



## OPEN A clinical machine learning model for 1-year functional outcome prediction in acute ischemic stroke: temporal validation across evolving guidelines

Peijia Liu<sup>1</sup>, Ying Cao<sup>1</sup>, Xingbang Zou<sup>1</sup>, Zhaobo Zhang<sup>1</sup>, Jinan Fang<sup>1</sup>, Peipei Liu<sup>2</sup>✉ & Peifang Liu<sup>1</sup>✉

Accurate prediction of 1-year excellent functional outcome (modified Rankin Scale [mRS] 0–1) in acute ischemic stroke (AIS) patients is vital for guiding long-term rehabilitation. However, existing tools primarily focus on short-term (3-month) outcomes and often lack validation in temporally distinct cohorts, particularly when clinical guidelines and treatment landscapes evolve. To address this, we trained six machine learning models on a derivation cohort ( $n=965$ , admitted 2020–2023) managed under the 2018 Chinese Guidelines for Diagnosis and Treatment of Acute Ischemic Stroke. The optimal logistic regression (LR) model included eight key predictors: admission NIHSS, admission mRS, age, neutrophil-to-lymphocyte ratio (NLR), glucose, blood urea nitrogen (BUN), D-dimer, and B-type natriuretic peptide (BNP). The LR model was rigorously assessed on an independent temporal validation cohort ( $n=144$ , admitted 2024) treated under the 2023 Guidelines, which expanded indications for reperfusion therapy. Although the validation cohort showed significantly higher thrombolysis rates and milder symptoms than the derivation cohort, the LR model demonstrated robust performance (AUC = 0.80, 95% CI: 0.72–0.87), significantly outperforming admission National Institutes of Health Stroke Scale (NIHSS) score (AUC = 0.73, 95% CI: 0.64–0.81). The model also showed substantial incremental value with a net reclassification improvement of 0.71 and an integrated discrimination improvement of 0.14 (both  $P < 0.001$ ). Finally, an open-access web-based predictor was deployed to facilitate clinical implementation within the first 24 h of admission. In summary, we developed and temporally validated a robust, interpretable prediction model for 1-year functional outcome in AIS, offering a practical tool for long-term prognosis.

**Keywords** Acute ischemic stroke, Machine learning, Functional recovery, Temporal validation, Long-term outcome

Acute ischemic stroke (AIS) remains a leading cause of mortality and long-term disability worldwide, imposing a heavy burden on public health systems<sup>1,2</sup>. Despite the widespread adoption of acute recanalization therapies, such as intravenous thrombolysis and endovascular thrombectomy, a significant proportion of survivors suffer from long-term functional dependence. Research indicates that persistent impairment rates at one year can reach 50% in the general population, and remain as high as 25% even among younger cohorts<sup>3,4</sup>. While the 3-month (90-day) functional outcome is the standard endpoint in most clinical trials, the 1-year outcome provides a more comprehensive assessment of sustained recovery and long-term quality of life<sup>5,6</sup>. Therefore, accurate prediction of 1-year functional status is clinically vital for managing patient expectations, optimizing long-term rehabilitation strategies, and ensuring the effective allocation of healthcare resources.

Currently, established prognostic scores such as ASTRAL, DRAGON, and iScore generally demonstrate good discriminative ability (AUC 0.75–0.85) in their respective derivation cohorts<sup>7–10</sup>. However, their applicability

<sup>1</sup>Department of Neurology, Second Affiliated Hospital of Harbin Medical University, Harbin 150088, China.

<sup>2</sup>School of Pharmacy, Anhui Xinhua University, Hefei 230088, China. ✉email: liupeipei@axhu.edu.cn, liupeifang@hrbmu.edu.cn

to modern stroke care is increasingly constrained by several limitations. First, most validated scales focus on 3-month outcomes, leaving a gap in predicting long-term trajectories. Second, they rely primarily on static baseline variables, often overlooking individual biomarkers (e.g., inflammation, lipid profiles) and dynamic in-hospital complications (e.g., pneumonia, hemorrhagic transformation) that significantly influence prognosis. Third, regarding the outcome endpoint, these tools typically define a “good outcome” as functional independence (modified Rankin Scale [mRS] 0–2). This definition aggregates patients with slight disability (mRS 2) together with those who have fully recovered<sup>11</sup>. As acute therapies advance, predicting an “excellent outcome” (mRS 0–1), that indicates a symptom-free return to pre-stroke life, represents a more refined, patient-centered standard that current scores may not adequately capture<sup>12,13</sup>.

To address the complexity of stroke prognosis, machine learning has been extensively explored, offering the potential to model non-linear relationships among high-dimensional clinical data<sup>14,15</sup>. A variety of algorithms, ranging from random forest (RF) to neural networks, have been compared against standard logistic regression (LR)<sup>16–19</sup>. However, distinct gaps remain in the translational value of these studies. Methodologically, the majority rely on random data splitting (e.g., k-fold cross-validation) for internal validation. This approach fails to simulate the real-world clinical scenario where a model trained on historical data is applied to future patients, often resulting in overly optimistic performance estimates. While multi-center external validation is ideal, it is frequently resource-intensive. In its absence, temporal validation, evaluating the model on a strictly subsequent cohort, serves as a far more rigorous alternative to test model stability over time<sup>20</sup>, yet this step is frequently overlooked.

This study aims to develop and rigorously validate a clinical prediction model for 1-year excellent functional outcome (mRS 0–1) in AIS patients. We adopt a strict temporal validation design to ensure the model’s reliability for future application. Using a derivation cohort of 965 patients hospitalized from 2020 to 2023, we compared multiple algorithms, including LR, RF, and multi-layer perceptron (MLP), to determine the optimal modeling strategy. We then validated the model’s performance on an independent, temporal cohort of 144 patients admitted in 2024. Finally, to bridge the gap between research and practice, we deployed the final model as an easy-to-use web-based tool, enabling clinicians to readily access personalized prognostic predictions.

## Methods

### Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. The research protocol was approved by the Ethics Committee of the Second Affiliated Hospital of Harbin Medical University (No. 2021-123-01). The requirement for informed consent was waived by the Ethics Committee due to the retrospective nature of the study.

### Study design and participants

We screened a total of 1,187 patients (all Chinese ethnicity) admitted between November 2020 and October 2024, at the Department of Neurology, Second Affiliated Hospital of Harbin Medical University. The inclusion criteria were as follows: (1) age > 18 years; and (2) a confirmed diagnosis of AIS according to the Chinese Guidelines for Diagnosis and Treatment of Acute Ischemic Stroke 2018 or 2023<sup>21,22</sup>. Patients with non-vascular neurological deficits, malignancies affecting survival, or extensive missing data were excluded. Utilizing a temporal split design to simulate prospective application, we divided patients into a derivation cohort (November 2020–October 2023) and an independent validation cohort (January 2024–October 2024). After excluding 78 patients (approximately 6.6%) lost to follow-up at 1 year, 1,109 patients were included in the final analysis: 965 in the derivation cohort and 144 in the validation cohort (Fig. 1).

### Outcomes

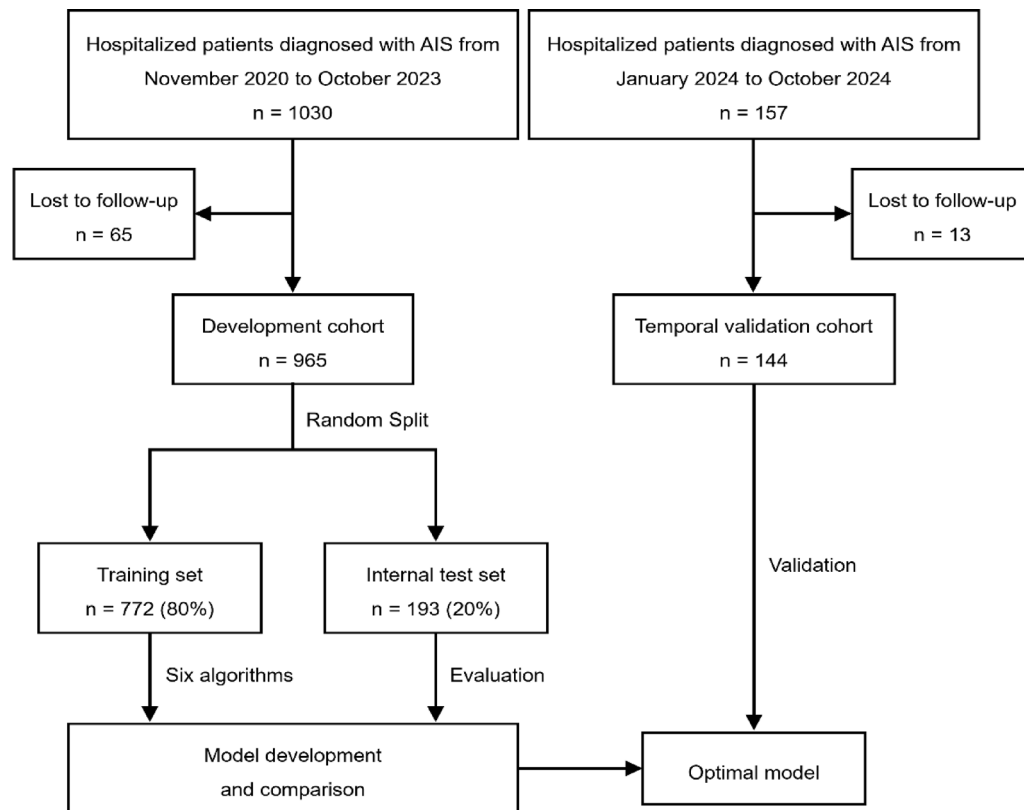
The primary endpoint was the functional outcome at 1 year (12 months) after stroke onset, assessed using mRS. Consistent with the goal of identifying optimal recovery, the outcome was dichotomized into excellent functional outcome (mRS 0–1) versus unfavorable outcome (mRS 2–6). While ordinal models utilize the full distribution of scores (mRS 0–6), a binary probability of excellent recovery is more intuitive for bedside counseling than interpreting shifts across multiple disability grades. Additionally, this approach avoids the strict proportional odds assumption, which is often violated in stroke prognostication. To ensure the reliability of the label, follow-up assessments were conducted via structured telephone interviews by two independent, trained raters blinded to model predictions. Detailed descriptions of the annotation process, rater expertise, inter-rater reliability (Kappa statistics), and conflict resolution strategies are provided in the Supplementary Methods.

### Data collection

Baseline clinical data were extracted from electronic health records by trained neurologists blinded to patient outcomes. A total of 76 candidate variables were collected and categorized into five domains: demographics and medical history, vital signs and clinical scores, laboratory parameters, imaging features, and treatment profiles. Laboratory markers were obtained from the venous blood sample drawn within 24 h of admission. To ensure data integrity and robustness, specific setting-dependent collection strategies (e.g., fasting vs. emergency requirements), plausibility checks, and measures to mitigate coding bias (e.g., upcoding) are detailed in the Supplementary Methods. A comprehensive list of all variables definitions is provided in Supplementary Table S1.

### Data preprocessing

In the derivation cohort, variables with a missing rate > 10% were excluded to minimize bias associated with imputation<sup>23</sup>. Additionally, variables with a single value frequency > 90% were removed as they functioned as near-zero variance predictors with limited discriminative power<sup>24</sup>. The remaining data were randomly split



**Fig. 1.** Flowchart of the study design.

into a training set (80%) for model development and an internal test set (20%) for performance evaluation. To prevent data leakage, all subsequent preprocessing steps were strictly fitted on the training set and applied to the internal test set and temporal validation cohort. Missing values were imputed using the median (for continuous variables) or mode (for categorical variables) derived from the training set. Continuous features were standardized using Z-score normalization based on the mean and standard deviation of the training data.

### Machine learning models development

Following preprocessing, we implemented a multi-stage feature selection strategy on the training set to identify robust predictors. This process integrated three quantitative approaches: univariate analysis, LASSO regression, and RF variable importance. The selected features were further refined by removing highly correlated variables (Spearman's  $r > 0.7$ ) to ensure clinical utility and minimize redundancy. Detailed methodology of the feature selection process is provided in Supplementary Methods.

Based on the selected features, we compared six machine learning algorithms: LR, RF, support vector machine (SVM), extreme gradient boosting (XGBoost), naive bayes, and MLP. To optimize performance and prevent overfitting, we applied grid search with 5-fold cross-validation on the training set to fine-tune the hyperparameters of each algorithm. The models were evaluated on the internal test set using a holistic assessment of the area under the receiver operating characteristic curve (AUC), accuracy, recall, specificity, and F1 score. The algorithm demonstrating the best balance of discrimination and stability was selected as the optimal model. Model development was conducted using the scikit-learn (version 1.7.2) and xgboost (version 3.1.2) libraries in Python.

### Temporal validation

The performance of selected model was further verified using the independent temporal validation cohort. Discrimination was quantified by AUC, and calibration was assessed via calibration plots with Brier score. Incremental prognostic value over the baseline admission NIHSS was quantified by calculating the net reclassification improvement (NRI) and integrated discrimination improvement (IDI). Finally, decision curve analysis (DCA) was performed to evaluate the model's clinical net benefit across varying threshold probabilities.

### Web-based predictor deployment

To facilitate the translation of our research findings into clinical practice, the optimal model was encapsulated into a user-friendly, open-access web application. The web tool was built using the streamlit (version 1.52.1) framework in Python.

## Statistical analysis

Descriptive statistical analyses were performed using R software (version 4.5.1). Continuous variables were expressed as medians with interquartile ranges (IQR), while categorical variables were presented as frequencies and percentages. Differences in baseline characteristics between the derivation and validation cohorts were compared using the Mann-Whitney U test for continuous variables and the Chi-square test for categorical variables. All statistical tests were two-sided, and a  $P$ -value  $< 0.05$  was considered statistically significant.

## Results

### Demographic and clinical characteristics

A total of 1,109 eligible AIS patients were analyzed, comprising a derivation cohort ( $n = 965$ , admitted 2020–2023) and a temporal validation cohort ( $n = 144$ , admitted 2024). The baseline characteristics are detailed in Table 1. While sex distribution and major comorbidities (e.g., hypertension and diabetes) were comparable, significant temporal heterogeneity was observed in disease severity and treatment. The validation cohort presented with milder symptoms, characterized by significantly lower median admission NIHSS scores (3.0 [IQR 2.0–9.0] vs. 4.0 [IQR 2.0–10.0],  $P = 0.009$ ), reduced B-type natriuretic peptide (BNP) levels (77.5 vs. 122.0 pg/mL,  $P = 0.001$ ), and fewer brainstem infarctions (11.1% vs. 21.6%,  $P = 0.005$ ). Notably, intravenous thrombolysis rates surged from 27.6 to 59.7% ( $P < 0.001$ ), likely attributable to the expanded indications under the 2023 Chinese Stroke Guidelines<sup>22</sup>. Despite these disparities, the rate of 1-year excellent functional outcome (mRS 0–1) remained comparable between the derivation (46.2%) and validation cohorts (40.3%,  $P = 0.213$ ). These data underscore the critical necessity of temporal validation to rigorously test the model's robustness in an evolving medical landscape.

Data are presented as median (interquartile range [IQR]) for continuous variables and number (percentage) for categorical variables. Differences between the derivation and temporal validation cohorts were assessed using the Mann-Whitney U test for continuous variables and the Pearson's chi-squared test or Fisher's exact test for categorical variables. Abbreviations: NIHSS, National Institutes of Health Stroke Scale; mRS, modified Rankin

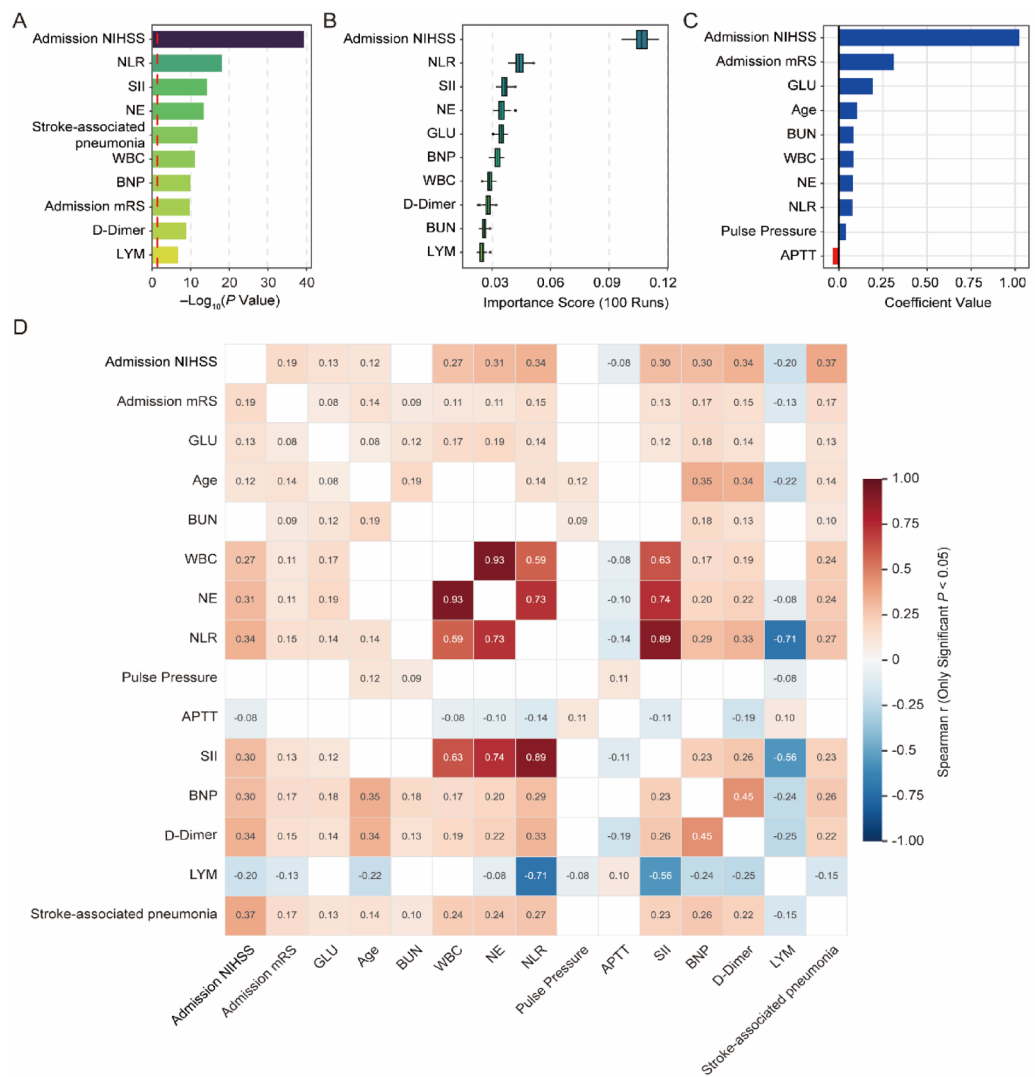
Variable	Overall (N = 1109)	Derivation cohort (N = 965)	Temporal validation cohort (N = 144)	P-value
Demographics				
Age, years	65.0 (57.0, 71.0)	64.0 (57.0, 71.0)	66.0 (61.0, 72.0)	0.015
Sex (Male), n(%)	781 (70.4%)	684 (70.9%)	97 (67.4%)	0.444
Smoking, n(%)	455 (41.0%)	415 (43.0%)	40 (27.8%)	< 0.001
Comorbidities				
Hypertension, n(%)	671 (60.5%)	595 (61.7%)	76 (52.8%)	0.052
Diabetes, n(%)	284 (25.6%)	250 (25.9%)	34 (23.6%)	0.627
Previous stroke/TIA, n(%)	400 (36.1%)	361 (37.4%)	39 (27.1%)	0.021
Clinical assessment				
Admission NIHSS, score	4.0 (2.0, 10.0)	4.0 (2.0, 10.0)	3.0 (2.0, 9.0)	0.009
Admission mRS, score	0.0 (0.0, 1.0)	0.0 (0.0, 1.0)	1.0 (0.0, 2.0)	< 0.001
SBP, mmHg	149.0 (135.0, 164.0)	149.0 (135.0, 164.0)	148.5 (136.0, 165.0)	0.498
Laboratory tests				
Glucose, mmol/L	6.2 (5.2, 8.1)	6.2 (5.3, 8.3)	5.5 (5.0, 6.8)	< 0.001
BUN, mmol/L	6.1 (4.9, 7.6)	6.1 (5.0, 7.6)	5.8 (4.6, 7.2)	0.054
D-dimer, ng/mL	165.0 (90.0, 368.0)	166.0 (89.0, 390.0)	157.5 (98.2, 313.2)	0.839
BNP, pg/mL	115.0 (49.0, 355.0)	122.0 (52.0, 422.0)	77.5 (33.5, 237.8)	0.001
NLR	4.1 (2.5, 6.9)	4.1 (2.5, 6.9)	4.0 (2.7, 6.7)	0.987
LDL-C, mmol/L	2.7 (2.1, 3.3)	2.7 (2.1, 3.3)	2.9 (2.2, 3.5)	0.056
Imaging and etiology				
TOAST classification, n(%)				0.009
LAA	684 (61.7%)	583 (60.4%)	101 (70.1%)	
CE	112 (10.1%)	106 (11.0%)	6 (4.2%)	
SVO	269 (24.3%)	241 (25.0%)	28 (19.4%)	
SOE	29 (2.6%)	22 (2.3%)	7 (4.9%)	
SUE	15 (1.4%)	13 (1.3%)	2 (1.4%)	
Treatment				
Intravenous thrombolysis, n(%)	352 (31.7%)	266 (27.6%)	86 (59.7%)	< 0.001
Antiplatelet therapy, n(%)	753 (67.9%)	630 (65.3%)	123 (85.4%)	< 0.001
Outcome				
1-Year mRS 2–6, n(%)	504 (45.4%)	446 (46.2%)	58 (40.3%)	0.213

**Table 1.** Baseline characteristics of the study population.

Scale; SBP, systolic blood pressure; BUN, blood urea nitrogen; NLR, neutrophil-to-lymphocyte ratio; LDL-C, low-density lipoprotein cholesterol; BNP, B-type natriuretic peptide; TOAST, Trial of ORG 10,172 in Acute Stroke Treatment; LAA, large artery atherosclerosis; CE, cardioembolism; SVO, small vessel occlusion; SOE, stroke of other determined etiology; SUE, stroke of undetermined etiology.

**Feature selection**

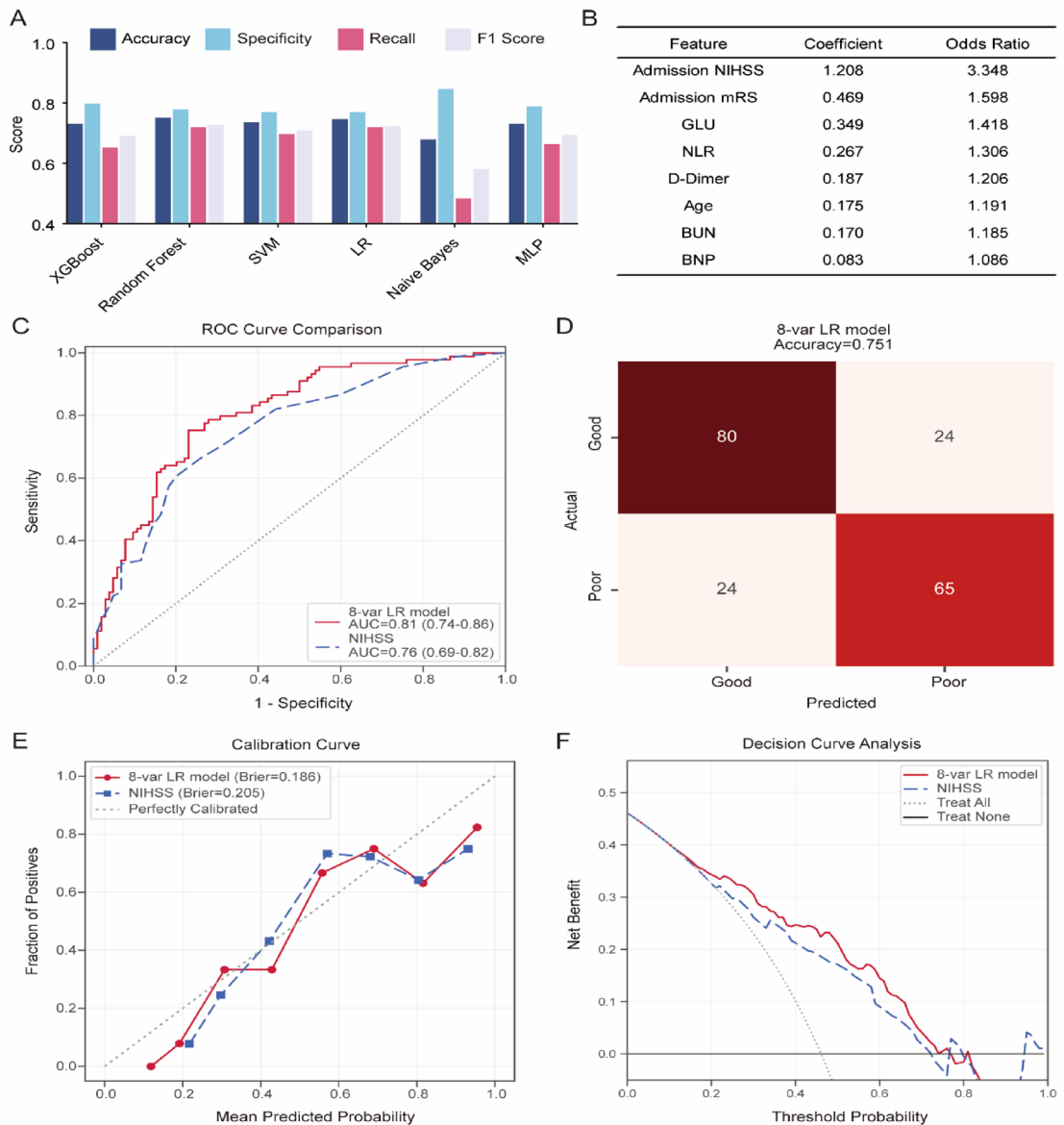
From an initial pool of 76 clinical variables (Supplementary Table S1), 58 remained after excluding features with high missing rates or low variance (Supplementary Table S2). Through our multi-stage screening, admission NIHSS consistently ranked as the top predictor (Fig. 2). Following the aggregation of top 10 candidate features, Spearman’s correlation analysis was conducted to address multicollinearity (Fig. 2D). Notably, neutrophil-to-lymphocyte ratio (NLR) was retained over highly correlated markers such as systemic immune-inflammation index (SII,  $r = 0.89$ ) and neutrophils ( $r = 0.73$ ) due to its superior RF importance (Fig. 2B). While BNP showed lower regularization weight, it was included due to its significant fold change (Supplementary Table S3) and established clinical relevance<sup>25</sup>. Ultimately, eight predictors were selected: admission NIHSS, admission mRS, age, NLR, glucose, blood urea nitrogen (BUN), D-dimer, and BNP. The detailed distributions and distributional shifts of these eight key predictors between the derivation and validation cohorts are visualized in Supplementary Fig. S1.



**Fig. 2.** Multi-stage feature selection and correlation analysis. (A–C) Top 10 features ranked based on (A) univariate analysis (the Mann-Whitney U test for continuous variables and the Chi-square test for categorical variables), (B) random forest variable importance, and (C) LASSO regression. (D) Spearman’s rank correlation heatmap of candidate features. NIHSS, National Institutes of Health Stroke Scale; mRS, modified Rankin Scale; GLU, blood glucose; BUN, blood urea nitrogen; WBC, white blood cell count; NE, neutrophil count; NLR, neutrophil-to-lymphocyte ratio; APTT, activated partial thromboplastin time; SII, systemic immune-inflammation index; BNP, B-type natriuretic peptide; LYM, lymphocyte count.

### Model development and performance comparison

Six machine learning algorithms were trained on the training set ( $n = 772$ ) and subsequently evaluated on the internal test set ( $n = 193$ ) to determine the optimal modeling strategy. The performance metrics revealed that the LR and RF models yielded comparable and superior predictive capabilities (Fig. 3A and Supplementary Table S4). The LR model achieved an accuracy of 0.746, recall of 0.719, specificity of 0.769, and an F1 score of 0.723. These metrics were statistically consistent with the complex ensemble RF model (accuracy = 0.751, recall = 0.719, specificity = 0.779, F1 score = 0.727), suggesting that increasing model complexity did not yield significant performance gains. Given the paramount importance of interpretability in clinical practice, LR was selected as the final model. Figure 3B details the standardized coefficients and odds ratios (OR) of the eight predictors in the LR model. Admission NIHSS emerged as the strongest independent predictor (OR = 3.348), followed by admission mRS (OR = 1.598) and glucose (OR = 1.418). Notably, inflammatory and coagulation markers (NLR, D-dimer, BUN, BNP), along with age, also contributed significantly to the risk profile (ORs ranging from 1.086 to 1.306), providing a multidimensional basis for prediction.



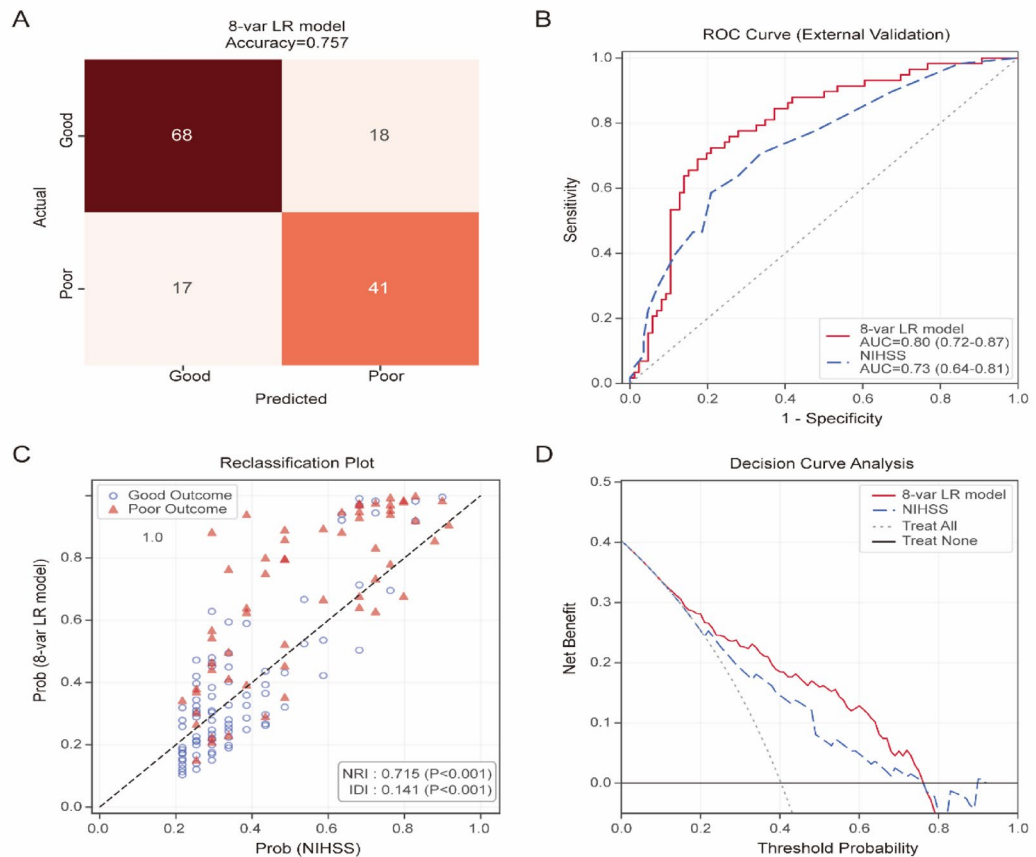
**Fig. 3.** Model development and performance evaluation in the derivation cohort. **(A)** Comparison of performance metrics across six machine learning algorithms in the internal test set. **(B)** coefficients and odds ratios for each predictor in the logistic regression (LR) model. **(C)** Receiver Operating Characteristic (ROC) curves comparing the LR model against the admission NIHSS score alone in the internal test set. **(D)** Confusion matrix of the LR model in the internal test set. **(E)** Calibration curves of the LR model and admission NIHSS score. **(F)** Decision Curve Analysis (DCA) comparing the net benefit of the LR model, admission NIHSS score, and “treat-all”/“treat-none” strategies.

Given that admission NIHSS is a robust independent predictor, we further compared the performance of our 8-variable LR model against the NIHSS score alone on the internal test set. The ROC analysis (Fig. 3C) highlighted the superior discrimination of the LR model, which achieved an AUC of 0.81 (95% confidence interval [CI]: 0.74–0.86) versus 0.76 (95% CI: 0.69–0.82) for the NIHSS alone. This represented a statistically significant improvement of 0.05 ( $P=0.012$ ). The confusion matrix (Fig. 3D) confirmed the model's balanced classification ability, correctly identifying 65 high-risk and 80 low-risk patients. Regarding calibration, the LR model demonstrated excellent agreement between predicted probabilities and observed outcomes (Fig. 3E), yielding a lower Brier score of 0.186 compared to the NIHSS baseline (0.205). Finally, DCA established the clinical utility of the LR model (Fig. 3F), providing a higher net benefit across a wide range of threshold probabilities (approximately 18% to 75%) than the single-variable NIHSS strategy.

### Temporal validation and incremental clinical value

To verify the robustness and generalizability of the model in a prospective-like setting, we evaluated its performance in an independent temporal validation cohort ( $n=144$ ). As illustrated by the confusion matrix in Fig. 4A, the overall classification accuracy of the 8-variable LR model in the validation set was 0.76. Further analysis of misclassified cases revealed that prediction errors were largely attributable to unmeasured post-admission events. Specifically, patients misclassified as having a good prognosis often experienced severe complications (e.g., infections or recurrence) during or after hospitalization, which worsened their 1-year functional status despite favorable baseline characteristics.

The ROC analysis (Fig. 4B) revealed that the LR model achieved an AUC of 0.80 (95% CI: 0.72–0.87), outperforming the admission NIHSS score alone (AUC=0.73, 95% CI: 0.64–0.81). This improvement in discrimination was statistically significant ( $\Delta\text{AUC}=0.07$ ,  $P=0.0125$ ), indicating that the multivariate model remained superior to the single-predictor baseline even in a temporally distinct population. Furthermore, reclassification analysis demonstrated that the LR model provided substantial added value, with an NRI of 0.72 ( $P<0.001$ ) and an IDI of 0.14 ( $P<0.001$ ) compared to the NIHSS score (Fig. 4C). These results indicate that the model significantly improves risk stratification accuracy by correctly reclassifying a notable proportion



**Fig. 4.** Robustness verification in the independent temporal validation cohort. **(A)** Confusion matrix of the LR model applied to the 2024 temporal validation cohort. **(B)** ROC analysis comparing the discrimination of the LR model versus the admission NIHSS score alone in the validation set. **(C)** Reclassification analysis illustrating the incremental clinical value (NRI and IDI) of the LR model over the NIHSS score. **(D)** Decision Curve Analysis (DCA) in the temporal validation cohort, indicating the range of threshold probabilities where the LR model provides superior net benefit.

Language / 语言  
 English  
 中文

## 1-Year Prognosis Prediction for Stroke

Based on 8-Variable Logistic Regression Model

Predicts 1-year mRS outcome for acute ischemic stroke patients.

⚠ For research use only. Not for definitive clinical diagnosis.

### Clinical Indicators

Admission NIHSS Score	5	-	+	Glucose (GLU, mmol/L)	6.00	-	+
Admission mRS (Pre-stroke)	0			Blood Urea Nitrogen (BUN, mmol/L)	5.50	-	+
Age (years)	65	-	+	D-Dimer (ng/mL)	200.00	-	+
Neutrophil-to-Lymphocyte Ratio (NLR)	2.50	-	+	BNP (pg/mL)	100.00	-	+

**Predict**

### Prediction Result

Probability of Poor Outcome: **30.2%**

**Good Outcome (mRS 0-1)**

Low Risk estimated by model.

① **Input data**

② **Run model**

③ **View results**

**Fig. 5.** Deployment of the web-based prognostic tool. Screenshot of the online calculator interface (available at: <https://stroke-1year-mrs-lzhwrswqlce2zxmoxfiky.streamlit.app/>).

of patients misclassified by NIHSS alone. DCA further supported the model's clinical applicability, showing a consistently higher net benefit across a wide range of decision thresholds compared to the NIHSS strategy (Fig. 4D). In summary, despite notable distributional shifts in baseline characteristics observed in the 2024 cohort, the LR model demonstrated stable and generalizable predictive performance.

### Web-based predictor for clinical implementation

To facilitate the translation of the validated 8-variable LR model into routine clinical practice, we developed an open-access, user-friendly web-based calculator (available at: <https://stroke-1year-mrs-lzhwrswqlce2zxmoxfiky.streamlit.app/>). As shown in Fig. 5, the interface is streamlined for bedside usage. Clinicians can input the patient's specific values for the eight key predictors (admission NIHSS, admission mRS, age, NLR, glucose, BUN, D-dimer, and BNP) via adjustable sliders or numeric entry fields. Upon submitting the data, the tool generates an immediate, individualized probability of achieving an excellent functional outcome (mRS 0–1) at 1 year. This digital tool serves as a rapid decision-support aid, enabling physicians to visualize long-term prognosis and optimize rehabilitation strategies and patient counseling early in the acute phase of hospitalization.

### Discussion

This study developed and validated a clinical machine learning model to predict 1-year excellent functional outcomes (mRS 0–1) in AIS patients. Unlike existing tools (e.g., ASTRAL, DRAGON) that focus on short-term outcomes using data predating widespread thrombectomy<sup>8,9</sup>, our study targets long-term recovery relevant to sustained quality of life. A key strength of our work is the rigorous temporal validation on an independent 2024 cohort. This design mimics a real-world prospective application, confirming the model's robustness against significant temporal shifts in patient characteristics. The deployment of our web-based predictor further bridges the gap between complex algorithms and bedside utility.

In feature selection, we employed a stringent multi-dimensional strategy combining statistical inference, regularization, and ensemble learning to ensure both data-driven accuracy and clinical interpretability. We excluded predictors like stroke-associated pneumonia due to diagnostic latency and subjectivity<sup>26,27</sup>. Instead, we selected objective biomarkers like NLR, D-Dimer, and BNP. NLR, a marker of systemic inflammation, reflects the secondary brain injury caused by neutrophil infiltration, which is known to worsen long-term recovery<sup>28</sup>. Similarly, elevated levels of D-dimer and BNP often signal cardioembolic etiologies or underlying cardiac dysfunction<sup>29,30</sup>. Integrating these quantifiable markers captures pathophysiological risks potentially overlooked by neurological scores alone, identifying high-risk subgroups within seemingly stable patients.

A pivotal finding is the superior robustness of our multivariate model over the admission NIHSS score in the temporal validation cohort. This cohort, representative of contemporary practice, presented with milder symptoms and higher reperfusion rates. In this “mild stroke” population, the discriminative power of the NIHSS diminished, likely due to its “floor effect”. While correlating well with infarct volume, NIHSS is less sensitive to subtle pathophysiological risks in minor deficits<sup>31</sup>. Moreover, NIHSS assessment carries inherent subjectivity and inter-rater variability<sup>32</sup>. Our model addresses these constraints by integrating objective, readily available biological markers (e.g., NLR, glucose) with the clinical score. For instance, patients with low NIHSS scores but high inflammatory or metabolic burdens are correctly reclassified as higher risk. Looking ahead, as public awareness improves and acute treatments are delivered more swiftly, the proportion of patients presenting with mild symptoms is expected to rise. Consequently, such integrated predictive tools will become increasingly crucial to complement and enhance traditional scale-based assessments, enabling more precise risk stratification in modern stroke care.

Beyond precise risk stratification, the predicted probability of 1-year excellent outcome offers tangible guidance for routine clinical practice. For instance, in terms of discharge planning and triage, identifying patients with a high likelihood of complete recovery (mRS 0–1) allows clinicians to confidently plan for early discharge with home-based rehabilitation, thereby optimizing bed turnover. Conversely, those with lower probabilities can be flagged early for comprehensive social work assessment to expedite transfer to skilled nursing facilities or subacute care settings. Furthermore, the model aids in the strategic allocation of rehabilitation resources; patients predicted to have unfavorable outcomes might be prioritized for enriched inpatient therapy or enrollment in clinical trials investigating neuroprotective therapies. Ultimately, by transforming a complex prognosis into a quantifiable probability, the web-based tool empowers clinicians to manage family expectations more effectively regarding the patient’s potential for independent living and return to work, thereby facilitating evidence-based shared decision-making.

It is also worth clarifying the distribution of outcomes in our cohort. The proportion of patients with unfavorable outcomes was approximately 45.4%, which may appear high relative to the low median admission NIHSS (Table 1). This may be primarily attributable to the strict definition of excellent outcome (mRS 0–1). Patients with slight disability (mRS 2), who are typically considered to have favorable outcomes, were classified as unfavorable in this study to target symptom-free recovery. Indeed, the incidence of post-discharge events (recurrent stroke and complications) was low and showed no statistically significant difference between the favorable and unfavorable outcome groups (Supplementary Table S5). This indicates that the outcome distribution is driven by our rigorous standard for recovery rather than a high burden of secondary medical events.

Several limitations must be acknowledged. First, this single-center study in Northeast China may reflect specific regional vascular risk profiles, necessitating multi-center validation for external generalizability. Second, reliance on admission data excludes the impact of post-discharge factors, such as rehabilitation adherence and socioeconomic support. Third, while discrimination was good (AUC ~ 0.80), future iterations could incorporate multi-omics or neuroimaging radiomics to capture granular biological heterogeneity and further enhance predictive precision.

## Conclusion

In summary, we successfully developed and rigorously validated a robust, interpretable machine learning model for predicting 1-year functional outcomes in patients with AIS. Moving beyond the admission NIHSS score, our 8-variable LR model integrates routine clinical and biomarker data to achieve superior, stable discrimination (AUC ~ 0.80) in a contemporary cohort characterized by milder strokes. Implemented as an accessible online platform, this tool facilitates early prognostic assessment and holds promise for guiding personalized long-term management strategies.

## Data availability

The data and code can be made available from the corresponding author upon reasonable request.

Received: 27 December 2025; Accepted: 23 March 2026

Published online: 27 March 2026

## References

1. Pu, L. et al. Projected global trends in ischemic stroke incidence, deaths and disability-adjusted life years from 2020 to 2030. *Stroke* **54**, 1330–1339 (2023).
2. Tu, W. J. & Wang, L. D. & Special Writing Group of China Stroke Surveillance Report. China stroke surveillance report 2021. *Mil Med. Res.* **10**, 33 (2023).
3. Cheng, Y. J. et al. Prolonged myelin deficits contribute to neuron loss and functional impairments after ischaemic stroke. *Brain* **147**, 1294–1311 (2024).
4. Kim, J. Y. et al. Long-term incidence of gastrointestinal bleeding following ischemic stroke. *J. Stroke.* **27**, 102–112 (2025).

5. Lu, Z. et al. Insulin resistance estimated by estimated glucose disposal rate predicts outcomes in acute ischemic stroke patients. *Cardiovasc. Diabetol.* **22**, 225 (2023).
6. Ebinger, M. et al. Association between dispatch of mobile stroke units and functional outcomes among patients with acute ischemic stroke in Berlin. *JAMA* **325**, 454–466 (2021).
7. Saposnik, G. et al. The iScore predicts effectiveness of thrombolytic therapy for acute ischemic stroke. *Stroke* **43**, 1315–1322 (2012).
8. Cooray, C. et al. External validation of the ASTRAL and DRAGON scores for prediction of functional outcome in stroke. *Stroke* **47**, 1493–1499 (2016).
9. Michel, P. et al. The Acute Stroke Registry and Analysis of Lausanne (ASTRAL). *Stroke* **41**, 2491–2498 (2010).
10. Strbian, D. et al. Predicting outcome of IV thrombolysis-treated ischemic stroke patients: The DRAGON score. *Neurology* **78**, 427–432 (2012).
11. Broderick, J. P., Adeoye, O. & Elm, J. Evolution of the modified rankin scale and its use in future stroke trials. *Stroke* **48**, 2007–2012 (2017).
12. Salim, H. A. et al. Endovascular therapy versus best medical management in distal medium middle cerebral artery acute ischaemic stroke: A multinational multicentre propensity score-matched study. *J. Neurol. Neurosurg. Psychiatry*. **96**, 239–248 (2025).
13. Mohammad, M. H. et al. Endovascular versus medical management in distal medium vessel occlusion stroke: The DUSK study. *Stroke* **55**, 1489–1497 (2024).
14. Daidone, M., Ferrantelli, S. & Tuttolomondo, A. Machine learning applications in stroke medicine: advancements, challenges, and future perspectives. *Neural Regeneration Res.* **19**, 769 (2024).
15. De Clares, J. B., Carneiro, T. S., Nunes Mendes, G. N., Dos Santos, N., Lima, J. C. & J. P. & A clinical-AI correlation for integrating artificial intelligence into stroke care: A systematized literature review and practice framework. *Int. J. Med. Informatics*. **208**, 106233 (2026).
16. Lee, M. et al. Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning. *Alz Res. Therapy*. **15**, 147 (2023).
17. Fernandez-Lozano, C. et al. Random forest-based prediction of stroke outcome. *Sci. Rep.* **11**, 10071 (2021).
18. Miyazaki, Y. et al. Logistic regression analysis and machine learning for predicting post-stroke gait independence: A retrospective study. *Sci. Rep.* **14**, 21273 (2024).
19. Vodencarevic, A. et al. Prediction of recurrent ischemic stroke using registry data and machine learning methods: The Erlangen stroke registry. *Stroke* **53**, 2299–2306 (2022).
20. Bedoya, A. D. et al. Machine learning for early detection of sepsis: An internal and temporal validation study. *Jamia Open*. **3**, 252–260 (2020).
21. Wangqin, R. et al. International comparison of patient characteristics and quality of care for ischemic stroke: Analysis of the China National Stroke Registry and the American Heart Association get with the guidelines—stroke program. *J. Am. Heart Association*. **7**, e010623 (2018).
22. Liu, L. et al. Chinese Stroke Association guidelines for clinical management of ischaemic cerebrovascular diseases: Executive summary and 2023 update. *Stroke Vasc Neurol.* **8** (2023).
23. Bennett, D. A. How can I deal with missing data in my study? *Aust. N. Z. J. Public Health*. **25**, 464–469 (2001).
24. Kuhn, M. & Johnson, K. *Applied Predictive Modeling* (Springer, 2013). <https://doi.org/10.1007/978-1-4614-6849-3>
25. Zhu, Z. et al. Elevated NT-proBNP predicts unfavorable outcomes in patients with acute ischemic stroke after thrombolytic therapy. *BMC Neurol.* **23**, 203 (2023).
26. Smith, C. J. et al. Diagnosis of stroke-associated pneumonia. *Stroke* **46**, 2335–2340 (2015).
27. Assefa, M., Tadesse, A., Adane, A., Yimer, M. & Tadesse, M. Factors associated with stroke associated pneumonia among adult stroke patients admitted to University of Gondar hospital, Northwest Ethiopia. *Sci. Rep.* **12**, 12724 (2022).
28. Gong, P. et al. The association of neutrophil to lymphocyte ratio, platelet to lymphocyte ratio, and lymphocyte to monocyte ratio with post-thrombolysis early neurological outcomes in patients with acute ischemic stroke. *J. Neuroinflammation*. **18**, 51 (2021).
29. Zhang, P., Wang, C., Wu, J. & Zhang, S. A. Systematic review of the predictive value of plasma D-dimer levels for predicting stroke outcome. *Front Neurol.* **12** (2021).
30. Kai, G. et al. Elevated BNP and NT-proBNP levels as prognostic biomarkers of short-term walking independence in post-acute stroke rehabilitation. *Eur. J. Prev. Cardiol.* **31**, zwae175098 (2024).
31. De Santis, F. et al. Acute treatment of disabling and nondisabling minor ischemic stroke: Expert guidance for clinicians. *Stroke*. **57** (2025).
32. Comer, A. R. et al. National Institutes of Health Stroke Scale (NIHSS) scoring inconsistencies between neurologists and emergency room nurses. *Front. Neurol.* **13**, 1093392 (2022).

## Author contributions

P.F.L. conceived the study, supervised the project, and acquired funding. P.P.L. designed the methodology and developed the software. P.J.L. and X.B.Z. performed the formal analysis. P.J.L., Y.C., X.B.Z., and J.F. conducted the investigation and data collection. Z.B.Z. validated the results. P.J.L. wrote the original draft of the manuscript. P.P.L. and P.F.L. reviewed and edited the manuscript. All authors read and approved the final manuscript.

## Funding

This study was supported by grant from the Joint Fund Cultivation Project of Natural Science Foundation of Heilongjiang Province of China (No. PL2024H099).

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-45800-x>.

**Correspondence** and requests for materials should be addressed to P.L. or P.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026