

RESEARCH

Open Access



Machine learning-based predictive model for the development of thrombolysis resistance in patients with acute ischemic stroke

Xiaorui Wang¹, Song Luo^{1*}, Xue Cui¹, Hongdang Qu¹, Yujie Zhao¹ and Qirong Liao¹

Abstract

Background The objective of this study was to establish a predictive model utilizing machine learning techniques to anticipate the likelihood of thrombolysis resistance (TR) in acute ischaemic stroke (AIS) patients undergoing recombinant tissue plasminogen activator (rt-PA) intravenous thrombolysis, given that nearly half of such patients exhibit poor clinical outcomes.

Methods Retrospective clinical data were collected from AIS patients who underwent intravenous thrombolysis with rt-PA at the First Affiliated Hospital of Bengbu Medical University. Thrombolysis resistance was defined as $([\text{National Institutes of Health Stroke Scale (NIHSS) at admission} - 24\text{-hour NIHSS}] \times 100\% / \text{NIHSS at admission}) \leq 30\%$. In this study, we developed five machine learning models: logistic regression (LR), extreme gradient boosting (XGBoost), support vector machine (SVM), the least absolute shrinkage and selection operator (LASSO), and random forest (RF). We assessed the model's performance by using receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA), and presented the results through a nomogram.

Results This study included a total of 218 patients with AIS who were treated with intravenous thrombolysis, 88 patients experienced TR. Among the five machine learning models, the LASSO model performed the best. The area under the curve (AUC) on the testing group was 0.765 (sensitivity: 0.767, specificity: 0.694, accuracy: 0.727). The apparent curve in the calibration curve was similar to the ideal curve, and DCA showed a positive net benefit. Key features associated with TR included NIHSS at admission, blood glucose, white blood cell count, neutrophil count, and blood urea nitrogen.

Conclusion Machine learning methods with multiple clinical variables can help in early screening of patients at high risk of thrombolysis resistance, particularly in contexts where healthcare resources are limited.

Keywords Acute ischaemic stroke, Intravenous thrombolysis, Machine learning, Prediction model

*Correspondence:

Song Luo

542462407@qq.com

¹Department of Neurology, The First Affiliated Hospital of Bengbu Medical University, Bengbu 233004, China



© The Author(s) 2024. **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/>.

Background

Ischemic stroke ranks as the second most prevalent cause of disability and mortality on a global scale, leading to significant financial burdens in healthcare expenditures, as well as productivity and economic setbacks [1, 2]. Approximately 17.8 million adults in China are estimated to have suffered a stroke in 2020 [3]. The utilization of recombinant tissue plasminogen activator (rt-PA) has substantially enhanced the effectiveness of treating acute ischemic stroke (AIS) [4]. However, a considerable portion of patients continue to experience unfavorable functional outcomes or mortality [5], commonly referred to as “thrombolysis resistance (TR)”, also known as “non-response to thrombolysis”, or “ineffective thrombolysis” or “poor response to thrombolytic therapy”. TR is a complex clinical phenomenon characterized by the persistence of thrombus despite standard thrombolytic therapy. It has been shown that platelet-rich thrombi had TR [6]. Extracellular traps of neutrophils are expected to be a new targeting measure for TR [7]. Addressing the TR necessitates a thorough examination of multiple factors and the implementation of personalized treatment strategies. However, TR is one of the factors that affect the prognosis of AIS, with other factors including blood pressure and subsequent severe brain damage, etc. [8, 9]. Therefore, the development of clinical prediction models is of paramount importance in optimizing patient selection and outcomes in intravenous thrombolysis (IVT) treatment. Currently, there is no uniformity in judging the effectiveness of thrombolysis, and previous studies have focused on the use of multimodal imaging techniques [10]. However, due to the strict treatment time window and relatively complex equipment requirements, it is difficult to promote and popularise it in primary hospitals. It is particularly important to find a scientific and convenient method to predict the effect of thrombolysis and prepare the bridging treatment after thrombolysis failure in advance.

Machine learning (ML) algorithms, known for their ability to optimize information for enhanced prediction accuracy, are becoming more prevalent in the field of medicine for quantifying risk, identifying predictors, and creating precise predictive models for diagnosis and prognosis [11, 12]. In this research, a predictive model for the risk of thrombosis was developed using machine learning techniques, aiding clinicians in timely preparation of high-risk patients for bridging endovascular treatment (EVT) and ultimately improving patient outcomes.

Methods

Patient cohorts

Patients diagnosed with AIS at our institution between January 2021 and December 2022 were selected for this study. All patients received rt-PA intravenous

thrombolysis within 4.5 h of the onset of symptoms, and were not bridged for endovascular treatment. The inclusion criteria were as follows: 1. The clinical manifestations and imaging data of the patient were consistent with acute ischemic stroke [13]; 2. Age ≥ 18 years; 3. Complete clinical data and signed the informed consent for intravenous thrombolytic therapy and enrollment; 4. The Ethics Committee of our hospital has approved this study, and followed the Declaration of Helsinki. The exclusion criteria were as follows: 1. Haemorrhagic cerebrovascular disease, mixed cerebral infarction; 2. Other organic diseases of the nervous system, such as encephalitis, tumour, traumatic brain injury, etc.; 3. Accompanied by other serious systemic diseases, such as recent serious infections, previous diseases of the haematological system, etc.; 4. Mental disorders or consciousness disorder after the onset of the disease and unable to cooperate with the examination (Fig. 1).

Data preprocessing

The variables identified by the box plot that contain outliers are excluded. In addition, for variables with a missing rate of less than 5%, we will impute the missing values using the mode for the median for continuous variables. For the sensitivity analyses, we pursued multiple imputation using the multivariate imputation by chained equations method.

Data collection and variable selection

Given the limitations of the sample size, univariate logistic regression was initially performed on the collected variables. Any variables exhibiting non-statistically significant differences were subsequently excluded.

Clinical data: gender, age, hypertension (receiving medications for hypertension or blood pressure $\geq 140/90$ mmHg on repeated measurements) [14]; diabetes mellitus (receiving medications for diabetes mellitus, fasting blood sugar ≥ 126 mg/dL or HbA1c $\geq 6.5\%$, or a casual plasma glucose > 200 mg/dL) [14]; atrial fibrillation (AF): receiving medication for the treatment of AF or sustained episodes of AF heart rate > 30 s recorded by 12-lead electrocardiogram (ECG), single-lead ECG, or Holter monitor [15]; Systolic blood pressure (SBP), diastolic blood pressure (DBP), and door-needle time (DNT) were collected at admission. National Institutes of Health Stroke Scale (NIHSS) [16] at admission and 24-hours NIHSS after thrombolysis. Note: NIHSS at admission and 24-hour NIHSS were assessed by the same physician.

Laboratory data were pre-thrombolytic indicators, including blood glucose (Glu), white blood cell count (WBC), neutrophil count (Neu), lymphocyte count (Lym), monocyte count (Mon), platelet count (PLT), D-dimer, fibrinogen (FIB), creatinine (Cr), blood urea nitrogen (BUN), total cholesterol (TC).

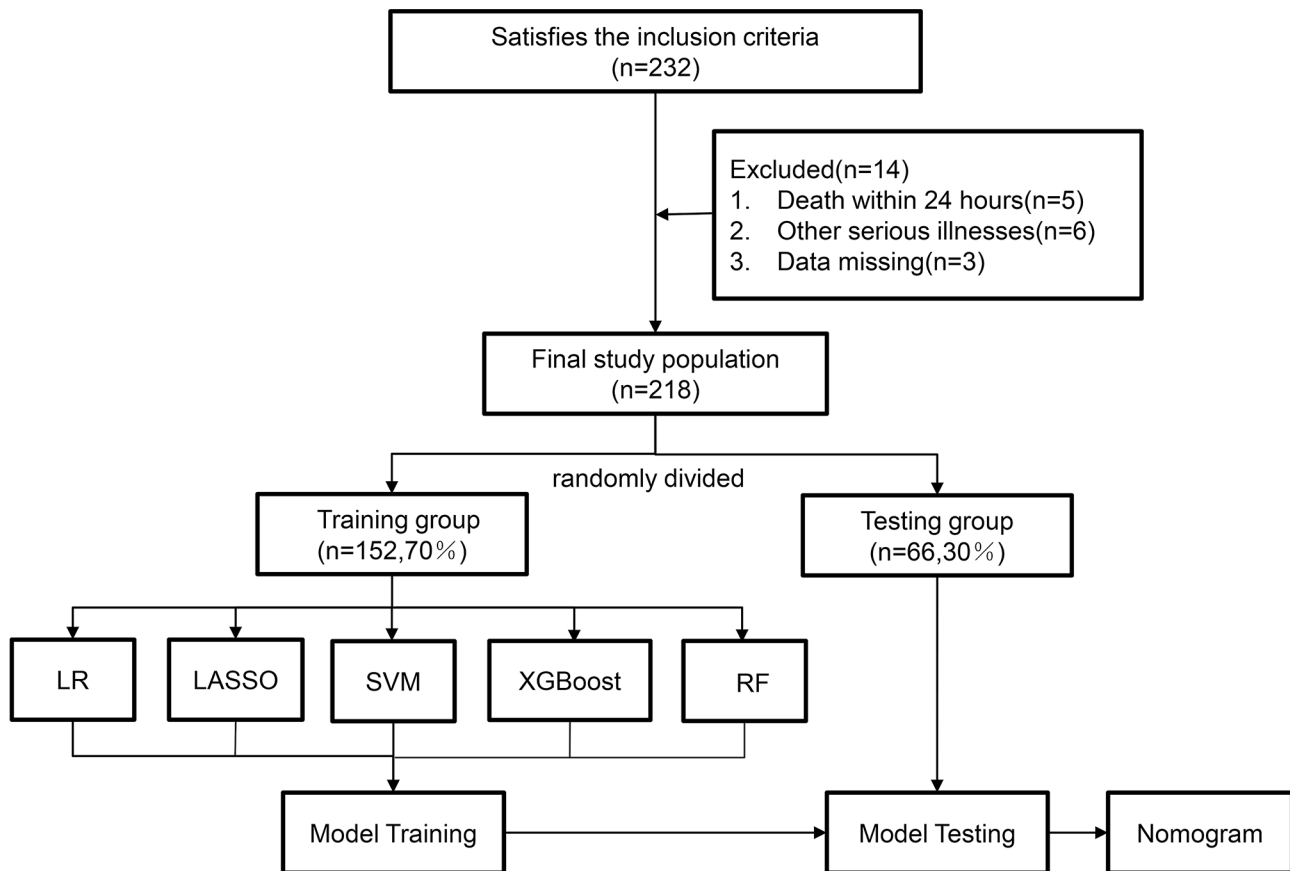


Fig. 1 Flowchart of the study population and process. LR, logistic regression; LASSO, the least absolute shrinkage and selection operator; SVM, support vector machine; XGBoost, extreme gradient boosting; RF, random forest; ML, machine learning

Definition of functional outcome.

Thrombolysis resistance was defined as $([\text{NIHSS at admission} - 24\text{-hour NIHSS}] \times 100\% / \text{NIHSS at admission}) \leq 30\%$ [17].

Model development and evaluation

The data were randomly divided into training and testing groups of 70% and 30%. We utilize the training group to ascertain the suitable parameters of the model and assess their efficacy by employing the testing group. There exist numerous approaches to machine learning, yet a definitive conclusion regarding the most effective one remains elusive. Five machine learning algorithms were used in this study: logistic regression (LR), the least absolute shrinkage and selection operator (LASSO), extreme gradient boosting (XGBoost), support vector machine (SVM), and random forest (RF), similar with previous study [18, 19]. This study employs a 10-fold cross-validation approach for model derivation, and utilizes a combination of the grid search algorithm to optimize the model hyperparameters. To evaluate the performance of each model, the features derived from the training group are applied to the testing group. In this study, the receiver operating characteristic (ROC) curves of the five

models were plotted and the performance of the models was evaluated based on the area under the curve (AUC), specificity, sensitivity, accuracy, positive predictive value (PPV), negative predictive value (NPV), precision, recall, and F1 value. Constructing a nomogram based on the output of the most influential variables from the best model. Finally, the accuracy, calibration, and clinical applicability of the model were evaluated in the testing group using ROC curves, calibration curves, and decision curve analysis (DCA) (Fig. 1).

Statistical analysis

Values are presented as the mean \pm standard deviation, median(interquartile range) for continuous variables, or as the number (%) of subjects for categorical variables, as appropriate. Comparisons of the characteristics between the two groups were performed by the t-test, chi-square test, Mann-Whitney U parametric test according to the type of the variable. The statistical significance was defined as a two-tailed *P*-value of <0.05 . Analyses were performed in SPSS (version 26, IBM, New York, NY, USA) and R (version 3.3.2; R Project for Statistical Computing) using the Caret ML library.

Results

Characteristics of study population

According to the inclusion criteria 232 patients with AIS treated with intravenous thrombolysis were selected and finally 218 patients were included after excluding 14 patients according to the exclusion criteria (Fig. 1). The median age was 68 years, 149 (68.3%) were male, 135 (61.9%) had hypertension, 46 (21.1%) had DM, 19 (8.7%) had AF, and 88 (40.4%) developed thrombolysis resistance. The study population was randomly divided into training ($n=152$) and testing ($n=66$) groups on 7:3 basis, and there was no statistical difference in the data characteristics between the two groups ($P>0.05$) (Table 1).

Prediction model performance

In this study, five models, including LR, LASSO, SVM, XGBoost and RF, were established, and the ROC curves of the five models were constructed (Fig. 2). In the training group, it can be seen that the LASSO model has the largest AUC (0.789), and its specificity of 0.777, PPV of 0.672, accuracy of 0.763, precision of 0.672, and F1 value of 0.705 are better than the other models; In the testing group, the LASSO model exhibited superior performance compared to other models in terms of AUC (0.765), specificity of 0.694, PPV of 0.676, accuracy of 0.727, and precision of 0.676. Consequently, LASSO is considered to

have good overall performance and selected as the best model (Fig. 2 and Table 2).

Figure 3A shows a plot of the coefficients of the LASSO variables. Each curve in the graph represents the trajectory of the coefficient of an independent variable. The vertical coordinate is the value of the coefficient. The lower horizontal coordinate λ is the parameter controlling the severity of the penalty. The upper horizontal coordinate is the number of non-zero coefficients in the model under the penalty parameter. Figure 3B shows the tuning parameters in the LASSO model, filtered by 10-fold cross-validation. Vertical dashed lines are plotted at 1 standard error (SE) of the minimum and maximum criteria (1-SE criterion) to select the optimal λ value for the LASSO model. The 20 features were reduced to 5 potential predictors of the occurrence of thrombolysis resistance according to the LASSO method.

Nomogram

A nomogram of thrombolysis resistance was constructed by screening the five optimal predictor variables, including NIHSS at admission, GLu, WBC, Neu, and BUN (Fig. 4). Each predictor variable was represented by a vertical line on the value axis, with the intersection of the line and scoring axis indicating the variable's score. The predicted probability of thrombolysis resistance was determined by summing the scores of each risk factor

Table 1 Clinical characteristics of the training and testing groups

	Overall	Training group	Testing group	$t/\chi^2/Z$	P value
N, n(%)	218	152(70.0)	66(30.0)		
Outcome, n(%)	88 (40.4)	58 (38.2)	30 (45.5)	0.737	0.391
Age, years	68.00 [58.00, 76.00]	68.00 [58.00, 75.00]	67.00 [58.00, 76.00]	-0.138	0.890
Gender, Mele, n(%)	149 (68.3)	100 (65.8)	49 (74.2)	1.154	0.283
Hypertension, n(%)	135 (61.9)	97 (63.8)	38 (57.6)	0.518	0.472
Diabetes mellitus, n(%)	46 (21.1)	30 (19.7)	16 (24.2)	0.323	0.570
Atrial fibrillation, n(%)	19 (8.7)	12 (7.9)	7 (10.6)	0.153	0.696
NIHSS at admission	6.00 [4.00, 11.75]	5.50 [3.00, 12.00]	6.50 [4.00, 10.75]	-0.748	0.455
Glucose, mmol/L	7.10 [6.10, 9.04]	7.08 [6.08, 8.23]	7.36 [6.10, 10.10]	-0.989	0.323
SBP, mmHg	146.50 [135.00, 164.75]	147.50 [135.00, 164.50]	142.50 [133.00, 164.75]	-0.624	0.533
DBP, mmHg	86.00 [78.00, 95.00]	86.00 [78.00, 95.00]	85.00 [78.50, 91.75]	-0.220	0.826
DNT, h	0.62 [0.50, 0.78]	0.60 [0.50, 0.80]	0.65 [0.50, 0.78]	-0.230	0.818
WBC, $10^9/L$	7.84 [6.02, 9.76]	7.94 [6.07, 9.71]	7.50 [6.00, 9.80]	-0.818	0.413
Neu, $10^9/L$	5.10 [3.77, 7.27]	5.10 [3.78, 7.52]	5.14 [3.76, 6.54]	-0.650	0.516
Lym, $10^9/L$	1.56 [1.15, 2.24]	1.55 [1.17, 2.20]	1.57 [1.12, 2.25]	-0.108	0.914
Mon, $10^9/L$	0.44 [0.36, 0.59]	0.44 [0.36, 0.59]	0.43 [0.33, 0.60]	-0.727	0.467
PLT, $10^9/L$	198.00 [161.00, 233.75]	194.50 [160.00, 233.50]	204.50 [167.75, 233.50]	-0.630	0.529
D-dimer, ug/ml	0.50 [0.25, 1.01]	0.49 [0.26, 1.06]	0.51 [0.22, 0.94]	-0.797	0.425
FIB, g/L	3.49 [2.82, 3.95]	3.45 [2.85, 3.90]	3.52 [2.78, 4.05]	-0.008	0.993
Cr, umol/L	66.00 [55.00, 77.75]	66.00 [54.00, 77.25]	65.00 [57.25, 77.50]	-0.212	0.832
BUN, mmol/L	6.34 [5.37, 7.64]	6.34 [5.34, 7.70]	6.36 [5.57, 7.19]	-0.050	0.996
TC, mmol/L	3.79 [3.32, 4.69]	3.92 [3.48, 4.69]	3.69 [3.16, 4.60]	-1.191	0.234

Abbreviations NIHSS, national institutes of health stroke scale; SBP, systolic blood pressure; DBP, diastolic blood pressure; DNT, door-needle time; WBC, white blood cell count; Neu, neutrophil count; Lym, lymphocyte count; Mon, monocyte count; PLT, platelet count; FIB, fibrinogen; Cr, creatinine; BUN, blood urea nitrogen; TC, total cholesterol

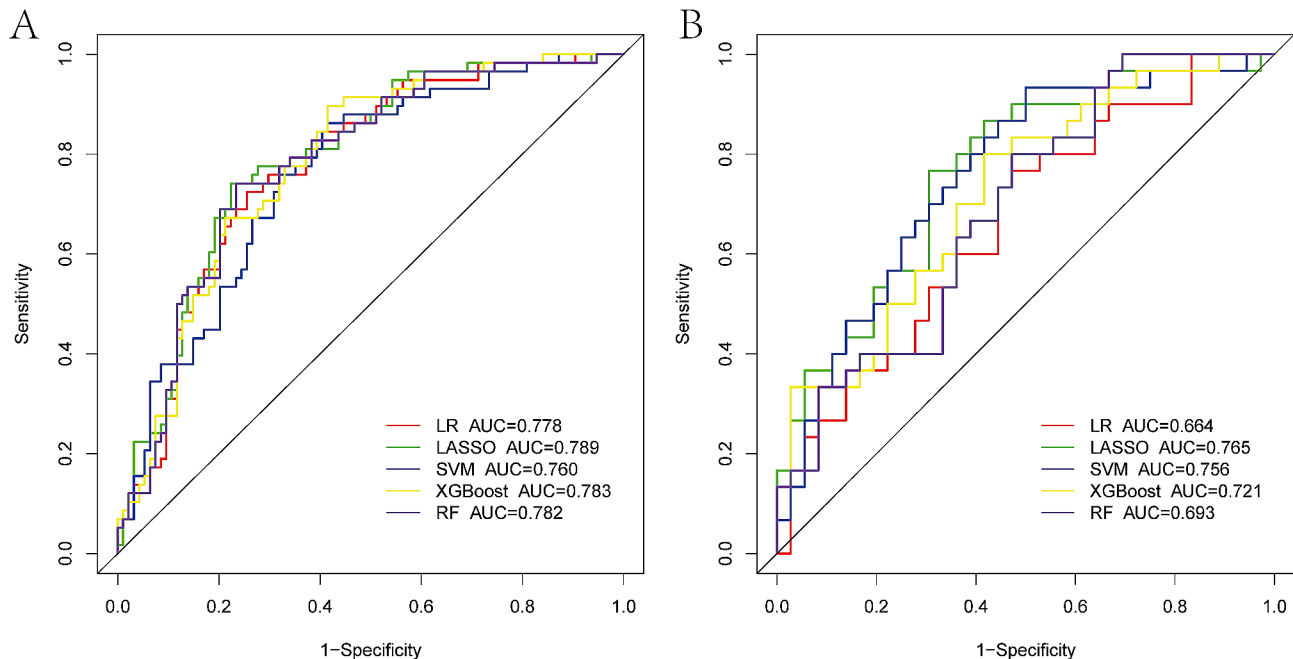


Fig. 2 Comparison of the receiver operating characteristic (ROC) curves for machine learning models in which thrombolysis resistance occurs. **(A)** Training group. **(B)** Testing group. LR, logistic regression; LASSO, the least absolute shrinkage and selection operator; XGBoost, extreme gradient boosting; SVM, support vector machine; RF, random forest; AUC, area under the curve

Table 2 Statistical comparison of machine learning models for the occurrence of thrombolysis resistance

Model	AUC(95%CI)	sensitivity	specificity	PPV	NPV	Accuracy	Precision	Recall	F1 value
Training group									
LR	0.778(0.704–0.852)	0.724	0.745	0.636	0.814	0.737	0.636	0.724	0.677
LASSO	0.789(0.717–0.862)	0.741	0.777	0.672	0.830	0.763	0.672	0.741	0.705
SVM	0.760(0.683–0.837)	0.862	0.585	0.562	0.873	0.691	0.562	0.862	0.680
XGBoost	0.783(0.710–0.855)	0.897	0.585	0.571	0.902	0.704	0.571	0.897	0.698
RF	0.782(0.707–0.856)	0.741	0.766	0.662	0.828	0.757	0.662	0.741	0.699
Testing group									
LR	0.664(0.532–0.796)	0.767	0.528	0.575	0.731	0.636	0.575	0.767	0.657
LASSO	0.765(0.649–0.881)	0.767	0.694	0.676	0.781	0.727	0.676	0.767	0.717
SVM	0.756(0.638–0.874)	0.933	0.500	0.609	0.900	0.697	0.609	0.933	0.737
XGBoost	0.721(0.599–0.844)	0.800	0.583	0.615	0.778	0.682	0.615	0.800	0.695
RF	0.693(0.566–0.819)	0.800	0.528	0.585	0.760	0.652	0.585	0.800	0.558

Abbreviations LR, logistic regression; LASSO, the least absolute shrinkage and selection operator; XGBoost, extreme gradient boosting; SVM, support vector machine; RF, random forest; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value

and projecting the total on the predictive probability axis. As the values of the 5 variables increase, the total score will become higher, the 5 variables are positively associated with a high risk of thrombolysis resistance.

Internal validation of the predictive model

The AUC values for the training and testing groups were 0.789 and 0.765, respectively. This suggests that the model has a reasonable ability to discriminate between subjects with thrombolysis resistance and those without it (Fig. 5A). DeLong's test was used to statistically compare the AUCs of the training and testing groups. The *P*-value of 0.725 is greater than the significance level of

0.05, indicating that the difference in AUCs between the two groups is not statistically significant. This suggests that the model's predictive accuracy is consistent across different datasets. The performance of the Calibration curve was internally validated by Bootstrap resampling (1000 times) analysis, The Hosmer-Lemeshow goodness-of-fit test was used to assess how well the model fits the observed data. *P*-values of 0.1225 for the training group and 0.4124 for the testing group, both greater than 0.05, indicate that the model fits the data well in both groups, which showed that the nomogram predicted the probability of thrombolysis resistance in good agreement with the actual probability. (Figure 5B and C). DCA were

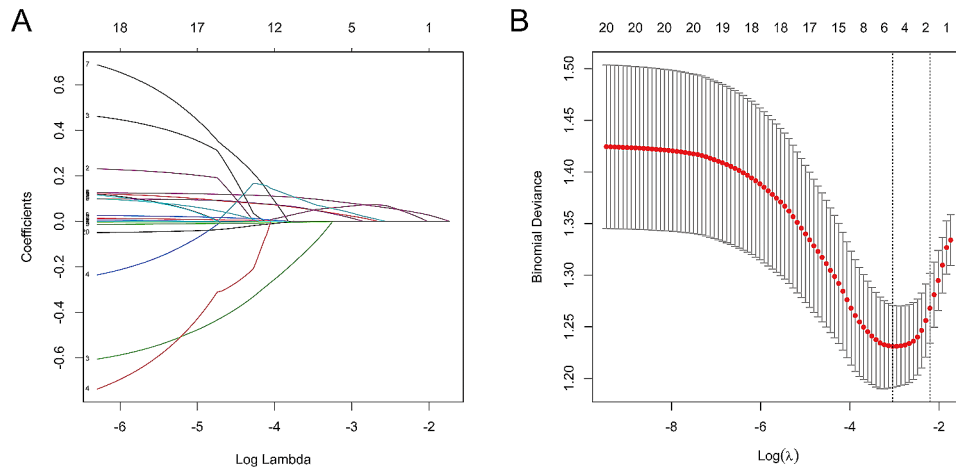


Fig. 3 Feature selection based on LASSO model. **(A)** Identification of the optimal penalization coefficient λ in the LASSO model. **(B)** Cross-validation for tuning parameter selection in the LASSO model. LASSO, the least absolute shrinkage and selection operator

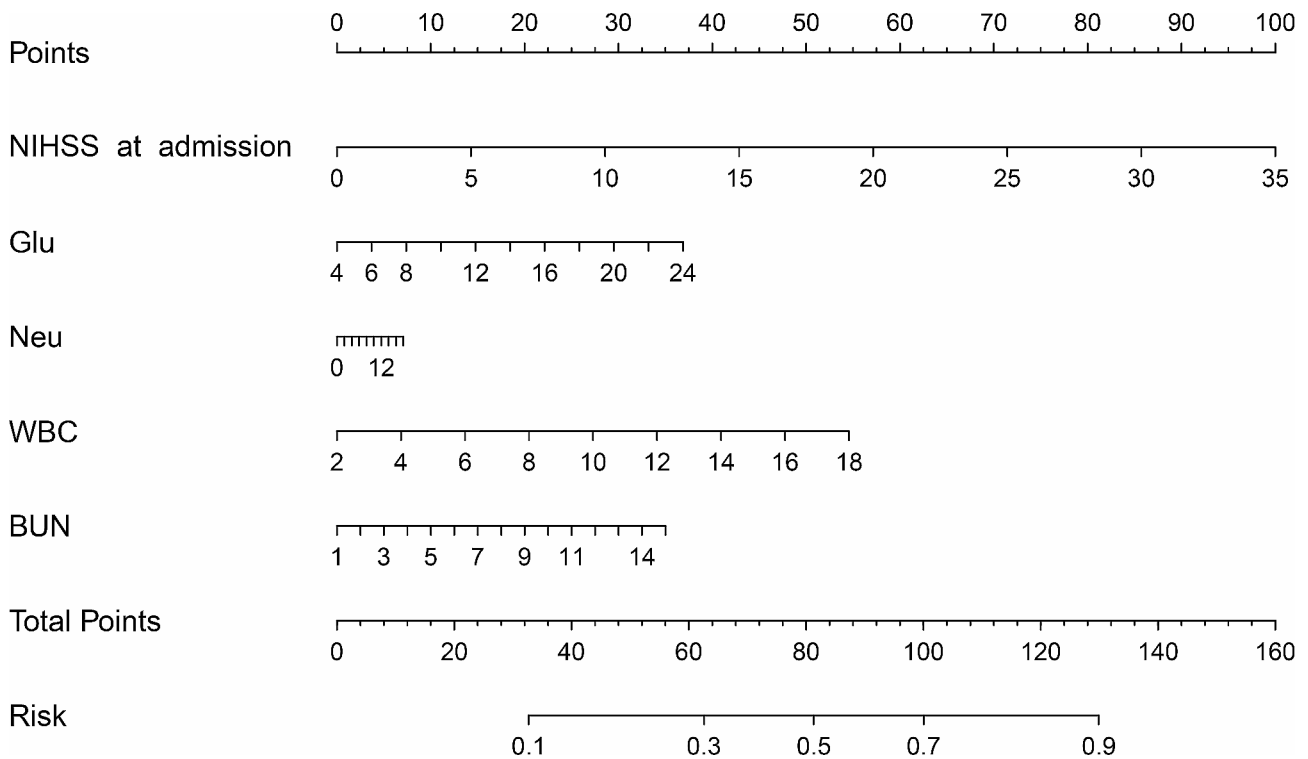


Fig. 4 Proposed nomogram for thrombolysis resistance. NIHSS, national institutes of health stroke scale; Glu, glucose; WBC, white blood cell count; Neu, neutrophil count; BUN, blood urea nitrogen

plotted to evaluate the clinical applicability of the model by comparing the net benefit of the model's predictions against a range of threshold probabilities. The positive net benefit in both the training and testing groups (Fig. 5D) indicates that the model's predictions provide a clinical advantage over a strategy of not using the model, suggesting that the model has good clinical utility.

Important variables of the machine learning models

After evaluating the significance of each feature, the 5 machine learning algorithms have ranked the top 5 variables based on their discriminatory performance. LR, LASSO, RE,SVM and XGBoost can estimate the contribution of each feature to the model by calculating the absolute value of the standardized regression coefficient, Gini coefficient, weight coefficient, and Gain value (Table 3). NIHSS at admission and Glu were among the top 5 variables in all models simultaneously, with Neu

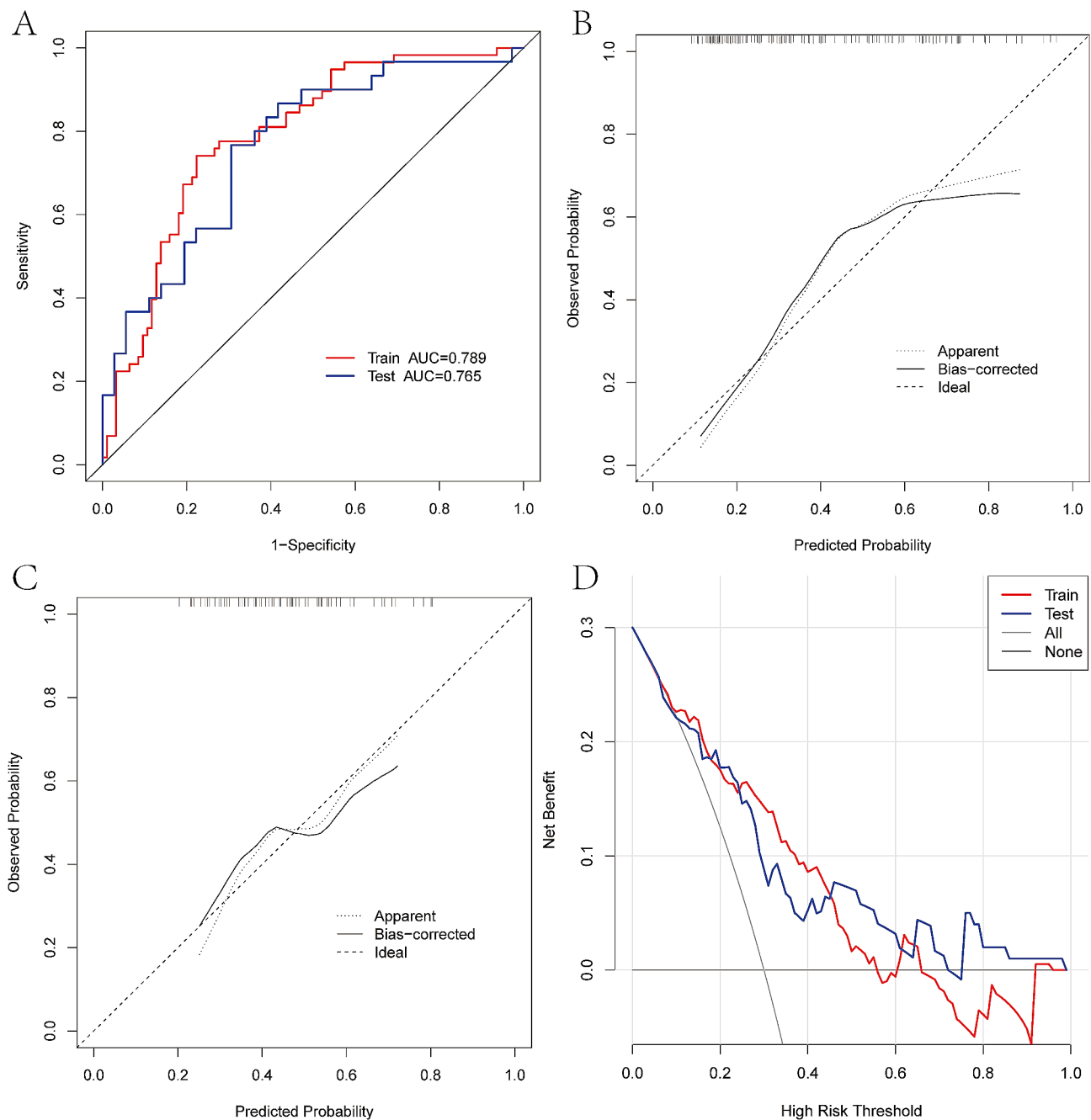


Fig. 5 Internal validation diagram for the model. **(A)** ROC curves for the training and testing groups; **(B)** Calibration curve for the training group, $P=0.1225$; **(C)** Calibration curve for the testing group, $P=0.4124$; **(D)** DCA for the training and testing groups

appearing in 4 models. The results of LASSO and RF were comparable, but only 1 variable differed, and WBC was identified as an important variable in both. Blood pressure-related indicators were found to be significant characteristics in LR and SVM. FIB was identified in XGBoost, RE, while BUN only appeared in LASSO.

Discussion

Currently, numerous studies have referenced about thrombolytic effect, but there is no precise definition or specific machine prediction model research for TR. In this study, the term “TR” was proposed based on the early thrombolysis effect in AIS patients, and presents a risk prediction model for identifying patients at risk for thrombolysis resistance by integrating easily accessible clinical variables with machine learning algorithms.

Table 3 Top 5 important features in each models

	LR*	LASSO	SVM	XGBoost	RF
1	Neu	NIHSS at admission	NIHSS at admission	FIB	NIHSS at admission
2	NIHSS at admission	WBC	Glu	Neu	Neu
3	Glu	Glu	DBP	TC	WBC
4	SBP	BUN	Hypertension	NIHSS at admission	Glu
5		Neu	SBP	Glu	FIB

Abbreviations LR, logistic regression; LASSO, the least absolute shrinkage and selection operator; XGBoost, extreme gradient boosting; SVM, support vector machine; RF, random forest; NIHSS, national institutes of health stroke scale; SBP, systolic blood pressure; DBP, diastolic blood pressure; WBC, white blood cell count; Neu, neutrophil count; FIB, fibrinogen; BUN, blood urea nitrogen; TC, total cholesterol; Glu, glucose.* :Only 4 variables were screened out after stepwise logistic regression

In clinical studies, there is a lack of standardization in the criteria used to evaluate the efficacy of early thrombolysis in patients with AIS. The modified Rankin Scale (mRS) scores and NIHSS score are two commonly used stroke assessment tools. The mRS score commonly used as the primary efficacy endpoint in clinical trials for its simplicity and ease of interpretation [20]. The mRS scores are typically assessed 3–6 months after thrombolysis and are susceptible to a number of factors, and the changes in 24-hour NIHSS scores were chosen in this study as a potentially more effective way to assess the early effects of thrombolysis. Some studies have used a decrease of ≥ 4 points or an increase of ≥ 4 points in the NIHSS from baseline to 24 hours after intravenous thrombolytic therapy to assess the early therapeutic effect of thrombolysis in patients [21, 22]. This study concluded that a 4-point decrease or increase in admission NIHSS for patients with different admission NIHSS does not represent the same degree of improvement or deterioration, and that the use of percentage change in NIHSS better explains the wide variation in patients' admission scores and more accurately determines the effectiveness of thrombolysis [23]. A study on mechanical thrombolysis found that $>30\%$ improvement in NIHSS score at the end of thrombolysis was a reliable prognostic predictor [24]. Meanwhile, data from a multicentre, prospective trial of intravenous alteplase for the treatment of AIS showed that there was no difference in the change in baseline versus 24-hour NIHSS difference between patients in the alteplase-treated group compared with the placebo group, whereas the change in the 24-hour NIHSS percentage was more pronounced, and 30% was the optimal cut-off value [17]. Consequently, TR was considered to be defined in this study as an improvement in NIHSS score of $\leq 30\%$ from the admission NIHSS 24 hours after thrombolysis.

A total of 218 patients were included in this study, and 88 (40.4%) patients developed thrombolysis resistance.

We conducted TR prediction models using ML algorithms, including LR, LASSO, XGBoost, SVM, and RF. Among the five models, LASSO was rated as the best model with superior AUC, specificity, PPV, accuracy, and precision. LASSO is based on a regression analysis method that does not rely on statistical significance for regularisation, but rather narrows down the coefficients of complex experimental variables and excludes relatively unimportant variables [25]. A Chinese study based on the GEO database to screen for differential genes developed a LASSO regression model to better identify the prognosis of AIS, and the AUC of its LASSO regression model was 0.969 [26]. A retrospective case-control study identifying AIS based on LASSO regression indicated that the model had a better identification ability (AUC=0.916) [25]. The LASSO regression model in this study had an AUC value of 0.765, which was a good predictor of whether thrombolysis resistance occurred in patients with AIS, suggesting the potential utility of LASSO regression in predicting disease prognosis or diagnosis. We found that the sensitivity value of 0.767 in the LASSO model is slightly lower than that of the RF model (0.800) (Table 2). RF model is typically capable of handling more complex data patterns and possess a certain degree of resistance to overfitting. Combining the AUC results, we thought the LASSO model is better than RF model. However, it is necessary to investigate this results with a large amount of data in the future study.

The nomogram (Fig. 4) made by the best model in this study showed that NIHSS at admission, Glu, WBC, Neu and BUN were significant predictors of the occurrence of TR. Several studies have confirmed that these five metrics are associated with poor prognosis in AIS patients [27–31], which is consistent with our findings, indicating that our model has scientific validity and credibility in real-world practice. A machine-learning based multicentre retrospective study concluded that the key features predicting functional outcome in AIS contain NIHSS at admission, and white blood cell count [32]. Meanwhile, a predictive model examining ineffective recanalisation in AIS considered NIHSS score, blood glucose, c-reactive protein, and creatinine as important predictors [33]. It suggests that inflammatory response may be a crucial factor in the development of AIS [34]. Previous studies on Cr in the model and on BUN in the present study has primarily focused on renal function, with findings indicating that elevated BUN levels upon admission may indicated haemodynamic deterioration [35], a known predictor of poor prognosis and death in AIS [36]. In this study, it was hypothesised that it might be related to the shift of blood to vital organs such as heart and brain during the onset of AIS, resulting in a decrease in blood flow to the kidneys. In contrast, the role of blood glucose as a predictor of thrombolytic prognosis may be related to the fact that

hyperglycaemia increases platelet mitochondrial reactive oxygen species production, which enhances procoagulant platelet formation and significantly increases infarct size [37]. Moreover, It is important to consider the influence of variables such as FIB, blood pressure-related markers, and TC, which have been included in other models, on the phenomenon of TR in this study. A predictive model on the prognosis of intravenous thrombolysis in patients with mild AIS identified higher NIHSS scores, DM, and deep infarcts in the middle cerebral artery (MCA) as potential predictors of a poor prognosis [38]. These findings align with those of the present model, which incorporates imaging data. For healthcare institutions lacking magnetic resonance imaging, the variables included in the LASSO model are more accessible, and patients at high risk for thrombolysis resistance are promptly transferred to hospitals equipped for endovascular therapy for further treatment. Yang H et al. [39]. used LASSO regression to construct the prognosis of patients with AIS found that NIHSS, SBP, lymphocyte percentage (LYM%), neutrophil-to-lymphocyte ratio (NLR), platelet-to-neutrophil ratio (PNR), and platelet-to-lymphocyte ratio (PLR) independently predicted early neurologic deterioration (END), which, after the incorporating the composite inflammatory index improved the AUC of the test set from 0.635 to 0.816. This study did not include the composite index for modeling for the time being, which is a good direction for research.

All the variables considered in this model are readily available at the time of admission, and when the values of the above five variables are entered into this model, clinicians will be able to quickly assess the effectiveness of thrombolysis in patients with AIS in advance, and decide whether to perform IVT/EVT or not. Limitations: There are numerous machine learning methods available, this study only focuses on 5 specific algorithms. To increase the stability of the model, additional machine learning algorithms could be incorporated. Additionally, The causes of poor functional prognosis or death after thrombolysis cannot be explained by a single theory, rather, they may be the result of a combination of factors. The study population could be expanded, more variables could be selected to build the model, and the model could be prospectively validated in an independent cohort. Furthermore, the definition of “thrombolysis resistance” proposed in this study has yet to be generalized to other studies. Additional research is necessary to verify the scientific validity and effectiveness of this definition.

Conclusion

In this study, we used a machine learning approach to screen five optimal predictors of thrombolysis resistance and constructed a nomogram to predict the risk of thrombolysis resistance. Among them, higher NIHSS at

admission, WBC, Glu, Neu, and BUN were the risk factors for the occurrence of thrombolysis resistance. This model has the potential to assist clinicians in the formulation of personalized treatment plans. Conducting external validation to evaluate the broader applicability and to detect any possible biases is a crucial subsequent phase for this research endeavor. From our results, we advocate for future investigations focusing on the application of ML to predict thrombolysis resistance to embrace a multi-faceted modeling approach. This strategy will enhance the robustness and reliability of predictive models in this domain.

Abbreviations

rt-PA	Recombinant tissue plasminogen activator
AIS	Acute ischaemic stroke
TR	Thrombolysis resistance
ML	Machine learning
EVT	Endovascular treatment
IVT	Intravenous thrombolysis
NIHSS	National Institutes of Health Stroke Scale
SBP	Systolic blood pressure
DBP	Diastolic blood pressure
DNT	Door-needle time
WBC	White blood cell count
Neu	Neutrophil count
Lym	Lymphocyte count
Mon	Monocyte count
PLT	Platelet count
FIB	Fibrinogen
Cr	Creatinine
BUN	Blood urea nitrogen
TC	Total cholesterol
LR	Logistic regression
LASSO	The least absolute shrinkage and selection operator
XGBoost	Extreme gradient boosting
SVM	Support vector machine
RF	Random forest
ROC	Receiver operating characteristic
AUC	Area under the curve
DCA	Decision curve analysis
SE	Standard error
CI	Confidence interval
PPV	Positive predictive value
NPV	Negative predictive value

Acknowledgements

The authors thank all patients who participated in the study.

Author contributions

Xiaorui Wang and Song Luo collected the data and were major contributors in writing the manuscript; Yujie Zhao analyzed and interpreted the patient data; Xue Cui, Hongdang Qu and Qirong Liao analyzes the data; Song Luo designed the work and revised it. All authors reviewed and approved the final manuscript.

Funding

The study was supported by the Clinical Medicine Research and Translational Project of Anhui Province (202304295107020076), the University Natural Science Research Project of Anhui Province (2022AH051480), the Bengbu Science and Technology Innovation Guidance Project (20230131), the Anhui Province Jianghuai Famous Doctor Cultivation Project (2022). The authors declare that they have no financial relationship with the organization that sponsored the research, and the funding body was not involved in study design, data collection, analysis and writing of the study.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the First Affiliated Hospital of Bengbu Medical University. Written informed consent was obtained from all participants included in the study. All procedures performed in the study involving human participants were in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 5 March 2024 / Accepted: 29 July 2024

Published online: 26 August 2024

References

- Saini V, Guada L, Yavagal DR. Global Epidemiology of Stroke and Access to acute ischemic stroke interventions. *Neurology*. 2021;97(20 Suppl 2):S6–16.
- Leng T, Xiong ZG. Treatment for ischemic stroke: from thrombolysis to thrombectomy and remaining challenges. *Brain Circ*. 2019;5(1):8–11.
- Tu WJ, Wang LD. Special Writing Group of China Stroke Surveillance Report. China stroke surveillance report 2021. *Mil Med Res*. 2023;10(1):33. 涂文杰、王立德,《中国脑卒中监测报告》特约撰稿组。中国卒中监测报告. *Mil Med Res*. 2023;10(1):33.
- Herpich F, Rincon F. Management of Acute ischemic stroke. *Crit Care Med*. 2020;48(11):1654–63.
- Tseng YJ, Hu RF, Lee ST, et al. Risk factors Associated with outcomes of recombinant tissue plasminogen activator therapy in patients with Acute ischemic stroke. *Int J Environ Res Public Health*. 2020;17(2):618.
- Parise P, Agnelli G. Thrombus resistance to lysis and reocclusion after thrombolysis: the role of platelets. *Blood Coagul Fibrinolysis*. 1991;2(6):749–58.
- Mengozi L, Barison I, Malý M, et al. Neutrophil Extracellular traps and Thrombolysis Resistance: New insights for Targeting therapies. *Stroke*. 2024;55(4):963–71.
- Gao L, Li Z, Yuan Z, et al. Major intracranial arterial stenosis influence association between baseline blood pressure and clinical outcomes after thrombolysis in ischemic stroke patients. *Brain Behav*. 2023;13(6):e3022.
- Sung PH, Chen KH, Lin HS, Chu CH, Chiang JY, Yip HK. The correlation between severity of neurological impairment and left ventricular function in patients after Acute ischemic stroke. *J Clin Med*. 2019;8(2):190.
- Hubert V, Hristovska I, Karpati S, et al. Multimodal Imaging with NanoGd reveals Spatiotemporal features of Neuroinflammation after experimental stroke. *Adv Sci (Weinh)*. 2021;8(17):e2101433.
- Aslam N, Khan IU, Bashamakh A, et al. Multiple sclerosis diagnosis using machine learning and deep learning: challenges and opportunities. *Sens (Basel)*. 2022;22(20):7856.
- Chen SD, You J, Yang XM, et al. Machine learning is an effective method to predict the 90-day prognosis of patients with transient ischemic attack and minor stroke. *BMC Med Res Methodol*. 2022;22(1):195.
- Chinese Society of Neurology, Chinese Stroke Society. Chinese guidelines for diagnosis and treatment of acute ischemic stroke 2018. *Chin J Neurol*. 2018;51:666–82.
- Elsaid AF, Fahmi RM, Shehta N, Ramadan BM. Machine learning approach for hemorrhagic transformation prediction: capturing predictors' interaction. *Front Neurol*. 2022;13:951401.
- Brundel BJM, Ai X, Hills MT, et al. Atrial fibrillation. *Nat Rev Dis Primers*. 2022;8(1):21.
- Kwah LK, Diong J. National Institutes of Health Stroke Scale (NIHSS). *J Physiother*. 2014;60(1):61.
- Agarwal S, Scher E, Lord A, et al. Redefined measure of early neurological improvement shows Treatment Benefit of Alteplase over Placebo. *Stroke*. 2020;51(4):1226–30.
- Abe D, Inaji M, Hase T, et al. A Prehospital Triage System to detect traumatic intracranial hemorrhage using machine learning algorithms. *JAMA Netw Open*. 2022;5(6):e2216393.
- Pirruccello JP, Lin H, Khurshid S, et al. Development of a prediction model for ascending aortic diameter among asymptomatic individuals. *JAMA*. 2022;328(19):1935–44.
- Lei Z, Li S, Feng H, et al. Effects of intravenous rtPA in patients with minor stroke. *Ann Med*. 2024;56(1):2304653.
- Gong P, Liu Y, Gong Y, 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*. 2021;18(1):51.
- Wang L, Cheng Q, Hu T, et al. Impact of stress hyperglycemia on early neurological deterioration in Acute ischemic stroke patients treated with intravenous thrombolysis. *Front Neurol*. 2022;13:870872.
- Kharitonova T, Mikulik R, Roine RO, et al. Association of Early National Institutes of Health Stroke Scale improvement with vessel recanalization and functional outcome after intravenous thrombolysis in ischemic stroke. *Stroke*. 2011;42(6):1638–43.
- Rudilosso S, Urta X, Amaro S, et al. Timing and relevance of clinical improvement after mechanical thrombectomy in patients with Acute ischemic stroke. *Stroke*. 2019;50(6):1467–72.
- Meng Z, Wang M, Guo S, et al. Development and validation of a LASSO Prediction Model for Better Identification of ischemic stroke: a case-control study in China. *Front Aging Neurosci*. 2021;13:630437.
- Yang WX, Wang FF, Pan YY, et al. Comparison of ischemic stroke diagnosis models based on machine learning. *Front Neurol*. 2022;13:1014346.
- Xia C, Wang X, Lindley RI, et al. Combined utility of blood glucose and white blood cell in predicting outcome after acute ischemic stroke: the ENCHANTED trial. *Clin Neurol Neurosurg*. 2020;198:106254.
- Benn M, Emanuelsson F, Tybjaerg-Hansen A, et al. Impact of high glucose levels and glucose lowering on risk of ischaemic stroke: a mendelian randomisation study and meta-analysis. *Diabetologia*. 2021;64(7):1492–503.
- Peng R, Liu K, Li W, et al. Blood urea nitrogen, blood urea nitrogen to creatinine ratio and incident stroke: the Dongfeng-Tongji cohort. *Atherosclerosis*. 2021;333:1–8.
- Cai W, Liu S, Hu M, et al. Functional dynamics of neutrophils after ischemic stroke. *Transl Stroke Res*. 2020;11(1):108–21.
- Quan K, Wang A, Zhang X, Wang Y. Leukocyte Count and adverse clinical outcomes in Acute ischemic stroke patients. *Front Neurol*. 2019;10:1240.
- Lee J, Park KM, Park S. Interpretable machine learning for prediction of clinical outcomes in acute ischemic stroke. *Front Neurol*. 2023;14:1234046.
- Meinel TR, Lerch C, Fischer U, et al. Multivariable prediction model for futile recanalization therapies in patients with Acute ischemic stroke. *Neurology*. 2022;99(10):e1009–18.
- Hu D, Ding C, Jiang X, et al. Elevated Levels of Inflammation Markers Predict Poor Outcomes in Acute ischemic stroke patients after intravenous thrombolysis. *J Stroke Cerebrovasc Dis*. 2021;30(3):105587.
- Aronson D, Mittleman MA, Burger AJ. Elevated blood urea nitrogen level as a predictor of mortality in patients admitted for decompensated heart failure. *Am J Med*. 2004;116(7):466–73.
- Baizabal-Carvallo JF, Alonso-Juarez M, Samson Y. Clinical deterioration following middle cerebral artery hemodynamic changes after intravenous thrombolysis for acute ischemic stroke. *J Stroke Cerebrovasc Dis*. 2014;23(2):254–8.
- Denorme F, Portier I, Kosaka Y, Campbell RA. Hyperglycemia exacerbates ischemic stroke outcome independent of platelet glucose uptake. *J Thromb Haemost*. 2021;19(2):536–46.
- Kim DH, Lee DS, Nah HW, et al. Clinical and radiological factors associated with unfavorable outcome after intravenous thrombolysis in patients with mild ischemic stroke. *BMC Neurol*. 2018;18(1):30.
- Yang H, Lv Z, Wang W, et al. Machine Learning models for Predicting Early neurological deterioration and risk classification of Acute ischemic stroke. *Clin Appl Thromb Hemost*. 2023;29:10760296231221738.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.