

Comparative Accuracy of Different Machine Learning Models in Predicting Early Neurological Deterioration after Intravenous rt-PA Thrombolysis in Patients with Acute Ischaemic Stroke

Yi Ba, Xiaolin Xu, Hong'an Gao and Chengguang Song

Department of Neurology, Benxi Central Hospital, Benxi, China

ABSTRACT

Objective: To compare six machine learning models for predicting early neurological deterioration (END) after intravenous rt-PA thrombolysis in acute ischaemic stroke, and to develop an interpretable clinical tool.

Study Design: Observational study.

Place and Duration of the Study: Department of Neurology, Benxi Central Hospital, Benxi, China, from January 2021 to December 2023.

Methodology: All consecutive adults receiving standard-dose rt-PA within 4.5 hours of onset were screened. END was defined as an increase in the National Institutes of Health Stroke Scale (NIHSS) score of ≥ 4 or death within 24 hours. Thirty-two baseline variables were collected; those showing $p < 0.10$ on univariate analysis (NIHSS, age, fibrinogen, and hypertension) entered model construction. An 80:20 stratified split produced training and validation cohorts. Decision tree, random forest, XGBoost, support vector classifier, multilayer perceptron, and logistic regression were tuned by grid search with fivefold cross-validation. Discrimination (area under the ROC curve and AUC), accuracy, sensitivity, specificity, and F1 score were calculated on the hold-out set. The best model underwent SHapley Additive exPlanation (SHAP) analysis to visualise feature important and protective or harmful thresholds. Internal robustness was confirmed with 1,000 bootstrap resamples.

Results: Among 209 eligible patients (END = 16, 7.7%), the XGBoost model achieved the highest discrimination (AUC 0.966), perfect sensitivity (1.000), accuracy (0.905), and specificity (0.897). The decision tree produced the top F1 score (0.750) but lower AUC (0.957). SHAP plots identified admission NIHSS, hypertension, age ≥ 72 years, and fibrinogen > 3.2 g/L as the principal drivers of risk, together accounting for 85 % of model weight.

Conclusion: A concise, four-variable XGBoost model reliably stratifies END risk after rt-PA, offering a transparent decision aid for clinicians to allocate intensified monitoring or adjunctive therapy.

Key Words: Machine learning, Stroke, Intravenous thrombolysis, rt-PA, Early neurological deterioration.

How to cite this article: Ba Y, Xu X, Gao H, Song C. Comparative Accuracy of Different Machine Learning Models in Predicting Early Neurological Deterioration after Intravenous rt-PA Thrombolysis in Patients with Acute Ischaemic Stroke. *J Coll Physicians Surg Pak* 2025; **35(11)**:1396-1401.

INTRODUCTION

Acute ischaemic stroke (AIS) is a major cause of high disability and mortality worldwide.^{1,2} Intravenous rt-PA (recombinant tissue plasminogen activator) thrombolysis is a pivotal recanalisation treatment during the acute phase of AIS and can improve patients' neurological outcomes to a certain extent.^{3,4}

However, in clinical practice, some patients experience early neurological deterioration (END) within 24 hours after thrombolysis, which is manifested as worsening of neurological deficits relative to baseline or death.⁵ Previous literature has indicated that the occurrence of END is closely associated with poorer functional outcomes and a higher risk of death; therefore, rapid and accurate identification of high-risk patients before or during the early stages of thrombolysis has become a key focus in clinical practice.^{6,7}

Building on traditional statistical analyses or clinical scoring systems (e.g., NIHSS, mRS, age), an increasing number of studies in recent years have begun to explore the value of machine learning algorithms in predicting stroke outcomes.⁸ Compared with traditional linear models, machine learning techniques are more able to capture non-linear relationships and interactions among variables, with a certain level of predictive performance even with relatively small, hetero-

Correspondence to: Dr. Chengguang Song, Department of Neurology, Benxi Central Hospital, Benxi, Liaoning, China
E-mail: aojiangzad271722@163.com

Received: May 29, 2025; Revised: August 21, 2025;

Accepted: September 27, 2025

DOI: <https://doi.org/10.29271/jcpsp.2025.11.1396>

geneous clinical datasets.^{9,10} However, the black box nature of machine learning algorithms often leads clinicians to remain cautious about their interpretability and reliability. Therefore, improving model interpretability while ensuring predictive performance has become an important goal in translational clinical research.

By comparing the differences in predictive performance and interpretability among the models, this study aimed to provide a basis for the rapid identification of high-risk patients in clinical practice, reduce missed diagnoses, and enhance the monitoring and intervention for high-risk populations, ultimately improving the overall patient outcomes.

METHODOLOGY

This study is a retrospective analysis that included patients with AIS who underwent intravenous rt-PA thrombolytic therapy in the Department of Neurology, Benxi Central Hospital, Benxi, China, from January 2021 to December 2023. The inclusion criteria included: confirmation of AIS by imaging (CT/MRI) and clinical diagnosis, meeting the criteria for rt-PA thrombolysis and completing the treatment, and the availability of complete clinical data. The exclusion criteria were patients who, upon admission, had other severe systemic or neurological diseases, and cases with missing baseline data or incomplete key variables.

Based on the previous literature¹¹ and clinical feasibility, END was defined as either an increase of ≥ 4 points in the NIHSS score from baseline within 24 hours after thrombolysis or death. Patients meeting these criteria were assigned to the END group, while the remaining patients were classified as the non-END group. Demographic data (e.g., age, gender), admission NIHSS scores, laboratory test indicators, and outcome information were collected for both groups. Data were subjected to routine cleaning, missing-value imputation, one-hot encoding, and z-score normalisation.

The complete dataset was randomly partitioned into training and validation sets at an 80:20 ratio using Python's `train_test_split` function. When computational resources allowed, five-fold or ten-fold cross-validation was further conducted within the training set to obtain more robust hyperparameter estimates. To compare the performance of several common machine learning algorithms in predicting the risk of END, six models were selected, including logistic regression, decision tree classifier, random forest, XGBoost (a representative gradient-boosting tree model), support vector machine (SVC), and neural network—multilayer perceptron (MLP). Within the training set, key hyperparameters (e.g., maximum tree depth, learning rate, regularisation coefficients, and the number of hidden-layer neurons) were tuned by grid search (`GridSearchCV`) or random search (`RandomisedSearchCV`) in combination with cross-validation. Because END cases were relatively scarce, class imbalance was mitigated by adjusting the `class_weight` parameter during training or by applying over-

sampling techniques such as SMOTE to improve learning performance on the imbalanced dataset.

On the validation set, the predictions of each model were compared with the actual group assignments. Primary evaluation metrics included area under the ROC curve (AUC)—overall discriminative ability—sensitivity (recall) and specificity, accuracy, F1 score, precision-recall (P-R) curve, and area under the precision-recall curve (AUPRC), which provide complementary information under class-imbalance conditions. The optimal model was selected by synthesising the above indicators.

After identifying the best-performing model, further visualisation and interpretability analyses were conducted. Confusion matrix was generated to display true positives, false positives, true negatives, and false negatives in the validation set, thereby quantifying misclassification patterns. SHAP interpretability analysis, using Shapley additive explanations, was applied to visualise and explain the contribution and direction of key features, thereby offering clinically relevant insights into high-risk factors.

All data preprocessing, statistical analyses, and model training were carried out in Python 3.10 (Anaconda 2024.02 distribution). Machine learning algorithms were implemented with `scikit-learn` 1.4.0 and `XGBoost` 2.0.3; statistical tests and confidence-interval computations were performed with `SciPy` 1.12.0 and `statsmodels` 0.15.0.

RESULTS

The overall frequency of END in this cohort study was relatively low. After data preprocessing and feature selection, statistical analysis showed significant differences between the END and non-END groups in admission NIHSS score, hypertension, age, and fibrinogen level ($p < 0.05$). These four variables were, therefore, incorporated as predictive features (Table I).

On the test set, six machine learning algorithms were evaluated. Although the decision tree classifier excelled in accuracy, specificity, and F1 score, the primary optimisation metric was AUC; therefore, XGBoost was ultimately selected as the optimal model. The XGBoost model achieved the highest AUC while maintaining a high sensitivity, thereby minimising missed END cases (Figure 1 showed that its ROC curve enclosed the largest area). P-R curves (Figure 2) further indicated that XGBoost and the MLP struck the best balance between precision and recall, a desirable property for datasets with imbalanced positive and negative classes. Taken together, these findings demonstrated that XGBoost offered the strongest overall discriminative capability, effectively balancing detection rate, and false-positive rate. Therefore, it is the most suitable model for predicting END in this study (Table II).

Among the six algorithm models, the XGBoost model achieved the highest AUC on the validation set (0.966) and was, therefore, selected as the optimal predictor. Its key performance metrics were: accuracy 0.905, recall (sensitivity) 1.000, specificity 0.897, precision 0.429, and F1 score 0.600.

Table I: Comparison of the clinical characteristics between the END and non-END groups.

Variables	END Group (n = 16)	Non-END Group (n = 193)	χ^2/t	p-values
TOAST classification				
LAA	12	111	4.578	0.205
CE	2	23		
SVO	1	55		
UD	1	4		
Pneumonia (comorbidity)	6	37	3.038	0.081
Infarct territory				
AC	12	123	1.080	0.583
PC	1	28		
AC+PC	3	42		
Alcohol consumption	5	69	0.013	0.909
Smoking	11	123	0.221	0.638
Moderate-to-severe arterial stenosis	11	113	0.639	0.424
Atrial fibrillation	3	25	0.432	0.511
Diabetes mellitus	3	47	0.257	0.612
Hypertension	14	101	7.399	0.007
Gender (male / female)	11/5	139/54	0.077	0.782
Age (years)	69.312 ± 9.250	63.725 ± 10.386	2.299	0.034
White blood cell count (×10 ⁹ /L)	8.046 ± 2.576	7.799 ± 2.539	0.369	0.717
Platelet count (×10 ⁹ /L)	253.188 ± 136.321	222.193 ± 58.444	0.903	0.38
Direct bilirubin (μmol/L)	2.794 ± 1.518	3.292 ± 2.682	-1.167	0.255
Albumin (g/L)	41.506 ± 2.709	40.783 ± 3.938	0.983	0.337
ALT(U/L)	21.75 ± 12.091	20.743 ± 10.191	0.323	0.75
Alkaline phosphatase (U/L)	90.500 ± 19.980	84.270 ± 22.642	1.185	0.251
Potassium (mmol/L)	4.119 ± 0.513	3.908 ± 0.388	1.606	0.127
Admission NIHSS score	9.812 ± 5.540	6.415 ± 5.152	2.37	0.03
Prothrombin activity (%)	110.375 ± 15.491	107.818 ± 15.066	2.977	0.009
INR	0.956 ± 0.079	0.972 ± 0.090	0.636	0.533
Prothrombin time (s)	12.800 ± 0.849	12.904 ± 0.892	-0.745	0.465
Activated partial thromboplastin time (s)	33.481 ± 3.104	33.662 ± 3.26	-0.47	0.644
Fibrinogen (g/L)	2.932 ± 0.472	3.274 ± 0.909	-2.256	0.018
Thrombin time (s)	17.162 ± 0.823	17.333 ± 1.776	-0.706	0.486
D-dimer (mg/L)	0.792 ± 0.827	0.799 ± 1.637	-0.705	0.487
Creatinine (μmol/L)	77.688 ± 19.12	75.575 ± 20.006	-0.025	0.98
Urea (mmol/L)	6.996 ± 1.819	6.131 ± 1.615	0.423	0.677

AC: Anterior circulation; PC: Posterior circulation; LAA: Large artery atherosclerosis; CE: Cardioembolism; SVO: Small vessel occlusion; UD: Undetermined aetiology.
Note: p-value calculation: continuous variables were compared using an independent-samples t-test (with Welch's correction when variances were unequal) or a Mann-Whitney U test if non-normally distributed; categorical variables were analysed using Pearson's χ^2 test (or Fisher's exact test when any expected cell count <5).

Table II: Comparison of the predictive accuracy on the validation set for six machine learning models.

Models	AUC	Accuracies	Sensitivities	Specificities	Precisions	F1 scores
Logistic regression	0.914	0.738	1.000	0.718	0.214	0.353
Decision tree	0.957	0.952	1.000	0.949	0.600	0.750
Random forest	0.962	0.901	0.667	0.923	0.400	0.500
XGBoost	0.966	0.905	1.000	0.897	0.429	0.600
Support vector machine	0.846	0.833	0.667	0.846	0.250	0.364
Neural network (MLP)	0.957	0.905	0.667	0.923	0.400	0.500

Table III: Confusion matrix of the optimal XGBoost predictive model.

Actual / Predicted	Predicted END (+)	Predicted non-END (-)
Actual END (+)	3 (TP)	0 (TN)
Actual non-END (-)	4 (FP)	35 (TN)

TP: True positive; FP: False positive; TN: True negative.

The optimal hyperparameters were a learning rate of 0.1 and a max_depth of 8. The confusion matrix (Table III) showed FN = 0, indicating no false negatives thereby a 100 % detection rate for END cases. There were four false positives, suggesting that clinical judgment is required to reduce over-alerts. The SHAP plot (Figure 3) provided a global interpretation of the XGBoost model. The beeswarm diagram ranked the top three contributors as admission NIHSS score, hypertension, and age. Admission NIHSS had the highest mean SHAP value, indicating that baseline neurological deficit was the dominant factor influencing END. Hyper-

tension, elevated fibrinogen, and older age contribute positively, further increasing the risk of END.

DISCUSSION

Developing machine learning models to predict END after intravenous rt-PA thrombolysis offers several clinical advantages. First, END generally occurs within 24 hours of thrombolysis and is a pivotal determinant of functional outcome in patients with AIS. Early identification of high-risk individuals during this window enables clinicians to intensify neurological monitoring, repeat neuroimaging promptly, and adjust blood-pressure control, anticoagulation, and other therapeutic strategies, thereby reducing secondary injury.^{12,13} Second, hospital resources for AIS are limited, whereas END is relatively uncommon (≈ 5 %-30 %).

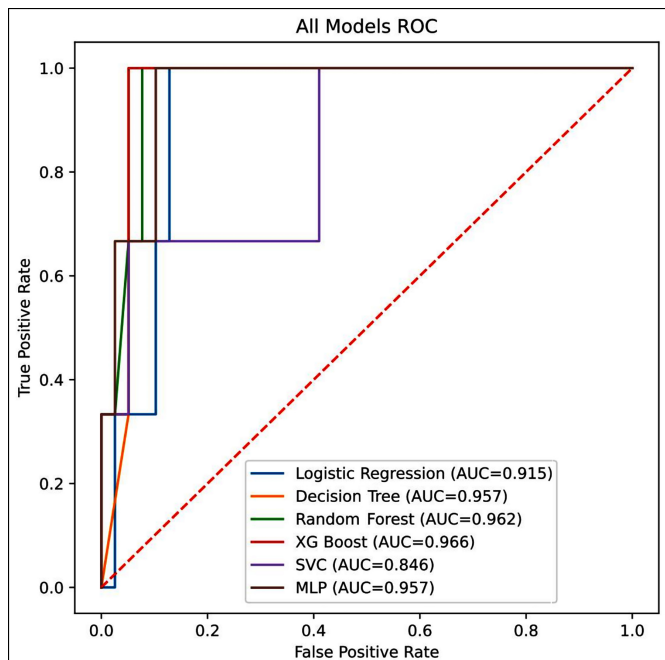


Figure 1: ROC curves for the six machine learning models.

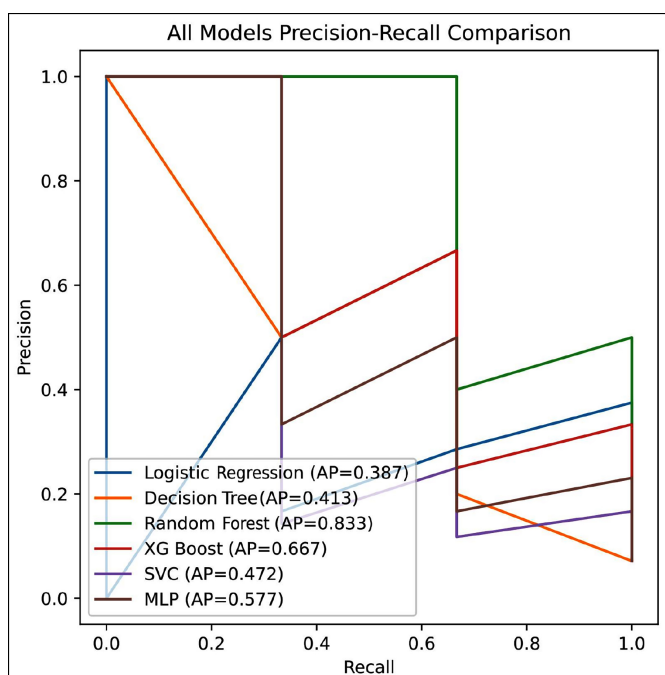


Figure 2: P-R curve comparison of the six models.

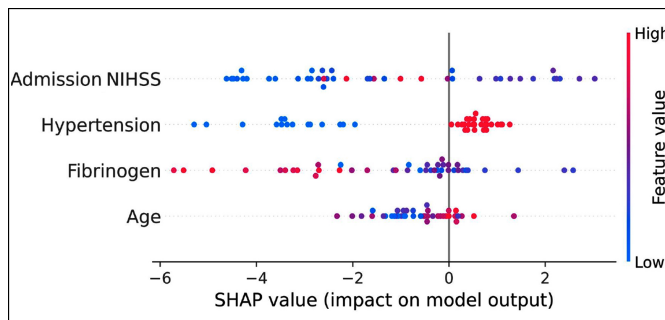


Figure 3: SHAP summary plot of the XGBoost model.

The traditional approach of universally intensive monitoring is therefore inefficient. Risk stratification allows monitoring resources to be concentrated on patients predicted to be at high risk, facilitating precision management.¹⁴ Third, compared with linear regression or single-score systems, machine learning models can process multidimensional clinical data and capture non-linear interactions, offering greater sensitivity and better generalisability for individualised prediction. Interpretability techniques such as SHAP further clarify each feature's relative contribution to outcomes, enhancing the model's acceptability in clinical decision-making.^{15,16} Finally, visual risk-assessment tools improve physician-patient communication, helping patients and their families understand their own risk and adhere to subsequent treatment and rehabilitation plans. Collectively, these benefits may lower disability and mortality rates, shorten hospital stays, and reduce healthcare costs.

Among the six algorithm models in this study, the linear logistic regression model reached 100% sensitivity with only a few parameters, yet its inability to capture non-linear and interaction effects kept precision at just 0.214, resulting in a high false-positive rate. Decision tree models are intuitive and easily interpreted; however, a single-tree structure is prone to over-fitting, so their generalisation performance was unstable. The ensemble random forest lowered variance through bagging and achieved an AUC of 0.962; however, in the presence of class imbalance, it leaned toward the majority class, leaving recall at 0.667. The kernel-based SVC handles high-dimensional features well; however, with the present data size, it required long training times and was extremely sensitive to hyperparameters, yielding an AUC of only 0.846. Although a deep MLP can, in theory, approximate any complex function, it requires much larger samples to avoid overfitting; consequently, its recall likewise remained at 0.667.

By contrast, XGBoost finely fits residuals *via* gradient boosting while introducing L1/L2 regularisation in each iteration to curb overfitting; combined with class-weighting mechanisms such as `scale_pos_weight`, it is naturally suited to minority-class problems, such as END. On the validation set, it delivered the highest AUC (0.966), achieved 100% recall, and limited false positives to four—striking a balance between zero missed diagnoses and an acceptable false-alarm rate. In addition, XGBoost automatically handles missing values, trains rapidly, and integrates seamlessly with SHAP, enabling individualised risk attribution without compromising clinical interpretability. Taken together, its overall performance, computational efficiency, and clinical usability surpassed those of the other models, establishing XGBoost as the optimal predictive tool in this study.

A closer examination of XGBoost's performance in the present setting shows that its advantages extend beyond leading metric values; they arise from a strong match between the algorithm's design and the characteristics of

clinical data. First, XGBoost employs a gradient-boosting framework that chains a series of weak decision trees in a residual-learning sequence, with each new tree fitting the hard samples left by the previous model. This layer-by-layer error-correction mechanism continually strengthens the recognition of the minority END class in a task that is highly imbalanced and limited in sample size, whereas bagging methods tend to dilute learning focus.^{17,18} Second, the objective function incorporates both first- and second-order gradients and embeds L1/L2 regularisation, accelerating convergence while effectively preventing overfitting. As a result, an AUC of 0.966 and 100% recall were achieved on a dataset of only 209 patients.¹⁹ Third, XGBoost's automatic missing-value branching strategy allows each tree to assign a default direction for missing records, avoiding information loss and imputation error that are common in clinical data. Combined with parameters—such as `scale_pos_weight`, the model maintains zero missed diagnoses while keeping false positives within an acceptable range.^{20,21} Fourth, built-in column subsampling and parallel computation make training far faster than traditional boosting or deep networks, facilitating rapid model updates within real-time clinical workflows. Its incremental learning capability also supports continual fine-tuning as new patient data accrue, preserving model freshness.^{22,23} Crucially, the tree structure integrates seamlessly with SHAP, enabling global and individual quantification of each feature's marginal contribution and dispelling black-box concerns. The high-impact features identified (admission NIHSS score, age, fibrinogen, and hypertension) coincide with established stroke risk factors, further enhancing the model's credibility and generalisability.^{24,25} In sum, XGBoost surpasses the other models across four dimensions—algorithmic mechanism, data adaptability, computational efficiency, and interpretability—making it the ideal choice for post-thrombolysis END risk stratification and laying a solid foundation for future clinical decision-support systems that integrate multimodal data such as imaging and genomics.

Although XGBoost achieved the best performance in this study, its use still has notable limitations. Being based on single-centre, retrospective data with a limited sample size, the model is vulnerable to selection bias. The feature set was restricted—only admission-time clinical variables were included; imaging, haemodynamic, and genetic data were not integrated, leaving some END-triggering mechanisms under-represented. Numerous hyperparameters also require manual tuning, potentially adding maintenance overhead during clinical deployment. Interpretability remains incomplete: while SHAP helps, it offers limited insight into temporal patterns or latent causal relationships.

Future work should validate generalisability in large, multi-center, and prospective cohorts, streamline updates via automated hyperparameter optimisation and incremental learning, and incorporate multimodal data and causal-inference frame-

works to further improve predictive accuracy, interpretability, and clinical usability.

CONCLUSION

Using four key admission variables—NIHSS score, age, fibrinogen level, and hypertension—for predicting END after rt-PA thrombolysis, XGBoost proved to be the optimal model, achieving the highest AUC (0.966), 100 % recall, and strong interpretability. It thus offers a reliable tool for early risk stratification and precise allocation of monitoring resources in rt-PA-treated patients, while also laying a methodological foundation for multimodal, continuously updated stroke-alert systems.

FUNDING:

This study was funded by Natural Science Foundation of Liaoning Province (2021-MS-378), Guidance Project of the Benxi Key Research and Development Programme (2023ZD JH005), and Intramural Research Project of Benxi Central Hospital (YN202307).

COMPETING INTEREST:

The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:

YB: Conceived and designed the study, supervised data acquisition, performed radiomics feature extraction and statistical analyses, and drafted the manuscript.

XX, HG: Curated the clinical and imaging dataset, assisted with radiomics pipeline implementation and interpretation of results.

CS: Provided methodological expertise for model construction and validation, conducted independent statistical checks and critically revised the manuscript.

All authors approved the final version of the manuscript to be published.

REFERENCES

1. AlRuwalli R, Al-Kuraishy HM, Alruwalli M, Khalifa AK, Alexiou A, Papadakis M, *et al.* The potential therapeutic effect of phosphodiesterase 5 inhibitors in the acute ischemic stroke (AIS). *Mol Cell Biochem* 2024; **479**(5):1267-78. doi: 10.1007/s11010-023-04793-1.
2. Anand SK, Benjamin WJ, Adapa AR, Park JV, Wilkinson DA, Daou BJ, *et al.* Trends in acute ischemic stroke treatments and mortality in the United States from 2012 to 2018. *Neurosurg Focus* 2021; **51**(1):E2. doi: 10.3171/2021.4.FOCUS21117.
3. Zhang S, Wang D, Li L. Recombinant tissue-type plasminogen activator (rt-PA) effectively restores neurological function and improves prognosis in acute ischemic stroke. *Am J Transl Res* 2023; **15**(5):3460-7.
4. Wu H, Liu Y, Miao G, Ge J, You S, Zhang X, *et al.* Effect of the interaction between atrial fibrillation and rt-PA dose on the prognosis of acute ischaemic stroke with intravenous thrombolysis. *Postgrad Med J* 2023; **99**(1172):588-94. doi: 10.1136/postgradmedj-2022-141710.

5. Xie X, Xiao J, Wang Y, Pan L, Ma J, Deng L, et al. Predictive model of early neurological deterioration in patients with acute ischemic stroke: A retrospective cohort study. *J Stroke Cerebrovasc Dis* 2021; **30(3)**:105459. doi: 10.1016/j.jstrokecerebrovasdis.2020.105459.
6. Tian T, Wang L, Xu J, Jia Y, Xue K, Huang S, et al. Prediction of early neurological deterioration in acute ischemic stroke patients treated with intravenous thrombolysis. *J Cereb Blood Flow Metab* 2023; **43(12)**:2049-59. doi: 10.1177/0271678X231200117.
7. Wang J, Zhang X, Tian J, Li H, Tang H, Yang C. Predictive values of systemic inflammatory responses index in early neurological deterioration in patients with acute ischemic stroke. *J Integr Neurosci* 2022; **21(3)**:94. doi: 10.31083/jjin2103094.
8. Lee M, Yeo NY, Ahn HJ, Lim JS, Kim Y, Lee SH, et al. Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning. *Alzheimers Res Ther* 2023; **15(1)**:147. doi: 10.1186/s13195-023-01289-4.
9. Wang X, Lyu J, Meng Z, Wu X, Chen W, Wang G, et al. Small vessel disease burden predicts functional outcomes in patients with acute ischemic stroke using machine learning. *CNS Neurosci Ther* 2023; **29(4)**:1024-33. doi: 10.1111/cns.14071.
10. Huang Q, Shou GL, Shi B, Li ML, Zhang S, Han M, et al. Machine learning is an effective method to predict the 3-month prognosis of patients with acute ischemic stroke. *Front Neurol* 2024; **15**:1407152. doi: 10.3389/fneur.2024.1407152.
11. Yu WM, Abdul-Rahim AH, Cameron AC, Korv J, Sevcik P, Toni D, et al. The incidence and associated factors of early neurological deterioration after thrombolysis: Results from SITS registry. *Stroke* 2020; **51(9)**:2705-14. doi: 10.1161/STROKEAHA.119.028287.
12. Shi HX, Li C, Zhang YQ, Li X, Liu AF, Liu YE, et al. Predictors of early neurological deterioration occurring within 24 hour in acute ischemic stroke following reperfusion therapy: A systematic review and meta-analysis. *J Integr Neurosci* 2023; **22(2)**:52. doi: 10.31083/jjin2202052.
13. Egashira S, Koga M, Toyoda K. Intravenous thrombolysis for acute ischemic stroke in patients with end-stage renal disease on hemodialysis: A narrative review. *J Cardiovasc Dev Dis* 2022; **9(12)**:446. doi: 10.3390/jcdd9120446.
14. Zhang X, Zhong W, Xue R, Jin H, Gong X, Huang Y, et al. Argatroban in patients with acute ischemic stroke with early neurological deterioration: A randomized clinical trial. *JAMA Neurol* 2024; **81(2)**:118-25. doi: 10.1001/jamaneurol.2023.5093.
15. Lu J, Zhou Y, Lv W, Zhu H, Tian T, Yan S, et al. Identification of early invisible acute ischemic stroke in non-contrast computed tomography using two-stage deep-learning model. *Theranostics* 2022; **12(12)**:5564-73. doi: 10.7150/thno.74125.
16. Jo H, Kim C, Gwon D, Lee J, Lee J, Park KM, et al. Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: An automated machine learning approach. *Scie Rep* 2023; **13(1)**:16926. doi: 10.1038/s41598-023-44201-8.
17. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Inter Learn Environ* 2023; **31(3)**:1-20. doi: 10.1080/10494820.2021.1928235.
18. Zhu X, Chu J, Wang K, Wu S, Yan W, Chiam K. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. *J Rock Mech Geotech Eng* 2021; **13(6)**:1231-45. doi: 10.1016/j.jrmge.2021.06.012.
19. Ge J, Zhao L, Yu Z, Liu H, Zhang L, Gong X, et al. Prediction of greenhouse tomato crop evapotranspiration using XGBoost machine learning model. *Plants (Basel)* 2022; **11(15)**:1923. doi: 10.3390/plants11151923.
20. Natras R, Soja B, Schmidt M. Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting. *Remote Sens* 2022; **14(15)**:3547. doi: 10.3390/rs14153547.
21. Zelli V, Manno A, Compagnoni C, Ibraheem RO, Zazzeroni F, Alesse E, et al. Classification of tumor types using XGBoost machine learning model: A vector space transformation of genomic alterations. *J Transl Med* 2023; **21(1)**:836. doi: 10.1186/s12967-023-04720-4.
22. Kavzoglu T, Teke A. Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arab J Sci Eng* 2022; **47(6)**:7367-85. doi: 10.1007/s13369-022-06560-8.
23. Palsa J, Adam N, Hurtuk J, Chovancov A E, Mados B, Chovanec M, et al. MLMD—A malware-detecting antivirus tool based on the XGBoost machine learning algorithm. *Appl Sci* 2022; **12(13)**:6672. doi: 10.3390/app12136672.
24. Noviany TR, Idroes GM, Maulana A, Hardi I, Ringga ES, Idroes R. Credit card fraud detection for contemporary financial management using XGBoost-driven machine learning and data augmentation techniques. *Indatu J Manag Account* 2023; **1(1)**:29-35. doi: 10.60084/ijma.v1i1.78.
25. Maulana A, Faisal FR, Noviany TR, Rizkia T, Idroes GM, Tallei T, et al. Machine learning approach for diabetes detection using fine-tuned XGBoost algorithm. *Infolitika J Data Sci* 2023; **1(1)**:2023. doi: 10.60084/ijds.v1i1.72.

• • • • •