affect. Disord. 279, 324–333.
Johnston, B.A., Steele, J.D., Tolomeo, S., Christmas, D., Matthews, K., 2015. Structural MRI-based predictions in patients with treatment-refractory depression (TRD). PLoS ONE 10 (7), e0132958.
Kaiser, R.H., Andrews-Hanna, J.R., Wager, T.D., Pizzagalli, D.A., 2015. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. JAMA Psychiatry 72 (6), 603–611.
Kašpárek, T., Mareček, R., Schwarz, D., Přikryl, R., Vaníček, J., Mikl, M., Češková, E., 2010. Source-based morphometry of gray matter volume in men with first-episode schizophrenia. Hum. Brain Mapp. 31 (2), 300–310.
Kautzky, A., Dold, M., Bartova, L., Spies, M., Kranz, G.S., Souery, D., Montgomery, S., Mendlewicz, J., Zohar, J., Fabbri, C., 2019. Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study. Acta Psychiatr. Scand. 139 (1), 78–88.
Kessler, R.C., 2012. The costs of depression. Psychiatr. Clin. 35 (1), 1–14.
Kishimoto, T., Nakamura, H., Kano, Y., Eguchi, Y., Kitazawa, M., Liang, K.C., Kudo, K., Sento, A., Takamiya, A., Horigome, T., 2022. Understanding psychiatric illness through natural language processing (UNDERPIN): rationale, design, and methodology. Front. Psychiatry 13, 954703.
Korgaonkar, M.S., Goldstein-Piekarski, A.N., Fornito, A., Williams, L.M., 2020. Intrinsic connectomes are a predictive biomarker of remission in major depressive disorder. Mol. Psychiatry 25 (7), 1537–1549.
Kurlansik, S.L., Maffei, M.S., 2016. Somatic symptom disorder. Am. Fam. Physician 93 (1), 49–54A.
Lee, D.Y., Kim, C., Lee, S., Son, S.J., Cho, S.M., Cho, Y.H., Lim, J., Park, R.W., 2022a. Psychosis relapse prediction leveraging electronic health records data and natural language processing enrichment methods. Front. Psychiatry 13, 844442.
Lee, E., Karim, H., Andreescu, C., Mizuno, A., Aizenstein, H., Lee, H., Lee, D., Lee, K., Cho, S.M., Kim, D., Park, R.W., Son, S.J., Park, B., 2022b. Network modeling of anxiety and psychological characteristics on suicidal behavior: cross-sectional study. J. Affect. Disord. 299, 545–552.
Lemke, H., Probst, S., Warneke, A., Waltemate, L., Winter, A., Thiel, K., Meinert, S., Enneking, V., Breuer, F., Klug, M., Goltermann, J., 2022. The course of disease in major depressive disorder is associated with altered activity of the limbic system during negative emotion processing. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 7 (3), 323–332.
Lichtenberg, P., Belmaker, R., 2010. Subtyping major depressive disorder. Psychother. Psychosom. 79 (3), 131–135.
Li, Y., Fan, Y., Sun, Y., Alolga, R.N., Xiao, P., Ma, G., 2021. Antihypertensive drug use and the risk of depression: a systematic review and network meta-analysis. Front. Pharmacol. 12, 777987.
Lu, F., Cui, Q., Chen, Y., He, Z., Sheng, W., Tang, Q., Yang, Y., Luo, W., Yu, Y., Chen, J., 2023. Insular-associated causal network of structural covariance evaluating progressive gray matter changes in major depressive disorder. Cereb. Cortex 33 (3), 831–843.
Lu, F., Cui, Q., Huang, X., Li, L., Duan, X., Chen, H., Pang, Y., He, Z., Sheng, W., Han, S., 2020. Anomalous intrinsic connectivity within and between visual and auditory networks in major depressive disorder. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 100, 109889.
Madsen, K.B., Momen, N.C., Petersen, L.V., Plana-Ripoll, O., Haarman, B.C., Drexhage, H., Mortensen, P.B., McGrath, J.J., Munk-Olsen, T., 2021. Bidirectional associations between treatment-resistant depression and general medical conditions. Eur. Neuropsychopharmacol. 51, 7–19.
Makadia, R., Ryan, P.B., 2014. Transforming the premier perspective hospital database into the Observational Medical Outcomes Partnership (OMOP) common data model. EGEMS (Wash. DC) 2 (1), 1110.
Malhi, G.S., Parker, G.B., Crawford, J., Wilhelm, K., Mitchell, P.B., 2005. Treatment-resistant depression: resistant to definition? Acta Psychiatr. Scand. 112 (4), 302–309.
McMakin, D.L., Olino, T.M., Porta, G., Dietz, L.J., Emslie, G., Clarke, G., Wagner, K.D., Asarnow, J.R., Ryan, N.D., Birmaher, B., 2012. Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment–resistant depression. J. Am. Acad. Child Adolesc. Psychiatry 51 (4), 404–411.
McIntyre, R.S., Filteau, M.J., Martin, L., Patry, S., Carvalho, A., Cha, D.S., Barakat, M., Miguelez, M., 2014. Treatment-resistant depression: definitions, review of the evidence, and algorithmic approach. J. Affect. Disord. 156, 1–7.
Moran, L.V., Ongur, D., Hsu, J., Castro, V.M., Perlis, R.H., Schneeweiss, S., 2019. Psychosis with methylphenidate or amphetamine in patients with ADHD. N. Engl. J. Med. 380 (12), 1128–1138.
Murley, A.G., Coyle-Gilchrist, I., Rouse, M.A., Jones, P.S., Li, W., Wiggins, J., Lansdall, C., Rodríguez, P.V., Wilcox, A., Tsvetanov, K.A., 2020. Redefining the multidimensional clinical phenotypes of frontotemporal lobar degeneration syndromes. Brain 143 (5), 1555–1571.
Papakostas, G.I., Petersen, T., Denninger, J., Sonawalla, S.B., Mahal, Y., Alpert, J.E., Nierenberg, A.A., Fava, M., 2003. Somatic symptoms in treatment-resistant depression. Psychiatry Res. 118 (1), 39–45.
Park, J., You, S.C., Jeong, E., Weng, C., Park, D., Roh, J., Lee, D.Y., Cheong, J.Y., Choi, J. W., Kang, M., 2021. A Framework (SOCRATex) for hierarchical annotation of unstructured electronic health records and integration into a standardized medical database: development and usability study. JMIR. Med. Inform. 9 (3), e23983.
Parker, G., Fink, M., Shorter, E., Taylor, M.A., Akiskal, H., Berrios, G., Bolwig, T., Brown, W.A., Carroll, B., Healy, D., Klein, D.F., 2010. Issues for DSM-5: whither melancholia? The case for its classification as a distinct mood disorder. Am. J. Psychiatry 167 (7), 745–747.
Perlis, R., Iosifescu, D., Castro, V., Murphy, S., Gainer, V., Minnier, J., Cai, T., Goryachev, S., Zeng, Q., Gallagher, P., 2012. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychol. Med. 42 (1), 41–50.
Perlis, R.H., 2014. Pharmacogenomic testing and personalized treatment of depression. Clin. Chem. 60 (1), 53–59.
Phillips, J.L., Batten, L.A., Tremblay, P., Aldosary, F., Blier, P., 2015. A prospective, longitudinal study of the effect of remission on cortical thickness and hippocampal volume in patients with treatment-resistant depression. Int. J. Neuropsychopharmacol. 18 (8), 1–9.
Pigott, H.E., 2015. The STAR* D trial: it is time to reexamine the clinical beliefs that guide the treatment of major depression. Can. J. Psychiatry 60 (1), 9–13.
Pigoni, A., Delvecchio, G., Madonna, D., Bressi, C., Soares, J., Brambilla, P., 2019. Can Machine Learning help us in dealing with treatment resistant depression? A review. J. Affect. Disord. 259, 21–26.
Qin, J., Shen, H., Zeng, L.L., Jiang, W., Liu, L., Hu, D., 2015. Predicting clinical responses in major depression using intrinsic functional connectivity. Neuroreport 26 (12), 675–680.
Radua, J., Phillips, M.L., Russell, T., Lawrence, N., Marshall, N., Kalidindi, S., El-Hage, W., McDonald, C., Giampietro, V., Brammer, M.J., 2010. Neural response to specific components of fearful faces in healthy and schizophrenic adults. Neuroimage 49 (1), 939–946.
Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V., McCoy, T., Perlis, R., 2016. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. Transl. Psychiatry 6 (10), e921.
Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Stewart, J.W., Nierenberg, A.A., Thase, M.E., Ritz, L., Biggs, M.M., Warden, D., Luther, J.F., 2006. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. N. Engl. J. Med. 354 (12), 1231–1242.
Sajjadian, M., Uher, R., Ho, K., Hassel, S., Milev, R., Frey, B.N., Farzan, F., Blier, P., Foster, J.A., Parikh, S.V., 2022. Prediction of depression treatment outcome from multimodal data: a CAN-BIND-1 report. Psychol. Med. 12, 5374–5384.
Shang, X., Peng, W., Hill, E., Szoeke, C., He, M., Zhang, L., 2019. Incidence of medication-treated depression and anxiety associated with long-term cancer, cardiovascular disease, diabetes and osteoarthritis in community-dwelling women and men. eClinicalMedicine, 15, 23–32.
Simon, G.E., Perlis, R.H., 2010. Personalized medicine for depression: can we match patients with treatments? Am. J. Psychiatry 167 (12), 1445–1455.
Singh, M.K., Kesler, S.R., Hosseini, S.H., Kelley, R.G., Amatya, D., Hamilton, J.P., Chen, M.C., Gotlib, I.H., 2013. Anomalous gray matter structural networks in major depressive disorder. Biol. Psychiatry 74 (10), 777–785.
Son, S.J., Hong, C.H., Kim, N.R., Choi, J.W., Roh, H.W., Lee, H., Seo, S.W., Choi, S.H., Kim, E.J., Kim, B.C., 2022. Structural covariance changes in major cortico-basal ganglia and thalamic networks in amyloid-positive patients with white matter hyperintensities. Neurobiol. Aging 117, 117–127.
Souery, D., Oswald, P., Massat, I., Bailer, U., Bollen, J., Demyttenaere, K., Kasper, S., Lecrubier, Y., Montgomery, S., Serretti, A., Zohar, J., 2007. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. J. Clin. Psychiatry 68 (7), 1062–1070.
Taliaz, D., Spinrad, A., Barzilay, R., Barnett-Itzhaki, Z., Averbuch, D., Teltsh, O., Schurr, R., Darki-Morag, S., Lerer, B., 2021. Optimizing prediction of response to antidepressant medications using machine learning and integrated genetic, clinical, and demographic data. Transl. Psychiatry 11 (1), 381.
Tollkötter, M., Pfleiderer, B., Sörös, P., Michael, N., 2006. Effects of antidepressive therapy on auditory processing in severely depressed patients: a combined MRS and MEG study. J. Psychiatr. Res. 40 (4), 293–306.
Tornero-Costa, R., Martinez-Millana, A., Azzopardi-Muscat, N., Lazeri, L., Traver, V., Novillo-Ortiz, D., 2023. Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review. JMIR Ment. Health 10 (1), e42045.
Vucenovic, A., Ali-Ozkan, O., Ekwempe, C., Eren, O., 2020. Explainable ai in decision support systems: a case study: predicting hospital readmission within 30 days of discharge. In: Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, pp. 1–4.
Wang, W., Kiik, M., Peek, N., Curcin, V., Marshall, I.J., Rudd, A.G., Wang, Y., Douiri, A., Wolfe, C.D., Bray, B., 2020. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS ONE 15 (6), e0234722.
Wang, K., Hu, Y., Yan, C., Li, M., Wu, Y., Qiu, J., Zhu, X., REST-meta-MDD Consortium, 2023. Brain structural abnormalities in adult major depressive disorder revealed by voxel-and source-based morphometry: evidence from the REST-meta-MDD Consortium. Psychol. Med. 53 (8), 3672–3682.
Watanabe, K., Kakeda, S., Katsuki, A., Ueda, I., Ikenouchi, A., Yoshimura, R., Korogi, Y., 2020. Whole-brain structural covariance network abnormality in first-episode and drug-naïve major depressive disorder. Psychiatry Res. Neuroimaging 300, 111083.
Williams, L.M., Korgaonkar, M.S., Song, Y.C., Paton, R., Eagles, S., Goldstein-Piekarski, A., Grieve, S.M., Harris, A.W., Usherwood, T., Etkin, A., 2015. Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. Neuropsychopharmacology 40 (10), 2398–2408.
Xu, L., Groth, K.M., Pearlson, G., Schretlen, D.J., Calhoun, V.D., 2009. Source-based morphometry: the use of independent component analysis to identify gray matter differences with application to schizophrenia. Hum. Brain Mapp. 30 (3), 711–724.

Yadaw, A.S., Li, Y.C., Bose, S., Iyengar, R., Bunyavanich, S., Pandey, G., 2020. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. Lancet Digit. Health 2 (10), e516–e525.
Yang, X.H., Wang, Y., Huang, J., Zhu, C.Y., Liu, X.Q., Cheung, E.F., Xie, G.R., Chan, R.C., 2015. Increased prefrontal and parietal cortical thickness does not correlate with anhedonia in patients with untreated first-episode major depressive disorders. Psychiatry Res. Neuroimaging 234 (1), 144–151.
Yang, X., Kumar, P., Nickerson, L.D., Du, Y., Wang, M., Chen, Y., Li, T., Pizzagalli, D.A., Ma, X., 2021. Identifying subgroups of major depressive disorder using brain structural covariance networks and mapping of associated clinical and cognitive variables. Biol. Psychiatry Glob. Open Sci. 1 (2), 135–145.
Yin, Y., Hou, Z., Wang, X., Sui, Y., Yuan, Y., 2015. Association between altered resting-state cortico-cerebellar functional connectivity networks and mood/cognition dysfunction in late-onset depression. J. Neural Transm. 122, 887–896.
Zhang, A., Li, G., Yang, C., Liu, P., Wang, Y., Kang, L., Wang, Y., Zhang, K., 2019. Alterations of amplitude of low-frequency fluctuation in treatment-resistant versus non-treatment-resistant depression patients. Neuropsychiatr. Dis. Treat. 15, 2119–2128.
Zhang, L., Bao, Y., Tao, S., Zhao, Y., Liu, M., 2022. The association between cardiovascular drugs and depression/anxiety in patients with cardiovascular disease: a meta-analysis. Pharmacol. Res. 175, 106024.
Zhang, Q., Hu, G., Tian, L., Ristaniemi, T., Wang, H., Chen, H., Wu, J., Cong, F., 2018. Examining stability of independent component analysis based on coefficient and component matrices for voxel-based morphometry of structural magnetic resonance imaging. Cogn. Neurodyn. 12, 461–470.