Neuropsychiatric Disease and Treatment

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Establishment and Validation of Machine Learning Model for Predicting Suicide Risk in Patients with Major Depressive Disorder

Hang Tan[1], Zhanjin Wang[2], Jie Ma[1], Zhan Wang[2], Xiangyang Zhang[3]

[1]Department of Nursing, The Affiliated Hospital of Qinghai University, Xining, People's Republic of China; [2]Department of Hepatobiliary and Pancreatic Surgery, The Affiliated Hospital of Qinghai University, Xining, People's Republic of China; [3]Hefei Fourth People's Hospital; Anhui Mental Health Center; Affiliated Psychological Hospital of Anhui Medical University, Hefei, People's Republic of China

Correspondence: Xiangyang Zhang, Hefei Fourth People's Hospital, Anhui Mental Health Center; Affiliated Psychological Hospital of Anhui Medical University, Hefei, People's Republic of China, Email zhangxy@psych.ac.cn; Zhan Wang, Department of Hepatobiliary and Pancreatic Surgery, The Affiliated Hospital of Qinghai University, Xining, People's Republic of China, Email ufofu01@163.com

Background: Suicide, a serious outcome in major depressive disorder (MDD), necessitates early risk detection for clinical intervention. This study developed machine learning models to predict suicide risk in MDD patients.The model incorporated both psychological (eg,the Hamilton Anxiety Rating Scale[HAMA], the Clinical Global Impression of Severity Scale [CGI-S]) and biological (eg, thyroid-stimulating hormone [TSH], Systolic blood pressure[SBP]) predictors, with CGI-S, HAMA, and SBP emerging as the top predictors.

Patients and Methods: We analyzed data from 1,718 first-episode medicated patients with MDD recruited from the psychiatric outpatient department of the First Affiliated Hospital of Shanxi Medical University (March 2016–June 2017). Feature selection was performed using Least absolute shrinkage and selection operator (LASSO) regression, and the importance of the selected features was ranked using SHAP values via the XGBoost algorithm. The features were incrementally incorporated into the model construction based on their importance. Eleven machine-learning algorithms were evaluated, and an optimized stacked ensemble model was developed using a stacking algorithm. Model performance was assessed using Receiver Operating Characteristic Curve, precision-recall (PR) curves, accuracy, recall, and F1 scores. Interpretability was enhanced using kernel-SHAP and LIME algorithms.

Results: The Cohort comprised 1,718 MDD patients (mean age 34.87 ± 12.43 years; 34.20% male). Eight key predictors were selected: CGI-S score, HAMA score, TSH, SBP, PANSS positive subscale score, Antithyroglobulin, Diastolic blood pressure (DBP) and age. The top predictors, included CGI-S, HAMA, and SBP, aligning with pathways involving autonomic dysregulation and anxiety-depression interplay. The stacked ensemble model demonstrated superior performance, achieving an Area Under Curve of 0.868 and a PR value of 0.665 on the test set, outperforming all the other models. Decision curve analysis (DCA) confirmed its clinical utility, showing the highest net benefit across a risk threshold range of 0.03–0.88. The SHAP method improved model interpretability and highlighted influential predictors.

Conclusion: The stacked ensemble model exhibited a strong predictive performance and clinical applicability for suicide risk assessment in patients with MDD. This tool may aid clinicians in the early identification and intervention of high-risk individuals and potentially improve patient outcomes.

Keywords: major depressive disorder, suicide attempt, machine learning, suicide prevention, predictive model

Introduction

Major depressive disorder (MDD) constitutes a substantial global health burden, primarily due to its robust association with suicidal behaviors. Epidemiological studies indicate that MDD accounts for approximately 50% of all suicide-related deaths, with affected individuals exhibiting a lifetime suicide rate of 2–12% and a notable 23.7% lifetime prevalence of suicide attempts.[1–3] These findings emphasize the urgent need for enhanced risk-assessment strategies in clinical practice.

Current suicide risk assessment tools, such as the Suicide Risk Assessment Tool, primarily focus on psychosocial factors, including demographic characteristics, substance use patterns, and psychological status. Although these instruments offer

https://doi.org/10.2147/NDT.S561526
Received: 18 August 2025
Accepted: 16 December 2025
Published: 19 January 2026

valuable screening capabilities, their dependence on subjective measures and limited integration of biological markers lead to suboptimal predictive performance.[4,5] Furthermore, the complex neurobiological mechanisms underlying suicidal behavior in MDD remain incompletely elucidated, underscoring the need for more comprehensive assessment approaches.

Recent studies have identified several promising biomarkers associated with suicide risk in patients with MDD. Liu et alreported a significant correlation between elevated thyroid-stimulating hormone (TSH) levels and increased suicide attempts in a large cohort (n=1,279) of Chinese outpatients with first-episode MDD.[6] Similarly, Zhou et al demonstrated the predictive value of total cholesterol (TC) levels for assessing suicide risk.[7] Neurobiological investigations have further revealed metabolic disturbances, including elevated fasting blood glucose (FBG) levels and reduced glucose metabolism in specific brain regions, in MDD patients with suicidal behavior.[8] Accumulating evidence suggests that metabolic and endocrine dysregulation—such as insulin resistance, dyslipidemia, and thyroid dysfunction—may constitute a shared pathophysiological pathway linking depression and suicidal behavior. These abnormalities can trigger chronic neuroinflammation and oxidative stress, thereby impairing the function of the prefrontal and limbic systems and affecting decision-making control and negative emotional processing.[9,10] Concurrently, such physiological dysregulation is often accompanied by autonomic nervous system (ANS) imbalance. A meta-analysis confirms that suicidal propensity is linked to reduced heart rate variability, a key index of ANS flexibility.[11] In patients with MDD, this imbalance often manifests as sympathetic overactivation (eg, elevated blood pressure) and reduced vagal tone, which are implicated in emotional dysregulation and impulsivity—core features of suicide risk.[12] Therefore, the inclusion of easily accessible biomarkers such as blood pressure, glucose, lipid profiles, and thyroid indices in predictive models is grounded in a solid neurobiological rationale. However, the diagnostic utility of individual biomarkers is limited by their poor specificity, indicating the need for multivariate predictive models.

Machine learning (ML) approaches provide robust solutions to this challenge by enabling the integration of complex multidimensional data. Chen et al successfully employed support vector machines (SVMs) to classify patients with MDD based on whole-brain functional connectivity (accuracy=88.50%) and stratify suicide risk (accuracies of 84.56% and 74.60% for different risk categories, respectively).[8] In a pioneering study, Yang et al developed ML models using epigenetic and transcriptomic data, achieving remarkable classification accuracies (up to 92.6%) for various diagnostic distinctions.[13] Although these advances are promising, barriers to practical implementation—particularly regarding the clinical feasibility of genomic profiling, which involves high costs, specialized laboratory requirements, and lack of routine availability in psychiatric practice—restrict their widespread adoption.

The present study addresses these limitations through several innovations. First, we developed an interpretable ML model incorporating explainable artificial intelligence (XAI) techniques—namely, SHapley Additive exPlanations (kernel-SHAP) and Local Interpretable Model-agnostic Explanations (LIME)—to enhance clinical translatability. Second, our analysis utilized a large, well-characterized cohort of patients with first-episode MDD, minimizing confounding through rigorous inclusion criteria. Finally, we focused on clinically accessible biomarkers and features to ensure their practical utility in real-world psychiatric clinical settings. Our approach aimed to provide clinicians with a robust, interpretable tool for identifying high-risk patients with MDD, thereby facilitating timely intervention and improving patient outcomes.

Materials and Methods

Study Design and Participants

We conducted a retrospective cohort study analyzing data from 1,718 first-episode, drug-naïve patients with MDD who received treatment at the psychiatric outpatient department of the First Hospital of Shanxi Medical University between March 1, 2016, and June 30, 2017. The study protocol was approved by the Institutional Review Board of the First Hospital of Shanxi Medical University (Approval No. 2016-Y27), and all participants provided written informed consent after receiving detailed explanations of the study objectives and procedures. Participants retained the right to withdraw from the study without any repercussions. The research team had no access to personally identifiable information during data collection and analysis. The hospital's data security office de-identified all personal information (eg, names and medical record numbers) before dataset release. No financial compensation was provided to participants, as this study involved secondary analysis of de-identified clinical data.

Diagnostic Criteria and Patient Selection

MDD diagnosis was established based on the following inclusion criteria: (1) Confirmation by two independent psychiatrists using the Structured Clinical Interview for DSM-IV (SCID), with inter-rater reliability assessed using Cohen's kappa coefficient (K=0.85); (2) No prior exposure to psychotropic medications; (3) Han Chinese ethnicity, aged 18–60 years (Han Chinese ethnicity was required to reduce genetic and sociocultural heterogeneity, which may confound biomarker and clinical associations).; (4) Willingness to participate and ability to provide informed consent; (5) Sufficient cognitive capacity to understand study procedures. The exclusion criteria were as follows: (1) comorbid psychiatric disorders (assessed using SCID), (2) severe physical illnesses (eg, uncontrolled endocrine disorders, malignancies), (3) pregnancy or lactation status, (4) history of substance abuse (excluding nicotine use), and (5) incomplete baseline clinical assessments.

Study Procedures

All eligible participants underwent comprehensive baseline evaluations, including: (1) demographic characteristics (age, sex, and education level); (2) clinical assessments (duration of the current episode, symptom severity); (3) laboratory tests (thyroid function and metabolic parameters); and (4) psychometric evaluations (depression and anxiety severity scales). A detailed flowchart (Figure 1) illustrates the participant recruitment and selection processes, including the number of excluded individuals and reasons for exclusion.

Clinical Measurements

We employed standardized psychometric instruments to assess key clinical features. The 17-item Hamilton Depression Rating Scale (HAMD-17) was used to quantify depressive symptom severity. This widely validated instrument assesses mood, guilt, suicidal ideation, and neurovegetative symptoms, with higher scores indicating greater severity of depression. The Hamilton Anxiety Rating Scale (HAMA) was used to assess anxiety symptomatology. Based on established cutoffs, we categorized patients into the following groups: comorbid anxiety (HAMA ≥18) and non-anxious (HAMA <18).[14] The Positive and Negative Syndrome Scale (PANSS) positive subscale was used to evaluate psychotic symptoms. Patients scoring ≥15 on this subscale were classified as having psychotic features, consistent with previous research in Chinese populations with MDD.[15] The Clinical Global Impression of Severity Scale (CGI-S) was used to assess disease severity.

Definition of Suicide Attempt

We defined suicide attempts as intentional self-injurious behaviors carried out with at least some intent to die.[16] During assessments, participants were directly asked, "Have you ever made an intentional attempt to end your life?" Specifically, a "yes" classification required meeting all of the following criteria: (1) a self-reported affirmative response to the aforementioned question; (2) corroboration of the event's intentionality and outcome (ie, self-injury) during a subsequent clinical interview by a trained psychiatrist; and (3) availability of documented details regarding the method, level of intent, and resulting medical severity. For those responding affirmatively, trained clinicians conducted semi-structured interviews to document characteristics of the attempts, including frequency, methods employed, temporal patterns, and medical severity.

Variables

All rating psychiatrists completed standardized training on scale administration, achieving excellent inter-rater reliability (intraclass correlation coefficient >0.85 for all scales). The Chinese versions of these instruments have demonstrated strong psychometric properties in validation studies.[17,18]

Based on prior research on mental health, we collected comprehensive biomarker data from all participants, including metabolic parameters (fasting blood glucose [FBG] measured via the glucose oxidase method, with levels >6.1 mmol/L indicating abnormal glucose metabolism), thyroid function markers (free triiodothyronine [FT3], free thyroxine [FT4], thyroid-stimulating hormone [TSH], antithyroglobulin antibody [TgAb], and thyroid peroxidase antibody [TPOAb]), and lipid profiles (total cholesterol [TC], high-density lipoprotein [HDL], triglycerides [TG], and low-density lipoprotein [LDL] measured via enzymatic colorimetry). Dyslipidemia was defined according to established criteria: TC ≥5.2 mmol/L, TG ≥1.7 mmol/L, LDL-C

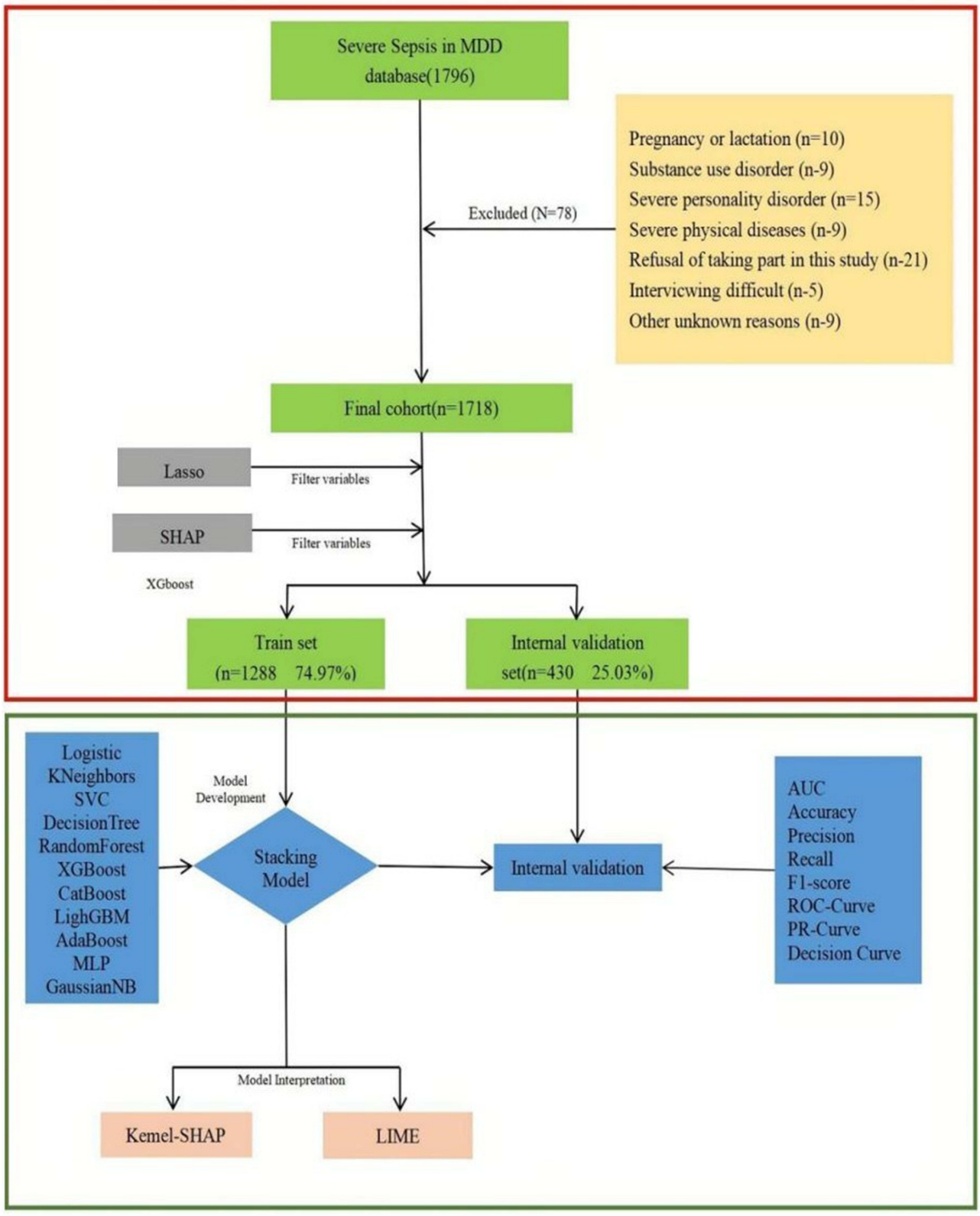

Figure 1 Flow chart.

≥3.4 mmol/L, or HDL-C <1.0 mmol/L.[19,20] Thyroid hormone reference ranges followed standard US guidelines: FT4 (10–23 pmol/L), TgAb (0–115 IU/L), TSH (0.27–4.20 mIU/L), FT3 (3.10–6.80 pmol/L), and TPOAb (0–34 IU/L).[21]

Referring to previous population-based literature, anthropometric measurements included body weight (kg) and height (m), from which we calculated body mass index (BMI) as weight/height^2 (kg/m^2), categorizing participants as underweight/normal (BMI <24 kg/m^2), overweight (24–28 kg/m^2), or obese (≥28 kg/m^2). Blood pressure assessment followed standardized protocols: after a 5-minute rest period, we measured diastolic blood pressure (DBP) and systolic blood pressure (SBP) using an automated electronic sphygmomanometer (HBP-9020; Omron, Japan), recording values in mmHg.[22,23] This comprehensive approach ensured robust characterization of the metabolic and cardiovascular risk factors in our study population.

Statistical Analysis

Variable Selection

All statistical analyses were performed using R software (version 4.3.1; R Foundation for Statistical Computing). Referring to previous literature, continuous variables were assessed for normality using the Shapiro–Wilk test, with normally distributed variables presented as mean ± standard deviation and non-normally distributed variables as median (interquartile range).[24,25] Between-group comparisons were conducted using independent samples t-tests for normally distributed continuous variables, Kruskal–Wallis tests for non-normally distributed continuous variables, and χ^2-tests or Fisher's exact tests for categorical variables, as appropriate (see Table 1). All statistical tests were performed using the compareGroups package (version 4.7.0) in R, with a two-tailed p-value <0.05 considered statistically significant.

Table 1 Characteristics of MDD Patients

Item	Non-Attempters N=1372	Attempters N=346	P
Age	33.0[23.0;45.0]	35.0[25.0;47.0]	0.023
Duration disease (months)	5.00[3.00;8.00]	6.00[3.00;9.00]	0.001
Age of onset	33.0[23.0;45.0]	34.0[25.0;47.0]	0.026
Gender			0.453
Female	476(34.7%)	112(32.4%)	
Male	896(65.3%)	234(67.6%)	
Education level (years)	12.76(9.87±15.65)	12.55(9.47±15.63)	0.365
Marriage			0.459
Yes	965(70.3%)	251(72.5%)	
No	407(29.7%)	95(27.5%)	
Severe anxiety			<0.001
Yes	93(6.78%)	111(32.1%)	
No	1279(93.2%)	235(67.9%)	
Psychiatric Symptom			<0.001
Yes	83(6.05%)	88(25.4%)	
No	1289(93.95%)	258(74.6%)	
HAMD	30.0[28.0;32.0]	32.0[30.0;34.0]	<0.001
HAMA	20.0[18.0;22.0]	23.0[21.0;26.0]	<0.001
Delusion	1.00[1.00;1.00]	1.00[1.00;3.00]	<0.001
Loosening	1.00[1.00;1.00]	1.00[1.00;2.00]	<0.001
Hallucination	1.00[1.00;1.00]	1.00[1.00;3.00]	<0.001
Exciting	1.00[1.00;1.00]	1.00[1.00;1.00]	<0.001
Suspicion	1.00[1.00;1.00]	1.00[1.00;3.00]	<0.001
Hostility	1.00[1.00;1.00]	1.00[1.00;2.00]	<0.001

(*Continued*)

Table 1 (Continued).

Item	Non-Attempters N=1372	Attempters N=346	P
PANSS positive subscale score	7.00[7.00;7.00]	8.00[7.00;17.8]	<0.001
CGI	6.00[5.00;6.00]	7.00[6.00;7.00]	<0.001
TSH	4.63[2.89;6.14]	6.76[4.54;8.89]	<0.001
TgAb	20.1[13.8;32.4]	27.9[18.1;144]	<0.001
TPOAb	16.4[12.2;29.0]	28.7[14.1;171]	<0.001
FT3	4.91[4.39;5.40]	4.92[4.34;5.44]	0.981
FT4	16.5[14.4;18.8]	16.5[14.4;18.6]	0.862
Glucose	5.28[4.92;5.71]	5.56[5.05;6.10]	<0.001
TC	5.11[4.36;5.81]	5.72[4.95;6.59]	<0.001
HDL-C	1.25[1.05;1.44]	1.12[0.89;1.30]	<0.001
TG	1.94[1.37;2.74]	2.16[1.46;2.93]	0.004
LDL-C	2.90[2.30;3.42]	3.21[2.60;3.74]	<0.001
BMI	24.2[23.2;25.6]	24.3[23.2;26.0]	0.811
SBP	120[111;126]	125[116;134]	<0.001
DBP	75.0[70.0;80.0]	78.0[74.0;84.0]	<0.001

Model Derivation and Validation

The machine learning prediction model was developed using Python 3.9 and R 4.3.1, following rigorous data preprocessing and feature engineering protocols. Feature selection was initially performed using least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation to identify the optimal regularization parameter (λ), resulting in the preliminary selection of 37 clinically relevant variables from the original dataset. Subsequently, we implemented an extreme gradient boosting (XGBoost) model to further refine the feature set. Using SHAP values, we quantitatively assessed and ranked feature importance, enabling stepwise variable inclusion based on their relative contributions to model performance. This iterative process was continued until the model achieved optimal predictive performance while maintaining parsimony, as evaluated via repeated cross-validation. Potential collinearity among predictors was assessed using variance inflation factors (VIF). Variables with VIF > 5 were excluded during LASSO regression to ensure feature independence. This step minimized multicollinearity and stabilized model performance.

We developed our predictive models using Python 3.9.10 with the scikit-learn library (version 1.3) and implemented a comprehensive machine learning framework incorporating 11 distinct algorithms: logistic regression (with L2 regularization), support vector machines (SVM with radial basis function kernel), k-nearest neighbors (KNN; k = 5), decision trees (maximum depth = 5), CatBoost (iterations = 1000), random forests (n_estimators = 100), XGBoost (learning_rate = 0.1), LightGBM (num_leaves = 31), AdaBoost (n_estimators = 50), multilayer perceptron neural networks (hidden_layer_sizes = (100)), and Gaussian naïve Bayes. To enhance predictive performance, we employed a stacking ensemble approach combining these diverse algorithms as base learners and logistic regression as the meta-learner to optimally weight their predictions. The stacking ensemble synergistically combines base learners (eg, XGBoost, SVM) via a meta-learner (logistic regression), which learns to weight their predictions optimally. This approach reduces variance and bias, outperforming individual models that may overfit to specific data patterns. Hyperparameter optimization was conducted via an exhaustive grid search combined with stratified 5-fold cross-validation, using the area under the receiver operating characteristic curve (AUC-ROC) as the primary performance metric for model selection.We employed stratified 5-fold cross-validation to ensure robust performance estimation, with each fold serving as a hold-out test set. This approach mitigates overfitting and provides a more generalizable performance metric than a single train-test split. This rigorous optimization process ensured that each algorithm's parameters were carefully tuned to maximize predictive accuracy while maintaining generalizability.

The predictive performance of all 12 machine learning models was rigorously evaluated using multiple complementary metrics. We assessed model discrimination using receiver operating characteristic (ROC) curve analysis, reporting the area under the curve (AUC) with 95% confidence intervals, and evaluated precision-recall (PR) curves to account for

class imbalance. For logistic regression models, AIC was used for model selection. For tree-based and ensemble models, cross-validated AUC and F1 score were prioritized due to their suitability for non-linear and imbalanced classification tasks. Additional performance metrics included accuracy, precision, recall, and F1 score, all of which were calculated using stratified 5-fold cross-validation to ensure reliable estimates. Following a comprehensive evaluation, we selected the optimal model based on balanced performance across all metrics. To enhance clinical interpretability, we applied interpretable machine learning (XAI) techniques, including kernel-SHAP (SHapley Additive exPlanations) and LIME algorithms, were employed to provide transparent, quantitative insights into how each feature influences model predictions, both globally and at the individual level. This dual-interpretation approach facilitates a transparent understanding of the model's decision-making process while maintaining the complexity required for accurate predictions.

Results

Study Cohort

The study cohort comprised 1,718 patients diagnosed with MDD, among whom 346 (20.14%) reported at least one suicide attempt, while 1,372 (79.86%) had no history of suicidal behavior. Between-group comparisons are summarized in Table 1, although some variables showed statistical differences, no single demographic or conventional clinical marker alone provided sufficient discriminative power for accurate suicide risk stratification, highlighting the need for multivariate machine learning approaches.

Model Evaluation

Our model development adopted a systematic stepwise variable inclusion approach to optimize predictive performance. Initial feature selection using LASSO regression with 10-fold cross-validation identified 37 potentially relevant variables from the original dataset. These variables were then sequentially incorporated into the model based on their SHAP (Shapley Additive exPlanations) values, which quantitatively assessed each feature's contribution to model predictions (Figure 2a–c). Model performance, as measured by AUC, exhibited dynamic changes with variable inclusion (Table 2), with these fluctuations visually represented in the performance heatmap (Figure 3). Notably, the stacked ensemble model achieved peak discriminative ability (AUC = 0.868, 95% CI: 0.842–0.894) and precision (0.849) in the validation set when incorporating the top eight most influential variables based on SHAP rankings. Further variable inclusion beyond this point yielded diminishing returns, with model performance stabilizing, as shown in the trend analysis of AUC and accuracy metrics (Figure 4). This plateau effect indicated that optimal predictive capacity could be maintained with only the top eight variables, which were consequently selected for final model implementation. Eight key predictors were

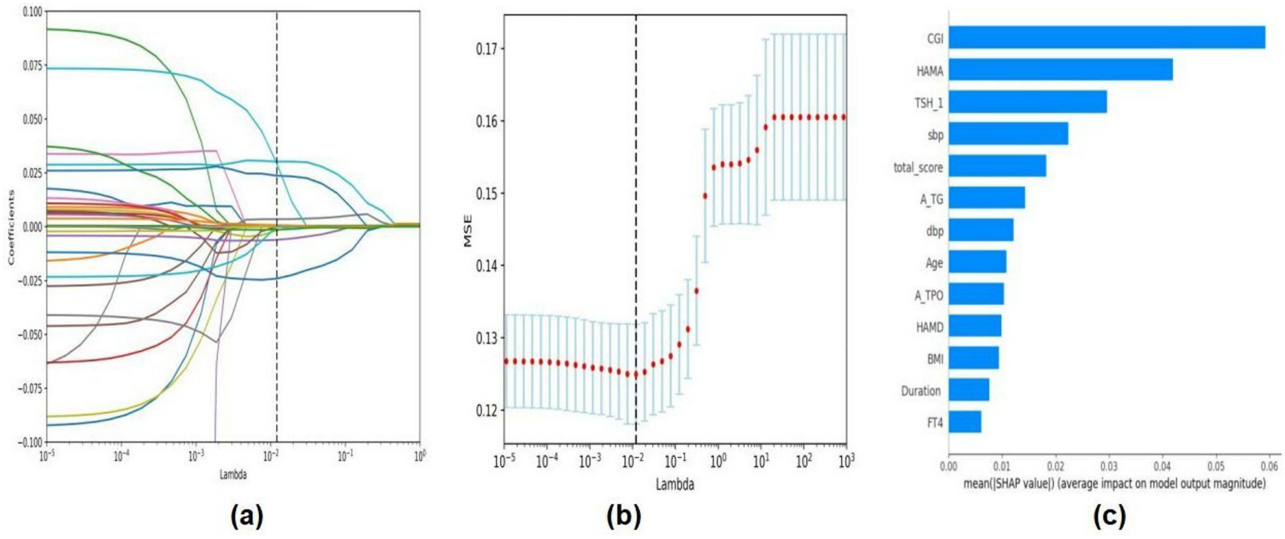

Figure 2 (**a**) LASSO regression coefficient path plot; (**b**) LASSO regularized path diagram; (**c**) Ranking of variables based on SHAP values.

https://doi.org/10.2147/NDT.S561526

Table 2 Changing AUC Values as Variables Increase for Different Models

Model	1	2	3	4	5	6	7	8	9	10	11	12	13
Logit	0.72	0.80	0.82	0.82	0.82	0.82	0.81	0.81	0.83	0.82	0.83	0.84	0.84
KNN	0.72	0.79	0.82	0.85	0.85	0.82	0.83	0.84	0.84	0.84	0.85	0.83	0.81
SVM	0.73	0.80	0.82	0.83	0.83	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.83
Cat	0.72	0.81	0.83	0.84	0.84	0.85	0.85	0.85	0.85	0.86	0.86	0.86	0.86
DT	0.71	0.81	0.80	0.79	0.79	0.78	0.79	0.79	0.81	0.80	0.79	0.78	0.78
RF	0.73	0.81	0.83	0.84	0.84	0.85	0.85	0.85	0.86	0.86	0.87	0.87	0.86
XGBoost	0.73	0.81	0.82	0.83	0.83	0.83	0.84	0.85	0.85	0.85	0.87	0.86	0.86
LightGBM	0.73	0.81	0.83	0.84	0.84	0.85	0.84	0.86	0.86	0.85	0.87	0.87	0.86
AdaBoost	0.72	0.81	0.82	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.85
MLP	0.72	0.80	0.82	0.81	0.81	0.81	0.82	0.82	0.83	0.84	0.84	0.80	0.81
GaussianNB	0.71	0.80	0.81	0.82	0.82	0.81	0.81	0.81	0.82	0.83	0.83	0.83	0.83
Stacking	0.73	0.81	0.83	0.85	0.85	0.85	0.87	0.87	0.87	0.86	0.87	0.86	0.86

ultimately selected for the final stacked ensemble model: CGI-S score, HAMA score, TSH, SBP, PANSS positive subscale, TgAb, DBP and age. Model performance metrics are shown in Table 3, Figures 5 and 6. Comprehensive performance metrics for all evaluated models in the internal validation set are presented in Table 3. In contrast, the discriminative performance of the stacked model was further illustrated through its ROC (Figure 5) and PR curves (Figure 6) in the validation cohort.

To enhance the interpretability of our stacked ensemble model—which incorporates 12 distinct machine learning algorithms—we implemented two complementary XAI techniques: Kernel-SHAP and LIME. Kernel-SHAP provides theoretically grounded feature importance estimates by integrating the mathematical rigor of Shapley values from game theory with the local approximation approach of LIME, enabling comprehensive model interpretation at both the

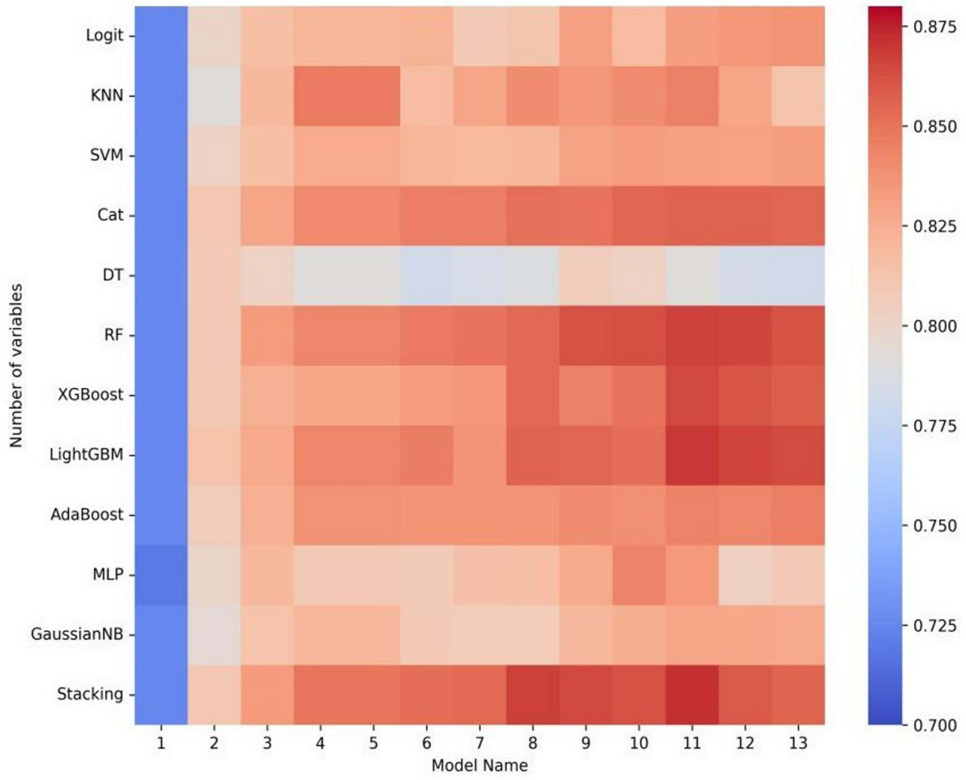

Figure 3 AUC of feature selection based on SHAP variable importance for various models.

https://doi.org/10.2147/NDT.S561526

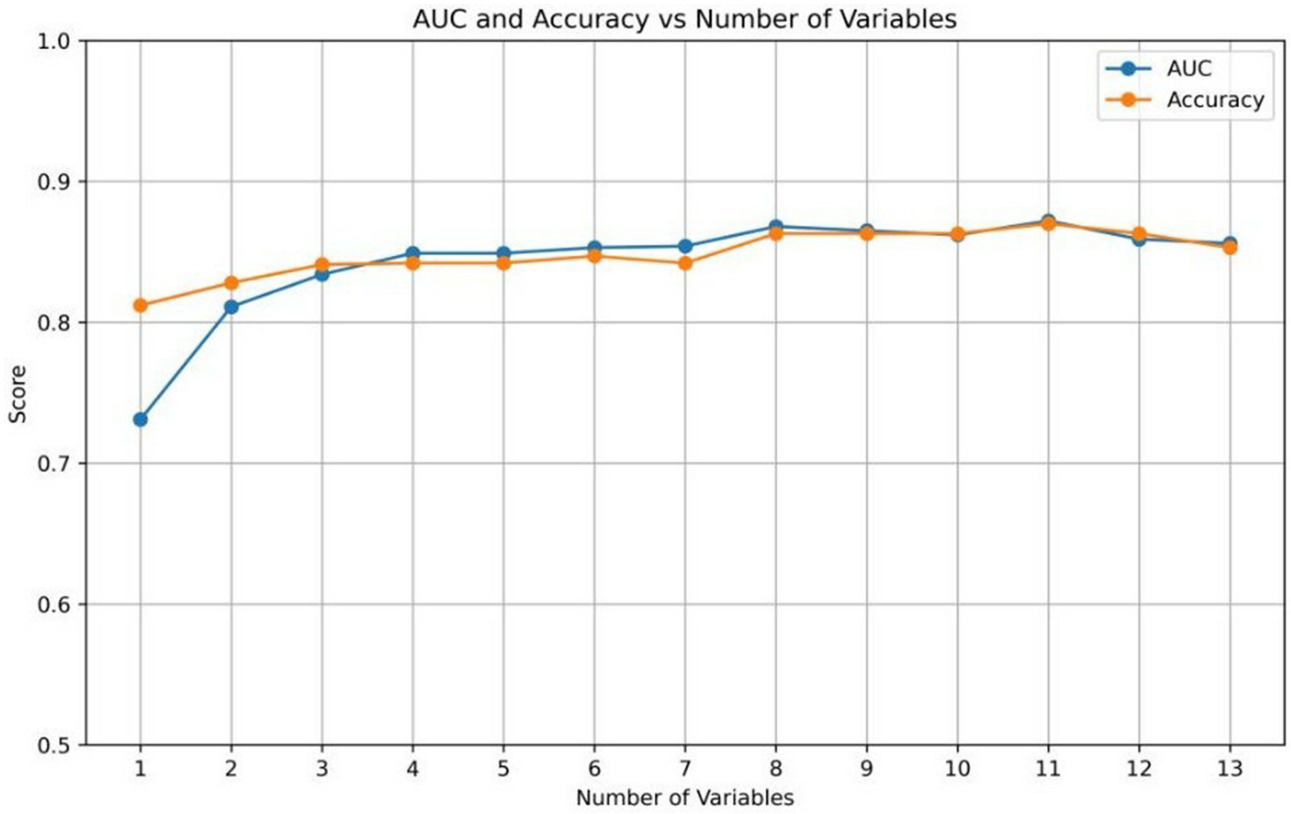

Figure 4 Changes in AUC and accuracy of the Stacking model as the number of variables increases.

population and individual levels. Our analysis identified eight clinically relevant predictors of suicidal behavior in patients with MDD, as ranked by SHAP values (see Figure 7) CGI-S score, HAMA score, TSH, SBP, DBP, TgAb levels, PANSS positive subscale score, and age. Individual prediction features were elucidated using the SHAP Beeswarm plot (Figure 7) and Force plot (Figure 8), which visually demonstrate how each feature contributes to specific risk assessments. As shown in Figure 7, HAMA score points become increasingly red toward the right, and the SHAP value of HAMA is positive, indicating that higher HAMA scores are associated with an increased risk of suicide. Figure 8 quantifies the magnitude of the effect of each influencing factor on suicide risk in patients with MDD. Figure 8 demonstrates that the Hamilton Anxiety Rating Scale (HAMA) score, SBP, and Clinical Global Impressions-Severity

Table 3 Internal Validation

Model	AUC	Precision	Accuracy	Recall	F1
Logit	0.82	0.43	0.76	0.77	0.55
KNN	0.84	0.84	0.84	0.20	0.32
SVM	0.82	0.40	0.74	0.77	0.52
Cat	0.85	0.48	0.78	0.73	0.55
DT	0.79	0.38	0.72	0.75	0.50
RF	0.85	0.51	0.82	0.59	0.55
XGBoost	0.85	0.38	0.71	0.85	0.53
LightGBM	0.86	0.53	0.82	0.61	0.56
AdaBoost	0.84	0.45	0.78	0.77	0.57
MLP	0.82	0.64	0.85	0.42	0.51
GaussianNB	0.81	0.46	0.80	0.38	0.42
Stacking	0.87	0.63	0.85	0.47	0.54

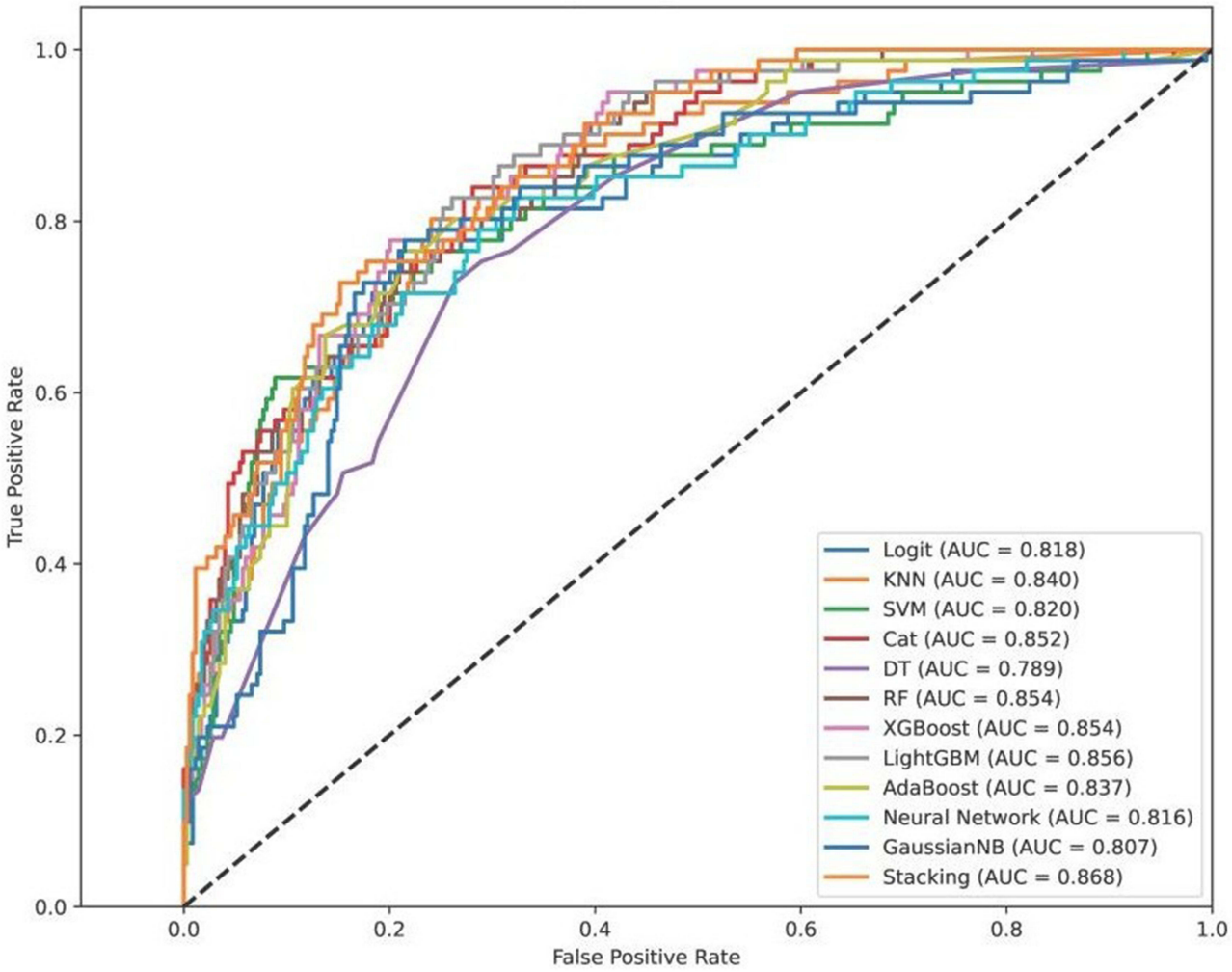

Figure 5 The ROC curves of each model on the internal validation set based on LASSO feature filtering.

Scale (CGI-S) score were the top three predictors of suicide risk in patients with major depressive disorder (MDD). For clinical translation, the LIME plot in Figure 9 provides a localized, intuitive explanation, showing that a specific patient with MDD had an 85% predicted risk of suicidal behavior, with factors such as CGI-S, HAMA score, SBP, TSH, age, and TgAb driving this prediction.

Validation of the Predictive Model

We validated the clinical utility of our predictive model using decision curve analysis (DCA), which quantifies net benefit across different risk thresholds by balancing true-positive predictions against false-positive results. As shown in Figure 10, the model demonstrated superior clinical net benefit compared to both "treat-all" and "treat-none" strategies across a clinically relevant risk threshold range of 3% to 88% (0.03–0.88 probability range). This wide effective threshold interval indicates robust clinical applicability, with maximum net benefit improvements of 38% observed at intermediate-risk thresholds (15–65%), where clinical decision-making is most uncertain. The sustained positive net benefit across nearly the entire risk spectrum indicates that our model could meaningfully inform suicide prevention strategies while minimizing unnecessary interventions, particularly in the critical mid-range probability zone, where clinicians often face diagnostic uncertainty. We evaluated the internal validation performance of the optimal model. On the test set, the model achieved a sensitivity of 0.4815 (48.15%) and a specificity of 0.9356 (93.56%). To assess the robustness of the model, we performed random 5-fold cross-validation on the dataset. Further validation on the 5-fold

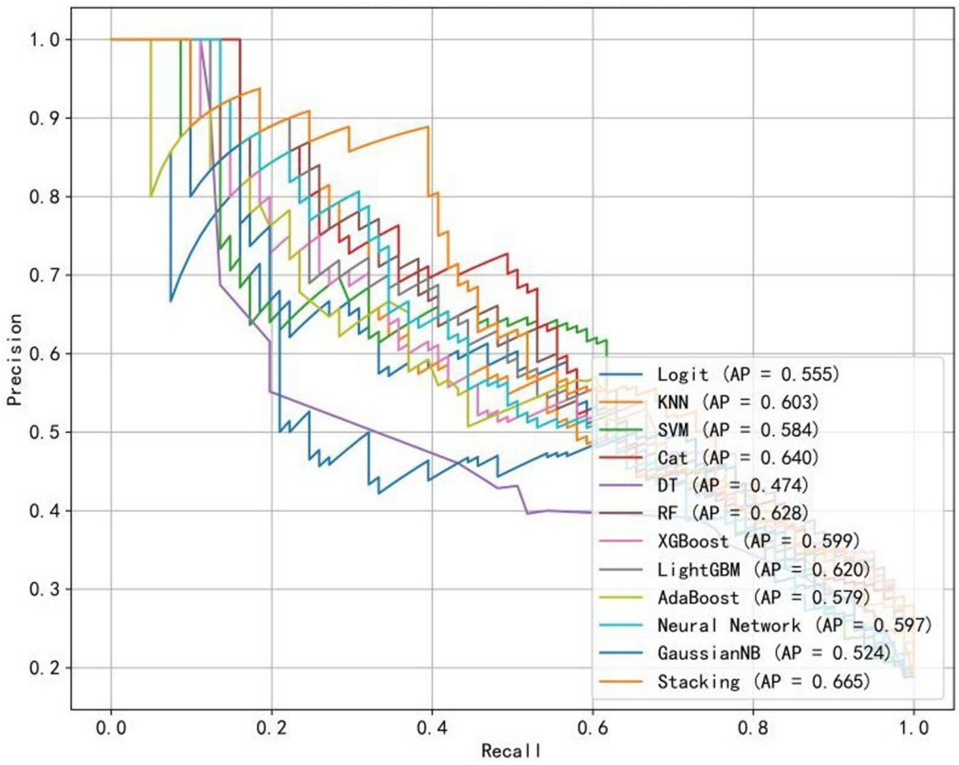

Figure 6 The PR curves of each model on the internal validation set based on LASSO feature filtering.

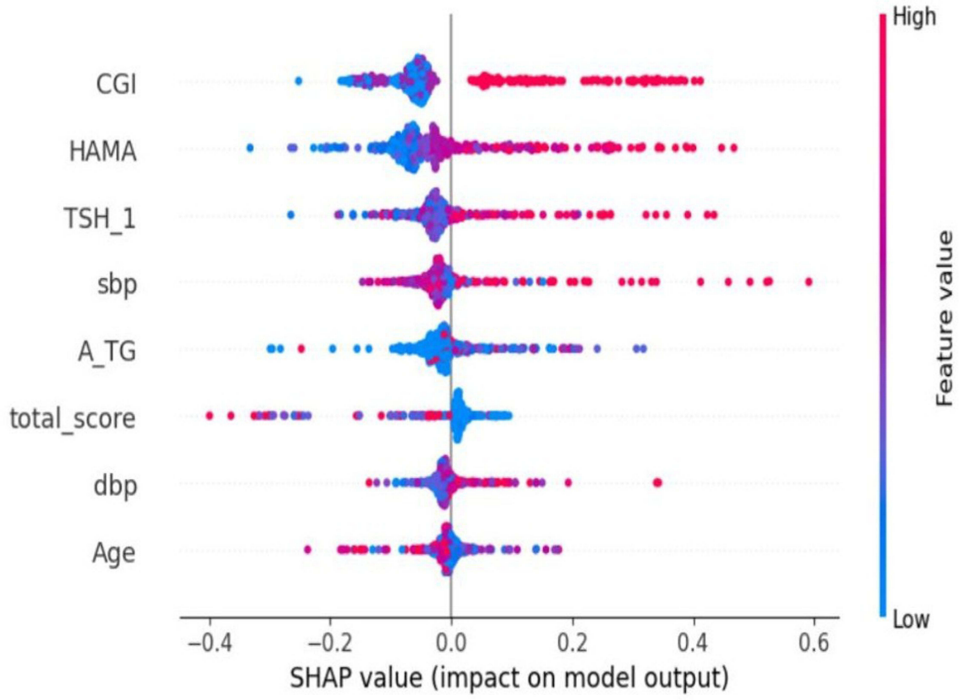

Figure 7 SHAP Beeswarm Plot of Feature attributes in the stacking ensemble model (each line represents a feature, and the x-axis represents SHAP values, indicating the impact of the feature on the outcome. Each point represents a sample. The redder the colour, the larger the feature value; the bluer the colour, the smaller the feature value).

https://doi.org/10.2147/NDT.S561526

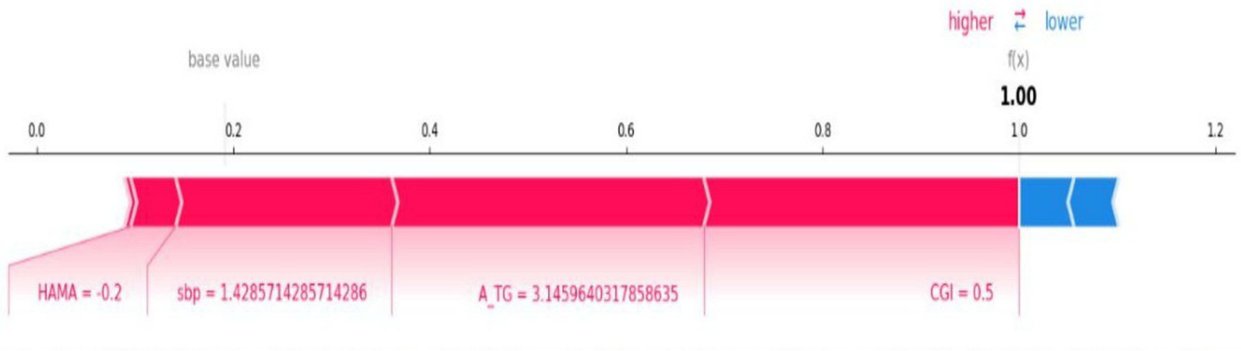

Figure 8 SHAP Force Plot for predicting suicide risk in patients with MDD (The SHAP values are indicative of the prediction-related features of individual patients and the contribution of each feature to the suicide behavior prediction. The bold-faced numbers represent the probabilistic predicted values (f(x)). In contrast, the base values correspond to the values predicted without providing input to the model. The f(x) is defined as the log odds ratio of each observation. The red features (on the left) indicate characteristics that have been demonstrated to increase the risk of suicide behavior, while the blue features indicate characteristics that have been demonstrated to decrease the risk of suicide behavior. The length of the arrows facilitates the visualization of the magnitude of the effect on the prediction. The magnitude of the effect is directly proportional to the length of the arrow).

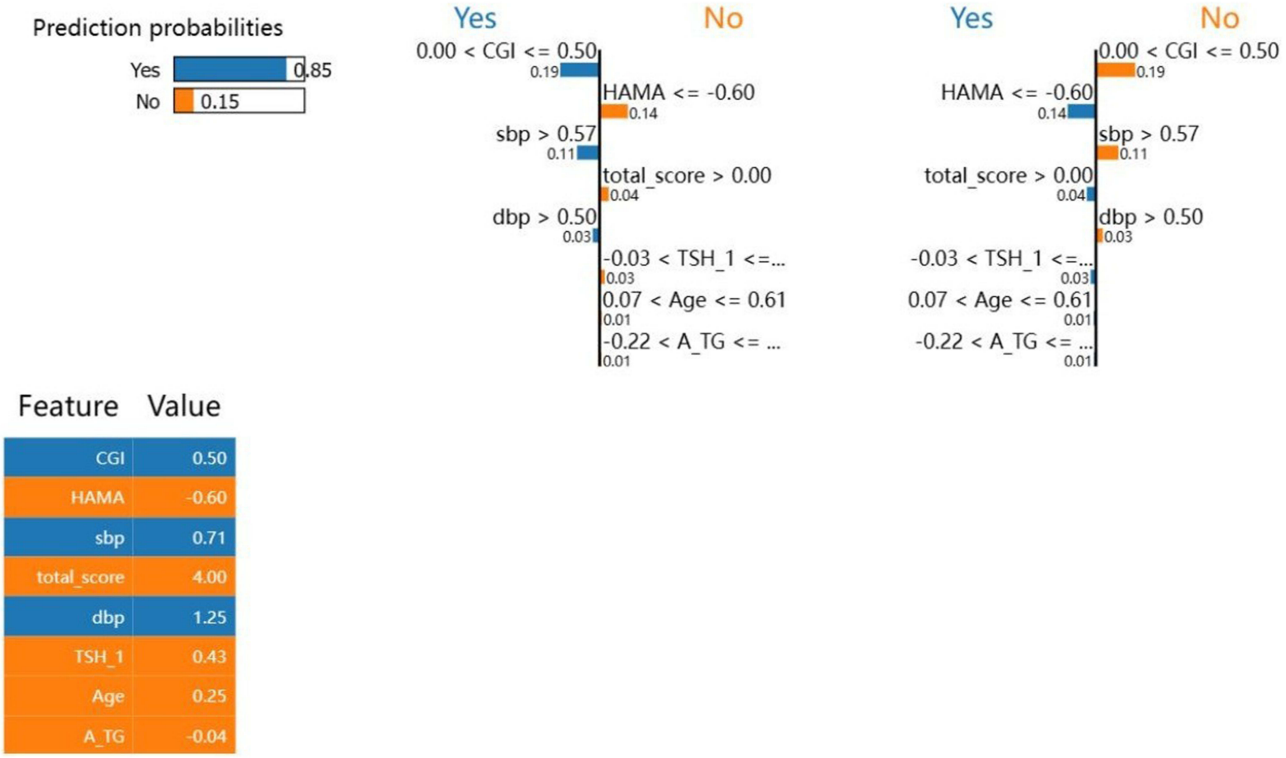

Figure 9 LIME plot for predicted risk of suicidal behavior of a particular patient with MDD.

splits showed stable performance across different data subsets, with a mean AUC of 0.859 (SD = 0.0182), as shown in Figure 11. These results further confirm the reliability of the model's performance. These results confirm that the model's predictive performance translates into tangible clinical value for risk stratification in MDD management.

Discussion

Patients with MDD face an elevated risk of recurrent suicidal behaviors following initial attempts, yet clinicians predominantly rely on empirical clinical assessments for risk stratification in non-acute cases.[26] Our study addresses this critical gap by developing a stacked ensemble machine learning model integrating 12 algorithms optimized via

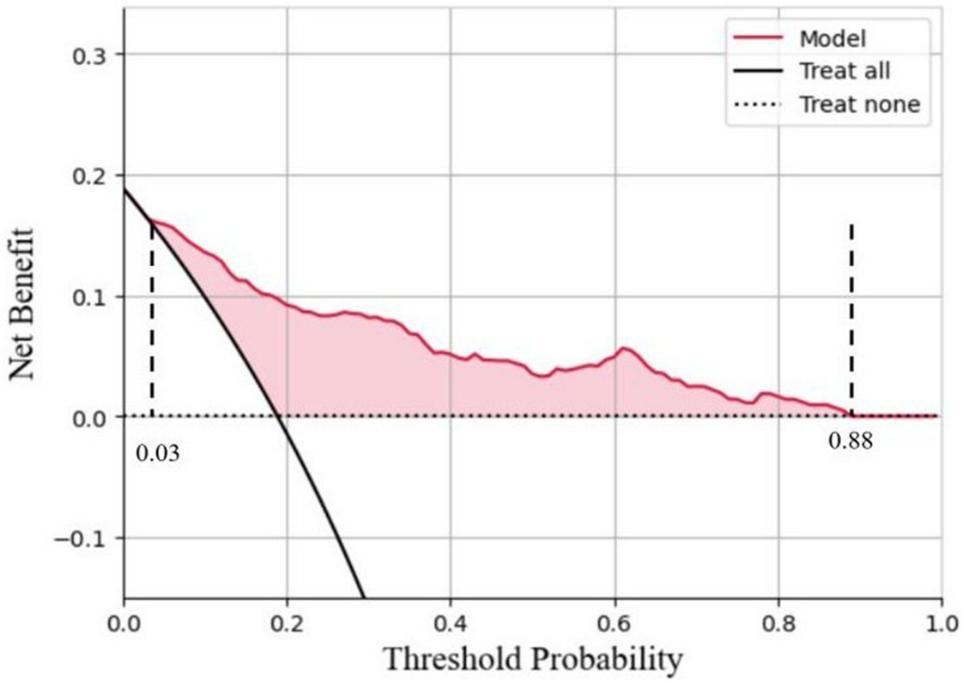

Figure 10 Clinical Decision Curve.

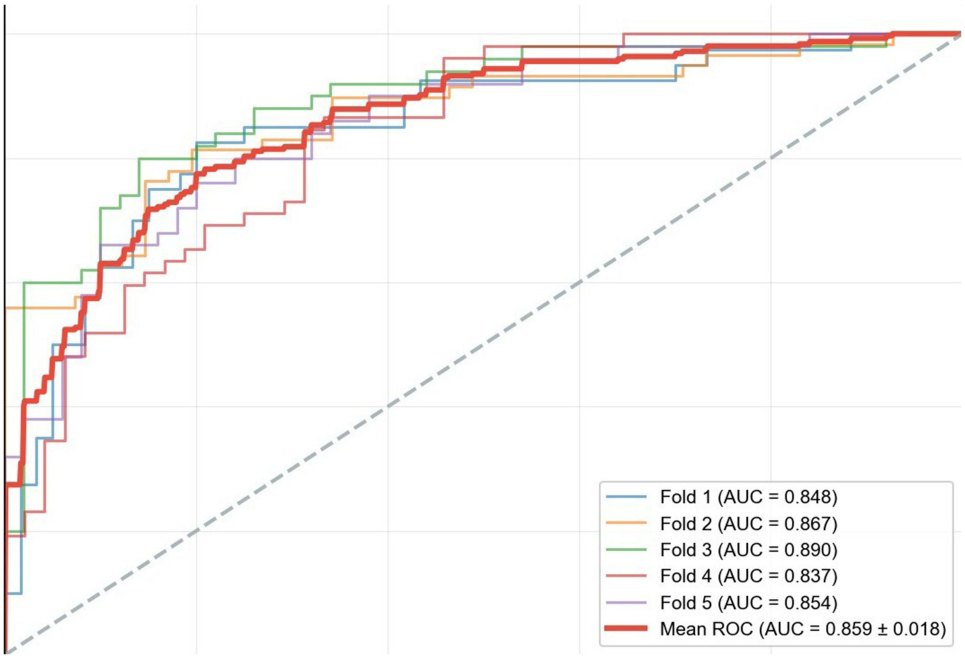

Figure 11 5-fold Cross-Validation ROC Curves.

LASSO regression and SHAP-based feature selection. By leveraging comprehensive clinical and biomarker data from first-episode, medication-naïve patients with MDD, the model demonstrated superior predictive accuracy (AUC=0.868) compared to individual algorithms such as XGBoost and LightGBM. This advancement suggests this model provides a clinically feasible tool that integrates routinely available data, enabling early identification of high-risk patients without requiring specialized tests. Its interpretability further supports trust and adoption among clinicians. And enabling targeted interventions to mitigate suicide-related morbidity.

Model interpretability remains paramount for the clinical adoption of artificial intelligence tools. Our approach prioritized transparency through three key strategies: (1) restricting input variables to routinely collected clinical parameters (eg, vital signs and serum biomarkers) and validated psychometric scales; (2) implementing SHAP and LIME frameworks to quantify feature contributions; and (3) validating clinical utility via decision curve analysis. The stacking model in this study exhibits high interpretability and provides clinicians with insights into how model predictions are generated. This will empower medical staff to formulate evidence-based interventions in a more timely and targeted manner, thereby reducing the incidence of suicide among patients with major depressive disorder (MDD). While individual clinical markers may show univariate associations with suicide risk, their predictive utility is limited when used in isolation. Our model demonstrates that it is the integration and weighted combination of these markers—particularly CGI-S, HAMA, and SBP—within a multivariate framework that achieves high discriminative accuracy. By intervening in relevant risk factors of patients with MDD to mitigate the tendency toward suicide attempts, a virtuous cycle may be formed, which can in turn lower the risk of suicide in these patients. This represents an advantage over previous studies, which failed to identify such risk factors.

The model maintained robust performance (net benefit range: 3–88% risk threshold) while identifying CGI-S scores, HAMA scores, and SBP as primary predictors. The identified predictors align with emerging pathophysiological models of suicidal behavior.[22–29] CGI-S, reflecting global illness severity, and CGI scores in patients with MDD are positively correlated with suicidal ideation and behavior, and suicide attempts are a symptom cluster in MDD.[30–32]The significance of HAMA scores supports the anxiety-depression suicide paradigm.[33] Additionally, our previous study showed that the incidence of suicide attempts in patients with MDD and comorbid anxiety symptoms was 9.51 times higher than in patients with MDD without anxiety symptoms.[34] Mechanistically, autonomic nervous system dysregulation, manifested through blood pressure abnormalities, may bridge depressive pathophysiology and suicidal vulnerability.[35–37] Our findings extend prior work by demonstrating that hypertension-associated parameters (SBP/DBP) provide incremental predictive value beyond traditional psychiatric assessments.[38,39] Notably, the SHAP force plots (Figure 8) and LIME explanations (Figure 9) provided granular insights into how specific biomarker profiles (eg, elevated TSH and TgAb levels) interact with clinical features to modulate individual risk trajectories. These readily accessible variables have the potential to predict suicidal behaviors in patients with MDD prospectively. A comparative analysis showed that the inclusion of biomarkers (eg, TSH, TgAb, SBP) improved the AUC from 0.79 (clinical scales only) to 0.87 (full model), representing a clinically meaningful increase in predictive accuracy.

The identification of systolic and diastolic blood pressure as key predictors in our model aligns with emerging neurobiological models of suicide risk that emphasize autonomic nervous system (ANS) dysregulation. Reduced heart rate variability, indicative of ANS inflexibility and particularly diminished vagal tone, is a well-established transdiagnostic biomarker for deficits in emotion regulation and behavioral control, core features of suicidal vulnerability.[11] In patients with MDD, specific patterns of autonomic activity have been empirically linked to the presence of suicidal ideation.[40] Elevated blood pressure may thus represent a measurable peripheral manifestation of sustained sympathetic overactivity, which can disrupt prefrontal-limbic circuitry. Furthermore, this autonomic imbalance is often part of a broader physiological state involving immune-metabolic dysregulation. Systemic inflammation and metabolic disturbances can both contribute to and be exacerbated by ANS dysfunction, creating a feed-forward loop that heightens neuroinflammation and impairs brain homeostasis, thereby increasing suicide risk.[41] Consequently, our model's incorporation of readily available cardiovascular indices provides a clinically feasible window into these underlying, interconnected neurobiological pathways.

This study advances the field through two methodological innovations: (1) the integration of serum biomarkers into interpretable ML frameworks, bridging the biological and clinical risk dimensions; and (2) the development of a continuous risk prediction tool with quantified uncertainty estimates. However, this study has several important limitations that warrant consideration and direct future research. The single-center retrospective design and homogeneous sample of Han Chinese first-episode drug-naïve patients, while methodologically controlled, may limit generalizability to other populations, healthcare settings, and recurrent or comorbid cases. Furthermore, reliance on static, single-time-point assessments cannot capture the dynamic nature of suicide risk, and the exclusion of variables like detailed psychosocial stressors or neuroimaging data may constrain predictive scope. Additionally, the current model exhibited a trade-off in

performance metrics, with sensitivity that may be suboptimal for some high-stakes clinical screening scenarios. While internal validation was rigorous, external validation in independent, prospective cohorts is essential to confirm real-world clinical utility. To address these constraints, future work should prioritize such multi-center validation, integrate dynamic digital phenotyping and broader multimodal data, technically optimize the model (eg, via advanced resampling techniques like SMOTE or cost-sensitive learning algorithms) to better balance sensitivity and specificity, elucidate the causal pathways underlying key predictors, and ultimately assess the model's implementation and impact within clinical workflows to translate predictive accuracy into tangible patient benefit.

Conclusion

This study demonstrates that a stacked ensemble machine learning model integrating clinical, biomarker, and psychometric data achieves superior suicide risk prediction (AUC=0.868) compared to conventional methods in patients with first-episode MDD. Key predictors, including CGI-S, HAMA scores, and blood pressure variability, align with the neurobiological pathways underlying autonomic nervous system dysregulation and anxiety-depression interaction. Two innovations enhance clinical translatability: interpretable machine learning (SHAP/LIME) and validated risk stratification across broad probability thresholds. Limitations include single-center recruitment and a cross-sectional design, necessitating future prospective validation across diverse populations with the integration of real-time digital biomarkers.

Institutional Review Board Statement

This study was conducted in accordance with the principles of the Declaration of Helsinki.

The core data of this study were derived from a primary clinical data collection project conducted at the Psychiatric Outpatient Department of the First Hospital of Shanxi Medical University, which enrolled 1,718 patients with major depressive disorder (MDD) between March 2016 and June 2017.

1.Ethical Approval for the Primary Project. This primary data collection project was approved by the Ethics Committee of the First Hospital of Shanxi Medical University (Approval No.: 2016-Y27). During the primary project, all participants provided written informed consent voluntarily after psychiatrists fully explained the study protocol and procedures; participants also retained the right to withdraw from the project at any time without penalty.

2.Ethical Compliance of This Secondary Analysis. Within the framework of the academic cooperation and sharing mechanism signed with the First Hospital of Shanxi Medical University, our team obtained the de-identified dataset of the primary project through a formal data transfer process. The de-identification process strictly adhered to national healthcare data security standards (eg, deletion of personally identifiable information such as name and ID number). Our current study is a retrospective secondary analysis of this de-identified dataset and does not involve prospective patient enrollment or intervention. Therefore, it poses no additional risk to the original participants and is fully compliant with the ethical guidelines of Dove Medical Press.

Abbreviations

MDD, Major Depressive Disorder; TSH, Thyroid-stimulating hormone; TC, Total cholesterol; FBG, Fasting blood glucose; ML, Machine learning; XAI, Explainable artificial intelligence; SVM, Support vector machine; Kernel-SHAP, SHapley Additive exPlanations. LIME, Local Interpretable Model-agnostic Explanations; HAMD, The Hamilton Depression Rating Scale; HAMA, The Hamilton Anxiety Rating Scale; PANSS, The Positive and Negative Syndrome Scale; CGI-S, The Clinical Global Impression-Severity score; FT3, Free triiodothyronine; FT4, Free thyroxine; TgAb, Antithyroglobulin; TPOAb, Thyroid peroxidase antibody; HDL, High-density lipoprotein; TG, Triglyceride; LDL, Low-density lipoprotein; BMI, Body mass index; DBP, Diastolic blood pressure; SBP, Systolic blood pressure; LASSO, Least absolute shrinkage and selection operator; XGBoost, Extreme gradient boosting; SVM, Support vector machines; KNN, K-nearest neighbors; ROC, Receiver Operating Characteristic Curve; AUC, Area Under Curve; PR, Public relations curves; DCA, The decision curve analysis.

Data Sharing Statement

The data supporting this study's findings are available upon request from the corresponding author, Xiangyang Zhang. However, the data are not publicly available because of restrictions such as containing information that could compromise the privacy of research participants.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was funded by the National Natural Science Foundation of China, grant number 82160131, National Natural Science Foundation of Qinghai, grant number 2021-ZJ-963Q. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the paper; or in the decision to publish the results.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Herpertz U. Prevention of suicidal behavior. *Dialogues Clin Neurosci*. 2016;18(2):183–190. doi:10.31887/DCNS.2016.18.2/uhegerl
2. Fang X, Yang J, Wu YJ, et al. Association between suicide risk severity and sarcopenia in non-elderly Chinese inpatients with major depressive disorder. *BMC Psychiatry*. 2020;20(1):345. doi:10.1186/s12888-020-02763-1
3. Dong M, Wang SB, Li Y, et al. Prevalence of suicidal behaviors in patients with major depressive disorder in China: a comprehensive meta-analysis. *J Affect Disord*. 2018;225:32–39. doi:10.1016/j.jad.2017.07.043
4. Belsher BE, Smolenski DJ, Pruitt LD, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. *JAMA Psychiatry*. 2019;76(6):642–651. doi:10.1001/jamapsychiatry.2019.0174
5. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. 2017;143(2):187–232. doi:10.1037/bul0000084
6. Liu W, Wu Z, Sun M, et al. Association between fasting blood glucose and thyroid-stimulating hormones and suicidal tendency and disease severity in patients with major depressive disorder. *Bosn J Basic Med Sci*. 2022;22(4):635–642. doi:10.17305/bjbms.2021.6754
7. Zhou S, Zhao K, Shi X, et al. Serum lipid levels and suicide attempts within 2 weeks in patients with major depressive disorder: is there a relationship? *Front Psychiatry*. 2021;12:676040. doi:10.3389/fpsyt.2021.676040
8. Chen S, Zhang X, Lin S, et al. Suicide risk stratification among major depressed patients based on a machine learning approach and whole-brain functional connectivity. *J Affect Disord*. 2023;322:173–179. doi:10.1016/j.jad.2022.11.022
9. Brundin L, Erhardt S, Bryleva EY, et al. The role of inflammation in suicidal behaviour. *Acta Psychiatr Scand*. 2015;132(3):192–203. doi:10.1111/acps.12458
10. Penninx BWJH, Lange SMM. Metabolic syndrome in psychiatric patients: overview, mechanisms, and implications. *Dialogues Clin Neurosci*. 2018;20(1):63–73. doi:10.31887/DCNS.2018.20.1/bpenninx
11. Beauchaine TP, Thayer JF. Heart rate variability as a transdiagnostic biomarker of psychopathology. *Int J Psychophysiol*. 2015;98(2 Pt 2):338–350. doi:10.1016/j.ijpsycho.2015.08.004
12. Chang HA, Chang CC, Chen CL, et al. Heart rate variability in patients with fully remitted major depressive disorder. *Acta Neuropsychiatr*. 2013;25(1):33–42. doi:10.1111/j.1601-5215.2012.00658.x
13. Bhak Y, Jeong HO, Cho YS, et al. Depression and suicide risk prediction models using blood-derived multi-omics data. *Transl Psychiatry*. 2019;9(1):262. doi:10.1038/s41398-019-0595-2
14. Yu WQ, Zhang GY, Jiang QF, et al. Prevalence and clinical profiles of comorbid anxiety in first episode and drug naïve patients with major depressive disorder. *J Affect Disord*. 2019;257(1):200–206. doi:10.1016/j.jad.2019.06.052
15. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261–276. doi:10.1093/schbul/13.2.261
16. Posner K, Oquendo MA, Gould M, et al. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *Am J Psychiatry*. 2007;164(7):1035–1043. doi:10.1176/ajp.2007.164.7.1035
17. Lin J, Wang X, Dong F, et al. Validation of the Chinese version of the Hamilton rating scale for depression in adults with epilepsy. *Epilepsy Behav*. 2018;89:148–152. doi:10.1016/j.yebeh.2018.10.009
18. Mingyuan Z, Yanling H. *Handbook of Psychiatric Rating Scales*. Hunan: Hunan Science and Technology Press; 2015:513.
19. Chisolm DJ, Keedy HE, Hart LC, et al. Exploring health literacy, transition readiness, and healthcare utilization in medicaid chronically ill youth. *J Adolesc Health*. 2021;69(4):622–628. doi:10.1016/j.jadohealth.2021.03.023
20. You Y, Wang R, Li J, et al. The role of dietary intake of live microbes in the association between leisure-time physical activity and depressive symptoms: a population-based study. *Appl Physiol Nutr Metab*. 2024;49(8):1014–1024. doi:10.1139/apnm-2023-0550

21. Garber JR, Cobin RH, Gharib H, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Endocr Pract*. 2012;18(6):988–1028. doi:10.4158/EP12280.GL
22. You Y, Chen Y, Ding H, et al. Relationship between physical activity and DNA methylation-predicted epigenetic clocks. *NPJ Aging*. 2025;11(1):27. doi:10.1038/s41514-025-00217-0
23. You Y, Wang D, Ding H, et al. Mediation role of telomere length in the relationship between physical activity and PhenoAge: a population-based study. *J Exerc Sci Fit*. 2025;23(3):149–156. doi:10.1016/j.jesf.2025.03.004
24. You Y, Ablitip A, Chen Y, et al. Saturation effects of the relationship between physical exercise and systemic immune inflammation index in the short-sleep population: a cross-sectional study. *BMC Public Health*. 2024;24(1):1920. doi:10.1186/s12889-024-19432-7
25. You Y, Ding H, Tang M, et al. Dose-response relationship between leisure-time physical activity and metabolic syndrome in short sleep US adults: evidence from a nationwide investigation. *Appl Physiol Nutr Metab*. 2025;50:1–10. doi:10.1139/apnm-2024-0347
26. De la Torre-Luque A, Pemau A, Ayad-Ahmed W, et al. Risk of suicide attempt repetition after an index attempt: a systematic review and meta-analysis. *Gen Hosp Psychiatry*. 2023;81:51–56. doi:10.1016/j.genhosppsych.2023.01.007
27. Song X, Liu X, Zhou Y, et al. Prevalence and correlates of suicide attempts in young patients with first-episode and drug-naïve major depressive disorder: an extensive cross-sectional study. *J Affect Disord*. 2023;340:340–346. doi:10.1016/j.jad.2023.08.006
28. Lin HL, Lee WY, Liu CH, et al. The clinical characteristics and manifestation of anxious depression among patients with major depressive disorders- results from a Taiwan Multicenter Study. *Psychiatry Invest*. 2024;21(6):561–572. doi:10.30773/pi.2023.0419
29. Ye G, Li Z, Yue Y, et al. Suicide attempt rate and the risk factors in young, first-episode and drug-naïve Chinese Han patients with major depressive disorder. *BMC Psychiatry*. 2022;22(1):612. doi:10.1186/s12888-022-04254-x
30. Jiang Y, Lu Y, Cai Y, et al. Prevalence of suicide attempts and correlates among first-episode and untreated major depressive disorder patients with comorbid dyslipidemia of different ages of onset in a Chinese Han population: a large cross-sectional study. *BMC Psychiatry*. 2023;23(1):10. doi:10.1186/s12888-022-04511-z
31. Turkoz I, Fu DJ, Bossie CA, et al. Relationship between the clinical global impression of severity for schizoaffective disorder scale and established mood scales for mania and depression. *J Affect Disord*. 2013;150(1):17–22. doi:10.1016/j.jad.2013.01.047
32. Otte C, Gold SM, Penninx BW, et al. Major depressive disorder. *Nat Rev Dis Primers*. 2016;2(1):16065. doi:10.1038/nrdp.2016.65
33. Baek JH, Heo JY, Fava M, et al. Anxiety symptoms are linked to new-onset suicidal ideation after six months of follow-up in outpatients with major depressive disorder. *J Affect Disord*. 2015;187:183–187. doi:10.1016/j.jad.2015.08.006
34. Zhou Y, Ren W, Sun Q, et al. The association of clinical correlates, metabolic parameters, and thyroid hormones with suicide attempts in first-episode and drug-naïve patients with major depressive disorder comorbid with anxiety: a large-scale cross-sectional study. *Transl Psychiatry*. 2021;11(1):97. doi:10.1038/s41398-021-01234-9
35. Siever L, Davis K. Overview: toward a dysregulation hypothesis of depression. *Am J Psychiatry*. 1985;142(9):1017–1031.
36. Dwight MM, Stoudemire A. Effects of depressive disorders on coronary artery disease: a review. *Harv Rev Psychiatry*. 1997;5(3):115–122. doi:10.3109/10673229709000297
37. Ford DE, Mead LA, Chang PP, et al. Depression is a risk factor for coronary artery disease in men: the precursors study. *Arch Intern Med*. 1998;158 (13):1422–1426. doi:10.1001/archinte.158.13.1422
38. Bosworth HB, Bartash RM, Olsen MK, et al. The association of psychosocial factors and depression with hypertension among older adults. *Int J Geriatr Psychiatry*. 2003;18(12):1142–1148. doi:10.1002/gps.1026
39. Meyer CM, Armenian HK, Eaton WW, et al. Incident hypertension associated with depression in the Baltimore Epidemiologic Catchment Area follow-up study. *J Affect Disord*. 2004;83(2–3):127–133. doi:10.1016/j.jad.2004.06.004
40. Chang HA, Chang CC, Chen CL, et al. Major depression is associated with cardiac autonomic dysregulation. *Acta Neuropsychiatr*. 2012;24 (6):318–327. doi:10.1111/j.1601-5215.2011.00647.x
41. Miller AH, Raison CL. The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol*. 2016;16(1):22–34. doi:10.1038/nri.2015.5