

# Predicting Post-Stroke Depression Risk in Elderly Patients Based on Machine Learning: A Retrospective Cohort Study Integrating Neuroimaging and Psychosocial Variables

Liangjin Zhang\*, Ping Yu\*, Liu Yang, Zian Zhong

Department of Neurology, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, Hubei, 441021, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Zian Zhong; Liu Yang, Department of Neurology, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, 136 Jingzhou Street, Xiangcheng District, Xiangyang, Hubei, 441021, People's Republic of China, Email 15570606720@163.com; 17351378351@163.com

**Objective:** To construct a machine learning-based risk prediction model for post-stroke depression (PSD) in elderly stroke patients by integrating neuroimaging and psychosocial variables, and to improve prediction accuracy for individualized prevention.

**Methods:** A retrospective cohort study included 691 elderly ( $\geq 60$  years) stroke patients from Xiangyang Central Hospital (2016–2023). Baseline clinical data, neuroimaging features (eg, lesion volume, CMBs), and psychosocial variables (eg, SSRS, HAMA-14, HAMD-17) were collected. Seven machine learning models were built; performance was evaluated via AUC, accuracy, and DCA. The optimal model was used to develop a nomogram.

**Results:** The logistic regression (LR) model outperformed others, with AUC=0.88 in the test set. Seven independent predictors were identified: NIHSS score, lesion volume, CMB number, DWI range, Fazekas grade, SSRS score, and HAMA-14 score. The LR-derived nomogram showed good calibration and discriminated PSD from non-PSD effectively.

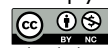
**Conclusion:** The LR model with 7 predictors is accurate and stable for elderly PSD prediction. Its nomogram aids early high-risk identification, supporting personalized intervention, though single-center limitations require multi-center validation.

**Keywords:** post-stroke depression, (PSD), elderly patients, machine learning, logistic regression, neuroimaging, psychosocial variables

## Introduction

Stroke is the second leading cause of death and disability globally, with high morbidity and mortality, and its socio-economic burden surpasses that of cardiovascular diseases and cancer.<sup>1,2</sup> Post-stroke depression (PSD) is the most common and severe neuropsychiatric complication of stroke. It is an affective disorder marked by persistent low mood, anhedonia, alongside accompanying symptoms such as sleep disturbances and fatigue that significantly hinder rehabilitation. Globally, approximately 30% of stroke patients develop depressive symptoms (some data indicate 30~50%), bringing cascading harms including worsened neurological deficits, reduced quality of life, and elevated risks of recurrence or death.<sup>3,4</sup> However, PSD is often overlooked due to prioritization of physical recovery, overlapping symptoms, insufficient clinical awareness, and patient concealment, resulting in missed opportunities for early intervention.

PSD arises from multi-dimensional risk factors including pre-stroke factors (genetics, gender, depression history), stroke-related factors (infarct location, neurological deficit severity), and post-stroke factors (social support, disability



level).<sup>5–7</sup> Pathophysiologically, brain structural damage and neurotransmitter imbalance are core triggers, whereas psychological stress from impaired function or social changes further heightens risk.<sup>8,9</sup> Existing PSD prediction models have critical shortcomings, including overreliance on single-dimensional variables without integrating neuroimaging and psychosocial data, limited efficiency as traditional methods fail to handle multi-source data and machine learning models are not optimized for the elderly, and insufficient clinical practicality due to hard-to-obtain indicators or unclear thresholds.

Machine learning excels at disease risk prediction through its robust data mining capabilities, meanwhile neuroimaging captures post-stroke brain structural and functional changes (infarct volume, functional connectivity) and psychosocial variables (social support, coping style) reflect psychological adaptability.<sup>10,11</sup> Their integration covers the physiological-psychological-social risks of PSD. This retrospective cohort study targets elderly stroke patients who are a high-risk group for PSD, integrates neuroimaging indicators and psychosocial variables, alongside applying machine learning algorithms (random forest, support vector machine, logistic regression) to construct a PSD risk early warning model. It aims to clarify the predictive value of multi-dimensional variables, enhance prediction accuracy, provide a basis for individualized prevention and intervention, and ultimately lower PSD incidence while improving patients' rehabilitation outcomes.

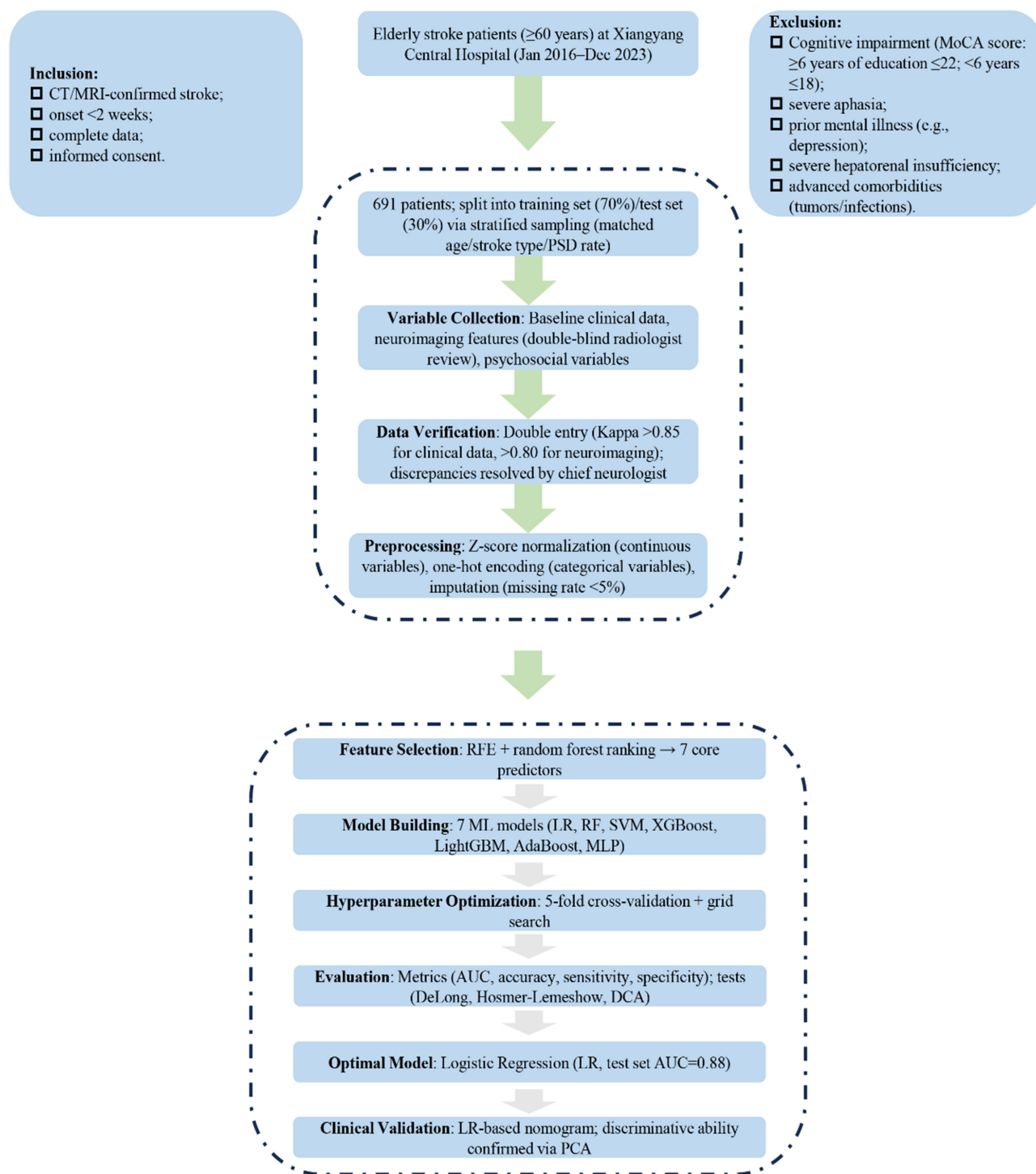
## Materials and Methods

### Study Population

This is a retrospective study focusing on elderly stroke patients admitted to Xiangyang Central Hospital from January 2016 to December 2023, covering an 8-year study period, from which eligible case data were collected using the convenience sampling method. The inclusion criteria included consistency with the standards outlined in the Chinese Guidelines for the Diagnosis and Treatment of Acute Ischemic Stroke and Chinese Guidelines for the Diagnosis and Treatment of Intracerebral Hemorrhage (confirmed by cranial CT and/or MRI), age  $\geq 60$  years (in line with the WHO definition of elderly for developing countries including China), stroke onset within  $< 2$  weeks (presenting in the acute phase), and complete case data with informed consent signed by the patient or their guardian. The exclusion criteria included cognitive impairment (defined as a Montreal Cognitive Assessment [MoCA] score  $\leq 22$  points for patients with  $\geq 6$  years of education, or  $\leq 18$  points for those with  $< 6$  years of education [primary school or below]), severe aphasia precluding communication and assessment, a history of mental illnesses such as depression with confirmed diagnosis or ongoing treatment, severe hepatic and renal insufficiency (including Child-Pugh Class C liver function and creatinine clearance rate  $< 30$  mL/min), and comorbid underlying diseases affecting prognosis (such as advanced tumors and severe infections).<sup>12,13</sup> To determine the sample size, we referred to the EPV principle of logistic regression (EPV 10–20), and combined with literature and hospital data, the preset detection rate of post-stroke depression in the elderly was 38%. Since the study included 12 research factors, 120 cases of depression were required based on an EPV of 10, corresponding to a total sample size of approximately 316 cases; considering the availability of cases over 8 years and the follow-up loss rate, 691 cases were ultimately included to meet statistical requirements. The study was approved by the Hospital Ethics Committee and strictly followed the ethical guidelines of the Helsinki Declaration; additionally, all extraction and analysis of case data adhered to relevant medical ethics and privacy protection regulations, with the patient inclusion process detailed in [Figure 1](#).

### Entry and Collation of Variables

The predictor variables of this study included baseline clinical data, neuroimaging features, and psychosocial variables, which were extracted from electronic medical records, image archives, and follow-up databases. Two trained researchers independently collected the data, which was then verified through double entry by two separate individuals; discrepancies were resolved by a chief neurologist (with a data consistency Kappa value  $> 0.85$ ). Baseline data encompassed demographic characteristics such as gender and age, underlying diseases including hypertension (either diagnosed or with blood pressure  $\geq 140/90$  mmHg) and diabetes mellitus (either diagnosed or with fasting blood glucose  $\geq 7.0$  mmol/L), core indicators like stroke type and admission NIHSS score, as well as initial laboratory results upon admission.



**Figure 1** Flowchart for Patient Enrollment and Predictive Model Development.

Neuroimaging data were based on CT/MRI scans performed within 48 hours of admission, with double-blind interpretation by two radiologists (Kappa value > 0.80); indicators such as lesion location and volume (calculated using the ABC/2 method and Tada formula) and the range of DWI hyperintensity were extracted and quantitatively measured via ImageJ 1.8.0.

Psychosocial variables were assessed by combining nursing records and follow-up evaluations conducted 3 months after treatment: the Social Support Rating Scale (SSRS) was used to classify social support levels (low: ≤ 22 points,

moderate: 23–44 points, high:  $\geq 45$  points), the 17-item Hamilton Depression Rating Scale (HAMD-17) (scores  $\geq 7$  indicating depression) and 14-item Hamilton Anxiety Rating Scale (HAMA-14) (scores  $\geq 14$  indicating anxiety) were employed to evaluate psychological status, and the Barthel Index (BI, 0–100 points) was used to assess daily living ability. Data with a missing rate  $< 5\%$  were supplemented through telephone follow-up. Databases were established using EpiData 3.1 with logical validation rules (eg, ages  $> 120$  years and NIHSS scores  $> 42$  points were identified as outliers).

## Construction and Evaluation of a Machine Learning Predictive Model

In this study, post-stroke depression in the elderly was defined as the outcome variable, with the diagnostic criterion being a score  $\geq 7$  on the 17-item Hamilton Depression Rating Scale (HAMD-17). The input features were strictly linked to the three categories of core indicators collected earlier to ensure data consistency and reliability, specifically including: baseline clinical data (age, gender, history of underlying diseases such as hypertension, type 2 diabetes mellitus, and coronary heart disease, National Institutes of Health Stroke Scale [NIHSS] score, initial fasting blood glucose and low-density lipoprotein test results upon admission), neuroimaging features (lesion location and volume extracted from cranial CT/MRI images within 48 hours of admission, Fazekas grading for leukoaraiosis, range of DWI hyperintensity involvement, and number of cerebral microbleeds), and psychosocial variables (Social Support Rating Scale [SSRS] score, Barthel Index [BI], and HAMA-14 score). Modeling was performed using Python 3.9 programming language, relying on toolkits such as Scikit-learn (including logistic regression, random forest, etc.), XGBoost (extreme gradient boosting), and LightGBM (light gradient boosting machine) to systematically construct 7 machine learning prediction models. The complete list of models included: logistic regression, random forest, support vector machine, extreme gradient boosting, light gradient boosting machine, adaptive boosting, and multilayer perceptron.

Data were divided into a training set (for model construction and parameter optimization) and a test set (for independent performance verification) at a ratio of 7:3 using stratified sampling. This approach ensured consistent age distribution, stroke type, and depression incidence between the two groups to reduce sampling bias. Standardized preprocessing was implemented before model training: continuous features (such as age and scale scores) were normalized using Z-score to eliminate the influence of dimensionality, while categorical features (such as gender and stroke type) were converted to numerical format using one-hot encoding. To reduce feature redundancy and improve model generalization ability, a subset of core features with the highest contribution to the depression outcome was selected through recursive feature elimination (RFE) combined with random forest feature importance ranking. Model performance was primarily evaluated using the area under the receiver operating characteristic curve (AUC), with comprehensive consideration of accuracy, sensitivity, specificity, F1-score, and Youden index. Hyperparameters (such as the number of decision trees and maximum depth for random forest, learning rate and number of tree nodes for extreme gradient boosting) were optimized using 5-fold cross-validation combined with grid search. Statistical differences in AUC among different models were compared using the DeLong test, with Bonferroni correction for multiple comparisons. Calibration curves and the Hosmer-Lemeshow test were used to assess the consistency between predicted probabilities and actual incidence of the models. Finally, decision curve analysis (DCA) was combined to select the prediction model with the optimal clinical net benefit.

## Statistical Analysis

In this study, SPSS 26.0 was used to perform basic data statistics and difference tests, while Python 3.9 was employed to complete machine learning model construction, feature selection, and performance verification, with two-tailed tests applied for all analyses. Continuous data were tested for distribution using the Shapiro–Wilk test, and categorical data were described as  $n$  (%). After grouping by HAMD-17 scores, intergroup comparisons were conducted using  $t$ -tests, Mann–Whitney  $U$ -tests, or other appropriate methods based on data characteristics, and correlation analyses were performed using Spearman's rank correlation or contingency table analysis. Data were managed with EpiData 3.1: missing values with a rate  $< 5\%$  were imputed using multiple imputation, and those with a rate  $\geq 5\%$  were noted. A  $P$ -value  $< 0.05$  was considered statistically significant.

## Results

### Baseline Characteristics of the Study Population

A total of 691 elderly stroke patients were enrolled in this study, including 483 in the training set and 208 in the testing set. As shown in [Table 1](#), the median age of the overall population was 70.00 years (IQR: 66.00, 75.00), with nearly equal gender distribution in the training set (male 50.3%, female 49.7%) and a higher proportion of males in the testing set (60.1%). No significant differences were observed between the depressive and non-depressive groups in terms of age, gender, hypertension, type 2 diabetes mellitus, coronary heart disease, smoking history, drinking history, and other general characteristics (all  $P > 0.05$ ). However, key clinical, neuroimaging, and psychosocial variables showed significant disparities: in both cohorts, the depressive group had higher NIHSS scores, larger lesion volumes, more cerebral microbleeds (CMBs), wider DWI involvement ranges, and higher Fazekas grades (all  $P < 0.001$ ). Additionally, the depressive group exhibited lower SSRS scores, lower BI scores, and higher HAMA-14 scores (all  $P < 0.001$ ). In the testing set, the proportion of hemorrhagic stroke in the depressive group was significantly higher (48.4% vs 27.1%,  $P = 0.03$ ). These results indicate that neurological deficit severity, brain structural damage, social support, and anxiety status may be closely associated with PSD in the elderly.

### Selection of Predictive Variables for the PSD Risk Model

Univariate and multivariate logistic regression analyses were performed to identify potential predictors of PSD. As shown in [Table 2](#), univariate analysis revealed that NIHSS score, lesion volume, CMBs number, DWI range, Fazekas grade (Grade 3–4 vs 1–2), SSRS score, and HAMA-14 score were significantly correlated with PSD (all  $P < 0.001$ ). After adjusting for confounding factors, multivariate logistic regression confirmed these seven variables as independent predictors of PSD (all  $P < 0.001$ ). Further variable screening was conducted using correlation analysis, LASSO regression, and Shapley value analysis ([Figure 2](#)). The correlation heatmap ([Figure 2A](#)) showed no severe multicollinearity among the candidate variables. The LASSO regression coefficient path plot ([Figure 2B](#)) and error curve ([Figure 2C](#)) helped eliminate redundant variables and determine the optimal lambda value. Shapley value-based weight analysis ([Figure 2D](#)) quantified the contribution of each variable, with HAMA-14 score, lesion volume, and CMBs number ranking among the top contributors. These results collectively confirmed that the seven independent variables identified by logistic regression are the core predictive factors for constructing the PSD risk model, covering clinical, neuroimaging, and psychosocial dimensions.

### Comparative Analysis of Predictive Performance Among Machine Learning Models

Seven machine learning models were constructed to predict PSD risk, and their performance was evaluated using multiple metrics in both training and testing sets ([Table 3](#)). Among all models, the logistic regression (LR) model achieved the best overall performance: in the training set, it had an accuracy of 0.90, a sensitivity of 0.89, a specificity of 0.87, and an area under the AUC of 0.90; in the testing set, these metrics were 0.89, 0.88, 0.85, and 0.88, respectively. Other models, including gradient boosting decision tree (GBDT), random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM), LightGBM, and multilayer perceptron (MLP), showed lower performance, with AUC values ranging from 0.72 to 0.81 in the training set and 0.70 to 0.79 in the testing set. The ROC curves ([Figure 3C and D](#)) clearly demonstrated that the LR model had the largest AUC in both datasets. As shown in [Figure 3A and B](#), the DCA indicated that the LR model provided a higher net clinical benefit than other models across a wide range of threshold probabilities. These findings suggest that the LR model has superior predictive accuracy, stability, and generalization ability compared to other machine learning models in predicting elderly PSD risk.

### Clinical Application Evaluation of the Optimal Predictive Model

Based on the LR model, a nomogram was established to visualize the PSD risk prediction ([Figure 4A](#)). The nomogram integrated the seven core predictive variables, allowing clinicians to calculate individual PSD risk by summing the points corresponding to each variable. Calibration curves for the training set ([Figure 4B](#)) and testing set ([Figure 4C](#)) showed good consistency between the predicted probabilities of the nomogram and the actual PSD incidence, indicating high

**Table 1** Baseline Characteristics of Variables in Depressive and Non-Depressive Patients in Training and Testing Cohorts

Variables	Training Cohort			P-value	Testing Cohort			P-value
	Overall (n=483)	Depression (n=75)	No_Depression (n=408)		Overall (n=208)	Depression (n=31)	No_Depression (n=177)	
Age (median [IQR])	70.00 [66.00, 75.00]	69.00 [66.00, 72.00]	70.00 [65.75, 75.00]	0.093	70.00 [66.00, 75.00]	70.00 [63.50, 74.50]	70.00 [66.00, 75.00]	0.651
Gender (%)								
Female	240 (49.7)	41 (54.7)	199 (48.8)	0.417	83 (39.9)	12 (38.7)	71 (40.1)	1
Male	243 (50.3)	34 (45.3)	209 (51.2)		125 (60.1)	19 (61.3)	106 (59.9)	
Hypertension (%)								
Yes	310 (64.2)	52 (69.3)	258 (63.2)		120 (57.7)	18 (58.1)	102 (57.6)	
No	173 (35.8)	23 (30.7)	150 (36.8)	0.378	88 (42.3)	13 (41.9)	75 (42.4)	1
T2DM (%)								
Yes	188 (38.9)	32 (42.7)	156 (38.2)		73 (35.1)	9 (29.0)	64 (36.2)	
No	295 (61.1)	43 (57.3)	252 (61.8)	0.552	135 (64.9)	22 (71.0)	113 (63.8)	0.574
CHD (%)								
Yes	104 (21.5)	20 (26.7)	84 (20.6)		36 (17.3)	7 (22.6)	29 (16.4)	
No	379 (78.5)	55 (73.3)	324 (79.4)	0.306	172 (82.7)	24 (77.4)	148 (83.6)	0.559
Stroke type (%)								
Hemorrhagic	134 (27.7)	19 (25.3)	115 (28.2)	0.714	63 (30.3)	15 (48.4)	48 (27.1)	0.03
Ischemic	349 (72.3)	56 (74.7)	293 (71.8)		145 (69.7)	16 (51.6)	129 (72.9)	
Smoking_history (%)								
Yes	124 (25.7)	13 (17.3)	111 (27.2)		62 (29.8)	8 (25.8)	54 (30.5)	
No	359 (74.3)	62 (82.7)	297 (72.8)	0.098	146 (70.2)	23 (74.2)	123 (69.5)	0.753
Drinking_history (%)								
Yes	82 (17.0)	16 (21.3)	66 (16.2)		22 (10.6)	3 (9.7)	19 (10.7)	
No	401 (83.0)	59 (78.7)	342 (83.8)	0.354	186 (89.4)	28 (90.3)	158 (89.3)	1
SBP (median [IQR])	143.57 [132.88, 154.04]	145.20 [135.86, 154.74]	143.28 [132.74, 153.61]	0.549	142.99 [132.03, 151.57]	145.81 [141.04, 153.44]	142.27 [131.36, 151.12]	0.079
DBP (median [IQR])	83.88 [77.15, 90.75]	82.92 [76.87, 91.55]	84.03 [77.21, 90.47]	0.713	85.25 [78.06, 90.63]	82.05 [77.10, 90.48]	85.41 [78.23, 90.71]	0.374
NIHSS (median [IQR])	8.26 [6.70, 9.89]	10.56 [8.55, 11.84]	7.95 [6.49, 9.56]	<0.001	8.10 [6.43, 9.66]	10.23 [8.48, 12.62]	7.87 [6.33, 9.32]	<0.001
Lesion location (%)								
Basal Ganglia	151 (31.3)	25 (33.3)	126 (30.9)	0.958	65 (31.2)	12 (38.7)	53 (29.9)	0.651
Frontal	86 (17.8)	14 (18.7)	72 (17.6)		48 (23.1)	8 (25.8)	40 (22.6)	
Others	54 (11.2)	8 (10.7)	46 (11.3)		15 (7.2)	2 (6.5)	13 (7.3)	
Temporal	192 (39.8)	28 (37.3)	164 (40.2)		80 (38.5)	9 (29.0)	71 (40.1)	
Sulcus width (median [IQR])	2.98 [2.61, 3.38]	3.01 [2.62, 3.38]	2.98 [2.61, 3.39]	0.886	3.08 [2.67, 3.44]	3.12 [2.67, 3.52]	3.06 [2.67, 3.40]	0.704
Ventricle width (median [IQR])	12.18 [11.30, 13.00]	12.14 [11.23, 13.07]	12.18 [11.31, 12.99]	0.955	12.10 [11.27, 12.89]	12.07 [11.54, 13.21]	12.10 [11.20, 12.82]	0.706

Lesion volume (median [IQR])	5.18 [4.11, 6.44]	7.85 [6.78, 9.56]	4.86 [3.93, 5.88]	<0.001	5.10 [3.95, 6.28]	7.76 [6.34, 9.20]	4.72 [3.86, 5.84]	<0.001
CMBs num (median [IQR])	1.34 [0.76, 2.04]	3.11 [2.38, 4.10]	1.12 [0.63, 1.69]	<0.001	1.40 [0.83, 2.10]	3.69 [2.85, 4.42]	1.29 [0.67, 1.65]	<0.001
DWI range (median [IQR])	2.46 [1.92, 3.01]	4.43 [3.67, 5.62]	2.33 [1.85, 2.78]	<0.001	2.42 [1.91, 2.96]	5.20 [4.57, 5.61]	2.22 [1.85, 2.71]	<0.001
Fazekas grade (%)								
Grade 1	210 (43.5)	4 (5.3)	206 (50.5)	<0.001	87 (41.8)	2 (6.5)	85 (48.0)	<0.001
Grade 2	229 (47.4)	32 (42.7)	197 (48.3)		104 (50.0)	13 (41.9)	91 (51.4)	
Grade 3	42 (8.7)	37 (49.3)	5 (1.2)		15 (7.2)	14 (45.2)	1 (0.6)	
Grade 4	2 (0.4)	2 (2.7)	0 (0.0)		2 (1.0)	2 (6.5)	0 (0.0)	
FBG (median [IQR])	5.45 [4.91, 5.93]	5.36 [4.90, 5.88]	5.46 [4.92, 5.94]	0.707	5.44 [5.00, 5.89]	5.45 [5.02, 5.70]	5.44 [5.01, 5.93]	0.662
LDL (median [IQR])	2.42 [2.13, 2.68]	2.25 [2.00, 2.60]	2.43 [2.17, 2.70]	0.002	2.40 [2.19, 2.67]	2.45 [2.30, 2.73]	2.38 [2.17, 2.66]	0.212
TC (median [IQR])	4.56 [4.15, 4.84]	4.47 [4.14, 4.77]	4.57 [4.17, 4.84]	0.203	4.54 [4.22, 4.88]	4.71 [4.23, 5.02]	4.53 [4.22, 4.86]	0.279
TG (median [IQR])	1.29 [1.10, 1.48]	1.24 [1.13, 1.44]	1.30 [1.10, 1.49]	0.248	1.27 [1.08, 1.48]	1.26 [1.08, 1.44]	1.28 [1.08, 1.48]	0.905
Hb (median [IQR])	128.63 [119.47, 136.69]	126.83 [118.87, 133.20]	128.75 [119.60, 137.67]	0.068	130.04 [122.05, 139.04]	130.16 [124.15, 139.63]	129.95 [121.82, 138.67]	0.686
SSRS score (median [IQR])	40.47 [35.08, 45.62]	30.96 [27.14, 35.42]	42.03 [36.97, 46.69]	<0.001	40.75 [36.18, 46.75]	31.72 [29.06, 35.17]	42.39 [38.15, 47.99]	<0.001
BI index (median [IQR])	77.41 [69.35, 86.60]	63.46 [56.05, 69.62]	79.70 [73.22, 88.16]	<0.001	77.80 [67.99, 86.21]	65.72 [54.72, 70.16]	79.22 [71.57, 87.46]	<0.001
HAMA14 score (median [IQR])	11.28 [8.55, 14.14]	15.76 [14.31, 18.52]	10.58 [7.93, 12.82]	<0.001	11.59 [8.70, 14.22]	18.61 [16.62, 20.94]	10.97 [8.46, 12.87]	<0.001

**Abbreviations:** IQR, Interquartile Range; T2DM, Type 2 Diabetes Mellitus; CHD, Coronary Heart Disease; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; NIHSS, National Institutes of Health Stroke Scale; CMBs, Cerebral Microbleeds; DWI, Diffusion-Weighted Imaging; FBG, Fasting Blood Glucose; LDL, Low-Density Lipoprotein; TC, Total Cholesterol; TG, Triglycerides; Hb, Hemoglobin; SSRS, Social Support Rating Scale; BI, Barthel Index; HAMA14, Hamilton Anxiety Rating Scale-14 Items.

**Table 2** Univariate and Multivariate Logistic Regression Analyses for Depression

Variables	Univariate Logistic Regression	P-value	Multivariate Logistic Regression	P-value
	OR (95% CI)		OR (95% CI)	
Age	0.97 (0.94–1.01)	0.093	–	–
Gender (Female vs Male)	1.24 (0.76–2.02)	0.417	–	–
Hypertension (Yes vs No)	1.31 (0.79–2.17)	0.378	–	–
T2DM (Yes vs No)	1.21 (0.73–1.99)	0.552	–	–
CHD (Yes vs No)	1.38 (0.82–2.32)	0.306	–	–
Smoking history (Yes vs No)	0.61 (0.32–1.16)	0.098	–	–
LDL (mmol/L)	0.85 (0.71–1.02)	0.002	0.91 (0.76–1.09)	0.315
Stroke type (Hemorrhagic vs Ischemic)	0.88 (0.52–1.49)	0.03	0.95 (0.55–1.64)	0.857
NIHSS score	1.32 (1.21–1.45)	<0.001	1.28 (1.15–1.43)	<0.001
Lesion volume (mL)	2.86 (2.31–3.55)	<0.001	2.64 (2.05–3.41)	<0.001
CMBs num	3.51 (2.78–4.43)	<0.001	3.27 (2.51–4.26)	<0.001
DWI range	4.12 (3.29–5.16)	<0.001	3.89 (3.05–4.97)	<0.001
Fazekas grade (Grade 3–4 vs 1–2)	11.75 (7.23–19.08)	<0.001	10.82 (6.45–18.15)	<0.001
SSRS score	0.82 (0.78–0.86)	<0.001	0.79 (0.74–0.84)	<0.001
HAMA14 score	1.47 (1.35–1.60)	<0.001	1.41 (1.28–1.55)	<0.001

**Notes:** All variables are strictly derived from Table 1; non-informative variables with no clinical relevance to depression (eg, TC, TG, DBP, ventricle width, Hb) are excluded.

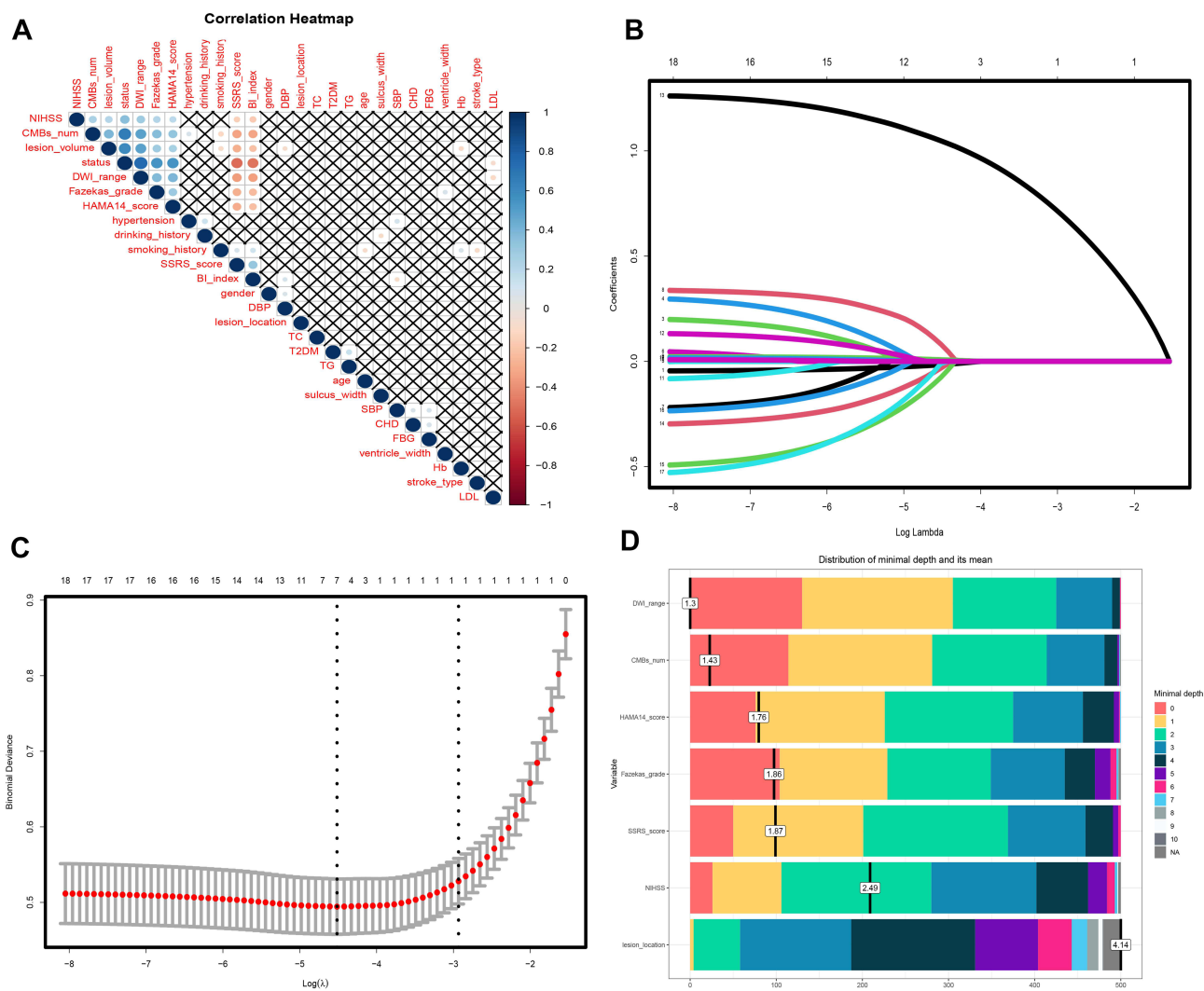
**Abbreviations:** T2DM, Type 2 Diabetes Mellitus; CHD, Coronary Heart Disease; LDL, Low-Density Lipoprotein; NIHSS, National Institutes of Health Stroke Scale; CMBs, Cerebral Microbleeds; DWI, Diffusion-Weighted Imaging; SSRS, Social Support Rating Scale; HAMA14, Hamilton Anxiety Rating Scale-14 Items.

calibration performance. Principal component analysis (PCA) further validated the model's discriminative ability: as shown in [Supplementary Figure 1](#), the PSD and non-PSD groups were clearly separated in both the training and testing set, with the first principal component explaining 35.9% and 38.0% of the total variance, respectively. These results confirm that the nomogram derived from the LR model has excellent clinical applicability, as it can accurately distinguish high-risk elderly stroke patients prone to PSD and provide intuitive risk assessment tools for clinical practice.

## Discussion

On a global scale, PSD afflicts approximately 30% to 50% of stroke patients, profoundly impeding neurological rehabilitation, diminishing quality of life, and elevating the risk of stroke recurrence and mortality.<sup>14–16</sup> Yet it is often overlooked in clinical practice due to prioritization of physical recovery, overlapping symptoms, insufficient clinical awareness, and patient concealment.<sup>17,18</sup> Against this backdrop the present retrospective cohort study focused on elderly stroke patients. These patients represent a high-risk group for PSD. The study integrated baseline clinical data, neuroimaging features and psychosocial variables. It aimed to construct machine learning-based PSD risk prediction models. The results showed that the logistic regression (LR) model emerged as the optimal predictor with an area under the AUC of 0.88 in the test set, and seven independent variables were identified as core predictors of PSD. This study addresses critical limitations of existing PSD prediction models such as overreliance on single-dimensional variables, poor adaptability to the elderly, and limited clinical practicality. By developing a clinically applicable nomogram, it provides a reliable tool for early risk stratification of PSD in elderly stroke patients, laying a foundation for individualized prevention and intervention to improve rehabilitation outcomes.

Seven independent predictors were identified. They are NIHSS, lesion volume, number of CMBs, DWI involvement range, Fazekas grade, SSRS and HAMA-14. The identification of these predictors aligns with the multi-dimensional pathogenesis of PSD. It is also supported by both pathophysiological mechanisms and existing research. The NIHSS score, a direct marker of neurological deficit severity, was confirmed as a robust predictor, which is consistent with findings from studies like Fang et al.<sup>19</sup> Pathophysiologically, severe neurological deficits often involve damage to brain regions critical for emotional regulation, such as the prefrontal cortex, limbic system, and hypothalamic-pituitary-adrenal axis.<sup>20,21</sup> This damage disrupts the synthesis and transmission of neurotransmitters like serotonin and norepinephrine that are essential for mood



**Figure 2** Multimodal Variable Selection for Depression Prediction in Elderly Stroke Patients.

Notes: (A) Correlation heatmap depicting pairwise correlations among clinical, imaging, and psychological variables relevant to depression prediction in elderly stroke patients. (B) LASSO regression coefficient path plot, showing the evolution of feature coefficients for depression prediction as log lambda increases. (C) LASSO regression error curve plot, displaying the relationship between log lambda and cross-validation error in identifying depression predictors. (D) Shapley value-based weight analysis plot, quantifying the contribution and importance of each candidate variable to depression prediction in elderly stroke patients.

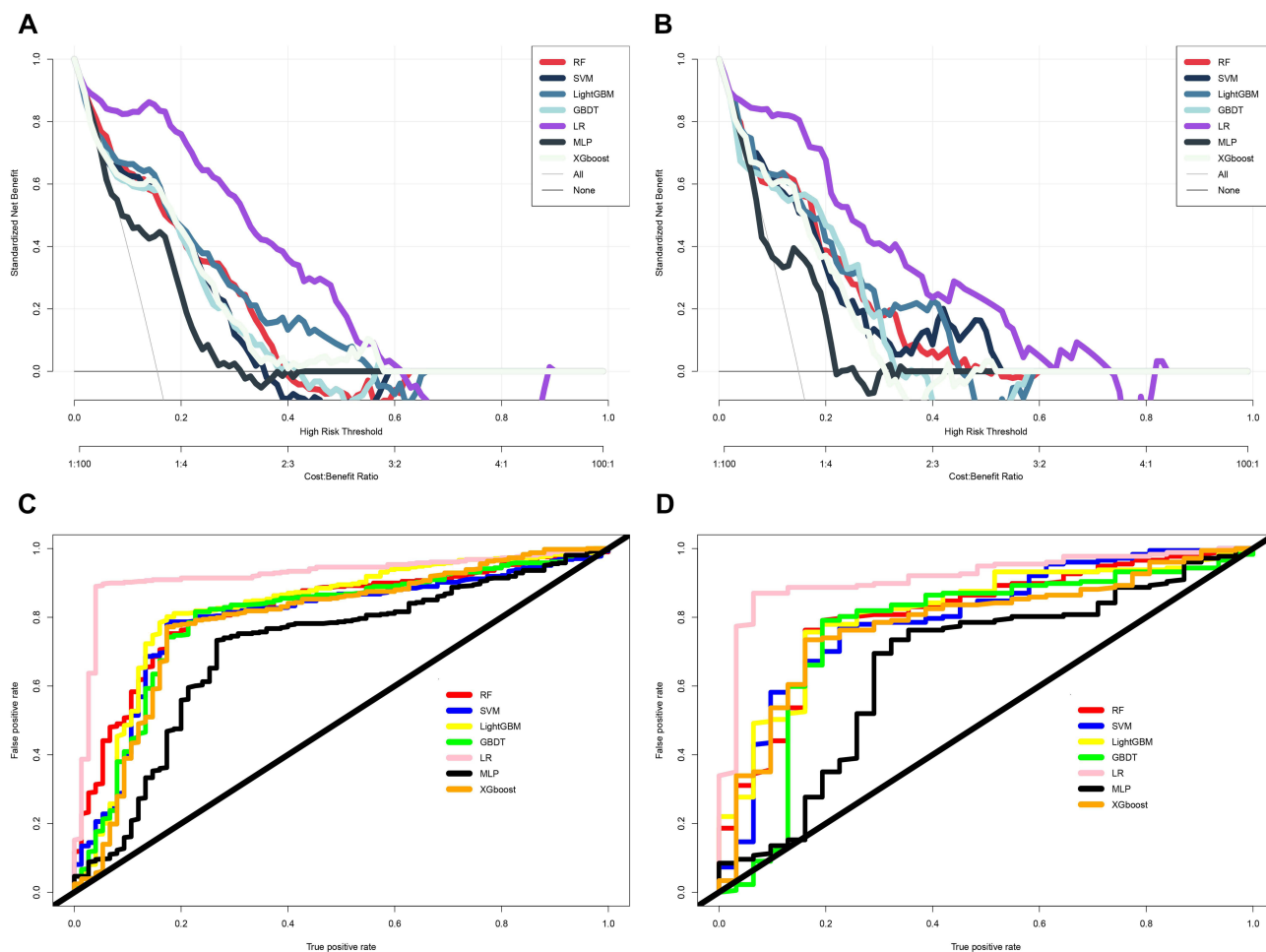
maintenance, thereby elevating depressive risk.<sup>22,23</sup> Lesion volume and CMB number, two key neuroimaging indicators in this study, reflect the extent of brain structural damage and small-vessel disease respectively. Larger lesions can directly destroy monoaminergic neurotransmitter pathways, while CMBs indicate chronic cerebral hypoperfusion and blood-brain barrier dysfunction, both of which exacerbate neuronal vulnerability and further increase PSD risk. This finding extends previous research that primarily focused on lesion location rather than quantitative indicators like volume or CMB count, highlighting the value of precise neuroimaging measurements in capturing subtle structural changes linked to PSD.<sup>24,25</sup> The Fazekas grade assesses the severity of WMLs, which are a key marker of cerebral small-vessel disease (cSVD), and it also emerged as a significant predictor of PSD. White matter lesions can impair functional connectivity between emotional regulation centers and other brain regions, a mechanism that aligns with the findings of Zhang et al<sup>24</sup>, who reported an association between vascular white matter changes and affective disorders in the elderly.

In terms of psychosocial factors, lower SSRS scores (indicating insufficient social support) and higher HAMA-14 scores (indicating significant anxiety) were strongly associated with PSD, which fits the biopsychosocial model of depression. Social support acts as a buffer against psychological stress. Such stress is caused by stroke-related disability

**Table 3** Multi-Dimensional Performance Metric Comparison of Various Machine Learning Models in Training vs Testing Datasets

Dataset	Model	Accuracy	Kappa	Sensitivity	Specificity	PPV	NPV	MCC	J_index	Bal_accuracy	Detection_prevalence	Precision	Recall	F_meas	Roc_auc
Training cohort	LR	0.9	0.55	0.89	0.87	0.97	0.47	0.56	0.76	0.88	0.86	0.97	0.89	0.93	0.9
	GBDT	0.84	0.38	0.83	0.82	0.96	0.32	0.44	0.65	0.82	0.8	0.96	0.83	0.89	0.81
	RF	0.83	0.35	0.82	0.81	0.95	0.29	0.41	0.63	0.81	0.79	0.95	0.82	0.88	0.79
	XGboost	0.82	0.34	0.81	0.8	0.95	0.28	0.4	0.62	0.8	0.78	0.95	0.81	0.87	0.78
	SVM	0.84	0.36	0.83	0.82	0.96	0.3	0.42	0.64	0.82	0.8	0.96	0.83	0.89	0.79
	LightGBM	0.85	0.39	0.84	0.83	0.96	0.33	0.45	0.66	0.83	0.81	0.96	0.84	0.9	0.81
Testing cohort	MLP	0.8	0.31	0.79	0.78	0.94	0.25	0.37	0.57	0.78	0.76	0.94	0.79	0.85	0.72
	LR	0.89	0.53	0.88	0.85	0.96	0.44	0.54	0.73	0.86	0.85	0.96	0.88	0.92	0.88
	GBDT	0.82	0.36	0.81	0.8	0.95	0.3	0.42	0.61	0.8	0.78	0.95	0.81	0.87	0.79
	RF	0.79	0.33	0.78	0.77	0.94	0.27	0.39	0.58	0.77	0.75	0.94	0.78	0.84	0.76
	XGboost	0.78	0.32	0.77	0.76	0.94	0.26	0.38	0.57	0.76	0.74	0.94	0.77	0.83	0.75
	SVM	0.8	0.34	0.79	0.78	0.95	0.28	0.4	0.59	0.78	0.76	0.95	0.79	0.85	0.77
Testing cohort	LightGBM	0.81	0.35	0.8	0.79	0.95	0.29	0.41	0.6	0.79	0.77	0.95	0.8	0.86	0.78
	MLP	0.77	0.29	0.76	0.75	0.93	0.24	0.35	0.55	0.75	0.73	0.93	0.76	0.82	0.7

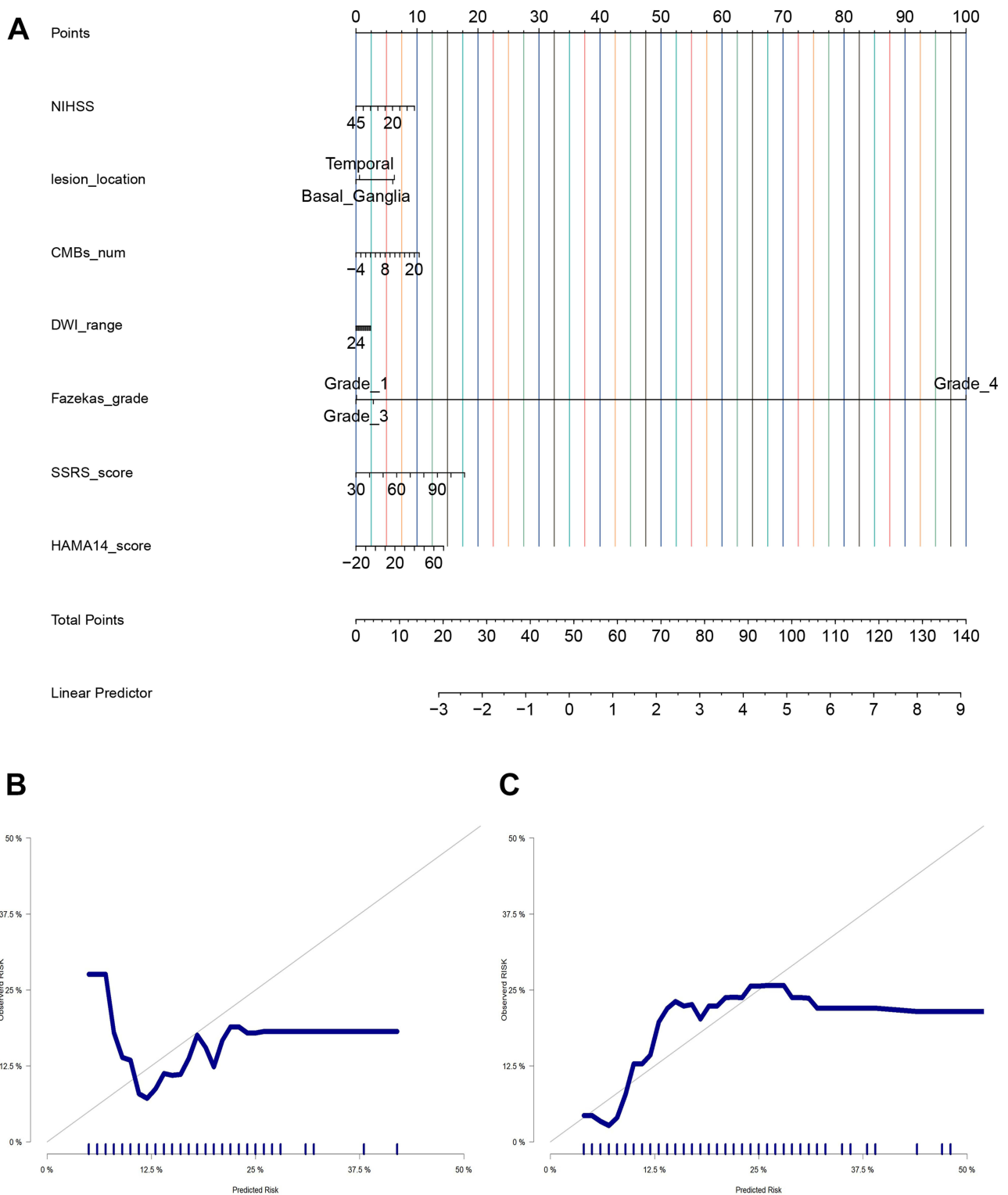
**Abbreviations:** LR, Logistic Regression; GBDT, Gradient Boosting Decision Tree; RF, Random Forest; XGboost, Extreme Gradient Boosting; SVM, Support Vector Machine; LightGBM, Light Gradient Boosting Machine; MLP, Multilayer Perceptron; Kappa, Cohen's Kappa; PPV, Positive Predictive Value; NPV, Negative Predictive Value; MCC, Matthews Correlation Coefficient; J index, Jaccard Index; Bal\_accuracy, Balanced Accuracy; F\_meas, F-measure; Roc\_auc, Receiver Operating Characteristic Area Under the Curve.



**Figure 3** Comparative Predictive Performance Evaluation of Machine Learning Models for Post-stroke depression. (A and B) DCA in Training and Testing dataset; (C and D) ROC curves in Training and Testing dataset.

and role transitions. Insufficient support reduces access to emotional and practical resources. This thereby increases depressive risk. This conclusion has been corroborated by Sun et al.<sup>26</sup> The strong predictive power of HAMA-14 scores underscores the high comorbidity of anxiety and depression after stroke. This comorbidity may stem from shared pathophysiological mechanisms such as neurotransmitter imbalance or overlapping neural circuitry involvement.<sup>27,28</sup> In clinical practice, anxiety often precedes depression in stroke patients because the uncertainty of recovery and functional impairment trigger persistent psychological distress that gradually evolves into depressive symptoms over time. Additionally, the test set revealed a significantly higher proportion of hemorrhagic stroke in the depressive group (48.4% vs 27.1%,  $P=0.03$ ), suggesting that hemorrhagic stroke may pose a greater PSD risk than ischemic stroke in the elderly. This observation is partially supported by Lin et al who reported a higher PSD incidence in hemorrhagic stroke patients.<sup>29</sup> The underlying reason may be that hemorrhagic stroke causes more sudden and severe damage to brain tissue compared to the gradual ischemia in ischemic stroke, leading to more profound disruption of emotional regulation pathways.<sup>30,31</sup> However, this finding requires validation in larger multi-center cohorts to confirm its generalizability.

Notably, the LR model outperformed other machine learning algorithms (including random forest, XGBoost, and support vector machine) in both predictive accuracy and clinical utility, and this superiority can be attributed to several key factors. First, the LR model offers strong interpretability by establishing clear linear relationships between predictors and PSD risk. This allows clinicians to easily understand how each variable contributes to the prediction, a critical advantage over “black-box” ensemble models like XGBoost and LightGBM whose internal decision-making processes are difficult to explain and may thus be less trusted by healthcare providers in clinical practice. Second rigorous feature



**Figure 4** Nomogram Model for Predicting Post-Stroke Depression (PSD) with **(A)** Risk Assessment Nomogram, **(B)** Training Set Calibration Curve, and **(C)** Testing Set Calibration Curve.

selection methods were applied. These methods combine recursive feature elimination with random forest importance ranking. They eliminated redundant variables. They also ensured no severe multicollinearity among the seven core predictors. This optimization is particularly beneficial for linear models like LR, as multicollinearity can distort coefficient estimates and reduce predictive stability. In contrast, complex ensemble models may overfit to noise in the

data, especially given the moderate sample size (691 cases) of this study, leading to reduced generalization ability in the test set. This finding differs from Gu et al who reported that the Gradient Boosting Decision Tree model achieved an AUC of 0.86 for PSD prediction.<sup>32</sup> The discrepancy may stem from differences in study populations (for example, Wang et al included non-elderly patients) or variable sets (they did not integrate both neuroimaging and psychosocial data).<sup>33</sup> Furthermore, the nomogram derived from the LR model demonstrated excellent calibration; calibration curves showed consistent predicted probabilities and actual PSD incidence in both the training and test sets, and principal component analysis confirmed clear separation between PSD and non-PSD groups. These results confirm that the LR model not only has high predictive accuracy but also provides an intuitive tool for bedside risk assessment.

When compared to existing PSD prediction models, the present study offers notable innovations while aligning with key findings in the field. Consistent with previous research, it confirmed the predictive roles of NIHSS score, social support, and anxiety, reinforcing the robustness of these variables across different populations. However, unlike most existing models that rely solely on clinical factors or neuroimaging data, this study integrated three dimensions of variables (clinical, neuroimaging, psychosocial) to cover the physiological-psychological-social spectrum of PSD pathogenesis. This approach addresses the multi-factorial nature of PSD and improves predictive comprehensiveness. Additionally, the inclusion of CMB number and Fazekas grade as key predictors provides new insights into the vascular basis of PSD in the elderly. Previous studies rarely incorporated these quantitative neuroimaging markers, and their inclusion highlights the importance of assessing small-vessel disease and white matter changes in risk stratification for elderly stroke patients who are more prone to such vascular pathologies. The high AUC (0.88) of the LR model in the test set also exceeds that of many existing models (which typically range from 0.70 to 0.85). This superior performance may be attributed to strict inclusion and exclusion criteria. These criteria exclude patients with pre-existing mental illnesses or cognitive impairment. Such exclusion has reduced confounding factors. The performance may also relate to rigorous data collection and verification processes. Trained researchers conduct double entry in these processes. Experts resolve discrepancies when they occur. These measures have ensured data quality. This study has several limitations that warrant consideration. First, it is a single-center retrospective study, which may introduce selection bias and limit the generalizability of results to other healthcare settings or populations. Second, psychosocial variables were assessed 3 months after stroke, and the long-term predictive performance of the model (for example, at 6 or 12 months post-stroke) remains unclear. Third, potential predictors such as genetic factors, gut microbiota composition, and medication use (including antidepressants and antihypertensives) were not included due to data availability, which may have affected the model's comprehensiveness. Fourth, the assessment of psychosocial variables relied on nursing records and follow-up evaluations, which may be subject to subjective bias. Future research should address these limitations by conducting multi-center prospective studies with longer follow-up periods, incorporating additional potential predictors, and validating the model in diverse elderly populations (such as different ethnicities or socioeconomic backgrounds). Furthermore, developing user-friendly tools (like mobile applications or web-based calculators) based on the nomogram could facilitate bedside risk assessment, and interventional studies are needed to verify whether early targeted interventions (such as psychological counseling and social support enhancement) for high-risk patients reduce PSD incidence and improve rehabilitation outcomes.

## Conclusion

Our study constructed and validated a machine learning-based PSD risk prediction model for elderly stroke patients via multi-dimensional variables. The LR model (7 predictors: NIHSS score, lesion volume, CMB number, DWI range, Fazekas grade, SSRS score, HAMA-14 score) achieved high accuracy (AUC=0.88 in test set) with good stability and clinical applicability. Its nomogram aids early identification of high-risk patients. This further supports individualized prevention and intervention. The ultimate goals are to reduce PSD incidence and improve rehabilitation effects. Innovatively integrating quantitative neuroimaging markers (CMB number, Fazekas grade), the study addresses unmet needs for reliable PSD prediction in this group, aligning with existing evidence. Limited by single-center retrospective design, future multi-center prospective studies and interventional trials are needed for validation.

## Disclosure

Liangjin Zhang and Ping Yu contributed equally as co-first authors. Liu Yang and Zian Zhong are co-corresponding authors. The authors report no conflicts of interest in this work.

## References

- Guzik A, Bushnell C. Stroke epidemiology and risk factor management. *Continuum*. 2017;23(1):15–39. doi:10.1212/CON.0000000000000416
- Feigin VL, Brainin M, Norrving B, et al. World stroke organization: global stroke fact sheet 2025. *Int J Stroke*. 2025;20(2):132–144. doi:10.1177/17474930241308142
- Guo J, Wang J, Sun W, Liu X. The advances of post-stroke depression: 2021 update. *J Neurol*. 2022;269(3):1236–1249. doi:10.1007/s00415-021-10597-4
- Cai W, Mueller C, Li YJ, Shen WD, Stewart R. Post stroke depression and risk of stroke recurrence and mortality: a systematic review and meta-analysis. *Ageing Res Rev*. 2019;50:102–109. doi:10.1016/j.arr.2019.01.013
- Getachew B, Mulu A. Prevalence and associated factors of post-stroke depression in Africa: a systematic review and meta-analysis. *BMC Neurol*. 2024;24(1):360. doi:10.1186/s12883-024-03867-x
- Das J, KR G. Post stroke depression: the sequelae of cerebral stroke. *Neurosci Biobehav Rev*. 2018;90:104–114. doi:10.1016/j.neubiorev.2018.04.005
- Shi Y, Yang D, Zeng Y, Wu W. Risk factors for post-stroke depression: a meta-analysis. *Front Aging Neurosci*. 2017;9:218. doi:10.3389/fnagi.2017.00218
- Frank D, Gruenbaum BF, Zlotnik A, Semyonov M, Frenkel A, Boyko M. Pathophysiology and current drug treatments for post-stroke depression: a review. *Int J Mol Sci*. 2022;23(23).
- Villa RF, Ferrari F, Moretti A. Post-stroke depression: mechanisms and pharmacological treatment. *Pharmacol Ther*. 2018;184:131–144. doi:10.1016/j.pharmthera.2017.11.005
- Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Internal Med*. 2018;284(6):603–619. doi:10.1111/joim.12822
- Husile H, Bao Q, Sarula S, et al. Accuracy of machine learning in predicting post-stroke depression: a systematic review and meta-analysis. *Brain and Behav*. 2025;15(5):e70557. doi:10.1002/brb3.70557
- García-Pérez P, Rodríguez-Martínez MC, Gallardo-Tur A, Blanco-Reina E, de la Cruz-Cosme C, Lara JP. Early Occupational Therapy Intervention post-stroke (EOTIPS): a randomized controlled trial. *PLoS One*. 2024;19(8):e0308800. doi:10.1371/journal.pone.0308800
- Lin C, Ren Y, Lu A. The effectiveness of virtual reality games in improving cognition, mobility, and emotion in elderly post-stroke patients: a systematic review and meta-analysis. *Neurosurgical Rev*. 2023;46(1):167. doi:10.1007/s10143-023-02061-w
- Wijeratne T, Sales C, Wijeratne C. A narrative review on the non-pharmacologic interventions in post-stroke depression. *Psychol Res and Behav Manag*. 2022;15:1689–1706. doi:10.2147/PRBM.S310207
- Gallucci L, Sperber C, Guggisberg AG, et al. Post-stroke cognitive impairment remains highly prevalent and disabling despite state-of-the-art stroke treatment. *Int J of Stroke*. 2024;19(8):888–897. doi:10.1177/17474930241238637
- Dec-ówiek M, Wrona P, Homa T, Słowik J, Bodzioch A, Słowik A. Incidence and factors associated with cognitive impairment 90 days after first ever ischemic stroke. *Int J Geriatric Psychiatry*. 2025;40(5):e70087. doi:10.1002/gps.70087
- Butsing N, Zauszniewski JA, Ruksakulpiwat S, Griffin MTQ, Niyomyart A. Association between post-stroke depression and functional outcomes: a systematic review. *PLoS One*. 2024;19(8):e0309158. doi:10.1371/journal.pone.0309158
- Muşat MI, Cătălin B, Hadjiargyrou M, Popa-Wagner A, Greşiță A. Advancing post-stroke depression research: insights from murine models and behavioral analyses. *Life*. 2024;14(9):1110. doi:10.3390/life14091110
- Yang F, Zhang P. Prevalence and predictive factors of post-stroke depression in patients with acute cerebral infarction. *Alpha Psychiatry*. 2024;25(5):592–597. doi:10.5152/alphapsychiatry.2024.231381
- Gu P, Ding Y, Ruchi M, et al. Post-stroke dizziness, depression and anxiety. *Neurological Res*. 2024;46(5):466–478. doi:10.1080/01616412.2024.2328490
- Chang X, He Y, Liu Y, et al. Serum brain derived neurotrophic factor levels and post-stroke depression in ischemic stroke patients. *J Affective Disorders*. 2024;361:341–347. doi:10.1016/j.jad.2024.06.050
- Zhan Q, Kong F. Mechanisms associated with post-stroke depression and pharmacologic therapy. *Front Neurol*. 2023;14:1274709. doi:10.3389/fneur.2023.1274709
- Jing D, Hou X, Guo X, et al. Astrocytes in post-stroke depression: roles in inflammation, neurotransmission, and neurotrophin signaling. *Cell Mol Neurobiol*. 2023;43(7):3301–3313. doi:10.1007/s10571-023-01386-w
- Zhang X, Tang Y, Xie Y, et al. Total magnetic resonance imaging burden of cerebral small-vessel disease is associated with post-stroke depression in patients with acute lacunar stroke. *Eur J Neurol*. 2017;24(2):374–380. doi:10.1111/ene.13213
- Cao WW, Wang Y, Dong Q, et al. Deep microbleeds and periventricular white matter disintegrity are independent predictors of attention/executive dysfunction in non-dementia patients with small vessel disease. *Int. Psychogeriatr*. 2017;29(5):793–803. doi:10.1017/S1041610216002118
- Sun X, Wang Z, Song Y, et al. Development and validation of a decision tree model for prediction of insomnia risk among ischemic stroke convalescence patients. *BMC Public Health*. 2025;25(1):2841. doi:10.1186/s12889-025-24025-z
- Harrison M, Ryan T, Gardiner C, Jones A. Psychological and emotional needs, assessment, and support post-stroke: a multi-perspective qualitative study. *Topic Stroke Rehabil*. 2017;24(2):119–125. doi:10.1080/10749357.2016.1196908
- Kuang J, Yang L, Lv R, Li J, Zhou K, Xu M. The mediating effect of post-stroke depression between social support and quality of life among stroke survivors: a meta-analytic structural equation modeling. *Int J Geriatric Psychiatry*. 2022;37(12). doi:10.1002/gps.5820
- Lin HL, Sung FC, Muo CH, Chen PC. Depression risk in post-stroke aphasia patients: a nationwide population-based cohort study. *Neuroepidemiology*. 2023;57(3):162–169. doi:10.1159/000530070
- Wen QH, Liu Y, Chen HD, et al. Relationship between depression after hemorrhagic stroke and auditory event-related potentials in a chinese patient group. *Neuropsychiatr Dis Treat*. 2022;18:1917–1925. doi:10.2147/NDT.S362824

31. Shi ZM, Jing JJ, Xue ZJ, et al. Stellate ganglion block ameliorated central post-stroke pain with comorbid anxiety and depression through inhibiting HIF-1 $\alpha$ /NLRP3 signaling following thalamic hemorrhagic stroke. *J Neuroinflammation*. 2023;20(1):82. doi:10.1186/s12974-023-02765-2
32. Gu Y, Xue J, Xia X, et al. Prediction of post stroke depression with machine learning: a national multicenter cohort study. *J Psychiatr Res*. 2025;187:123–133. doi:10.1016/j.jpsychires.2025.05.015
33. Wang Y, Pan XF, Pan A. Lipidomics in gestational diabetes mellitus. *Curr Opin Lipidol*. 2023;34(1):1–11. doi:10.1097/MOL.0000000000000858

## Neuropsychiatric Disease and Treatment

**Dovepress**  
Taylor & Francis Group

### Publish your work in this journal

Neuropsychiatric Disease and Treatment is an international, peer-reviewed journal of clinical therapeutics and pharmacology focusing on concise rapid reporting of clinical or pre-clinical studies on a range of neuropsychiatric and neurological disorders. This journal is indexed on PubMed Central, the 'PsycINFO' database and CAS, and is the official journal of The International Neuropsychiatric Association (INA). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/neuropsychiatric-disease-and-treatment-journal>