Advances in Prognosis Prediction for Acute Ischemic Stroke: A Case Study of Machine Learning Prediction Models (Postprint)

Authors: Du Huijie, Liu Xingyu, Xu Minghuan, Yang Xuezhi, Zhang Huiqin, Mo Jiali, Lu Yi, Jie Kuang, Kuang Jie

Date: 2024-05-27T00:00:00+00:00

Abstract

Acute ischemic stroke (AIS) is characterized by high rates of disability, mortality, and recurrence, imposing a substantial burden on patients and society. With the advent of the big data era, prediction models are increasingly being applied in clinical decision-making, prognosis management, and health resource allocation, rendering their value ever more critical. Machine learning methods constitute one of the important approaches for predicting outcomes in AIS patients and have been widely adopted. This review focuses on machine learning methods, summarizes the latest advances in AIS outcome prediction research, and outlines the current problems and challenges confronting machine learning prediction models, thereby providing novel insights and methodological references for the early assessment and prediction of outcomes in AIS patients.

Full Text

Advances in Prognostic Prediction of Acute Ischemic Stroke: Using Machine Learning Predictive Models as an Example

DU Huijie, LIU Xingyu, XU Minghuan, YANG Xuezhi, ZHANG Huiqin, MO Jiali, LU Yi, KUANG Jie*

Department of Epidemiology, School of Public Health, Jiangxi Medical College, Nanchang University; Jiangxi Provincial Key Laboratory of Preventive Medicine, Jiangxi Medical College, Nanchang University, Nanchang 330006, China

*Corresponding author: KUANG Jie, Associate professor; E-mail: kuangjie@ncu.edu.cn

DU Huijie and LIU Xingyu are co-first authors

Abstract

Acute ischemic stroke (AIS) is characterized by high rates of disability, mortality, and recurrence, posing a significant burden on patients and society. In the era of big data, predictive models are increasingly used in patient diagnosis, treatment decisions, prognosis management, and healthcare resource allocation, highlighting their growing importance. Machine learning methods have become a crucial tool for predicting the prognosis of AIS patients and have been widely applied. This review explores recent advancements in the study of AIS prognosis prediction, focusing on machine learning methods. It discusses current issues and challenges faced by machine learning models, aiming to provide new insights and references for methods of early assessment and prediction of prognosis outcomes in AIS patients.

Keywords: Ischemic stroke; Prognosis prediction; Machine learning; Prediction model; Review

Stroke is the second leading cause of death and a major cause of disability worldwide [1], representing a significant global health concern. In China, the crude mortality rate of stroke has shown a continuous upward trend, with the rate of increase far exceeding that of other countries [2]. Acute ischemic stroke (AIS) is the most common type of stroke, accounting for approximately 80% of all strokes, and is characterized by high rates of disability, mortality, and recurrence [3], imposing a heavy disease burden on patients and society. Therefore, accurate prediction of prognosis in AIS patients is crucial for enabling early individualized interventions and rehabilitation treatments in clinical practice. With the advent of the big data era, massive amounts of clinical data have accumulated in the medical field, and traditional prediction methods have become inadequate to meet clinical needs. Machine learning (ML) methods, with their advantages in handling complex data, have been widely applied in healthcare [4]. This review focuses on ML methods to summarize the latest advances in AIS prognosis prediction research, aiming to provide new ideas and references for methodological approaches to early assessment and prediction of prognosis outcomes in AIS patients.

1 Literature Search Strategy

We searched PubMed, Web of Science, Wanfang Data Knowledge Service Platform, and China National Knowledge Infrastructure (CNKI) databases from inception to December 2023. Chinese search terms included "ischemic stroke," "prognosis prediction," and "machine learning," while English search terms included "acute ischemic stroke," "prognosis prediction," and "machine learning." Inclusion criteria: literature related to ML-based AIS prognosis studies, focusing on representative publications from the past three years. Exclusion criteria: literature unrelated to the topic, unavailable full text, or poor quality. A total of 60 articles were finally included.

2 AIS Prognosis Research

Prognostic prediction typically refers to predicting the risk and likelihood of a specific health condition occurring within a defined time period based on patients' clinical or non-clinical characteristics [5]. AIS prognosis research involves prognostic assessment of AIS patients to determine their functional outcomes and survival status, aiming to identify factors associated with poor prognosis and develop prediction models to promptly detect patients at high risk of adverse functional outcomes. AIS is characterized by high rates of disability, mortality, and recurrence [3], and timely prognosis prediction helps clinicians implement targeted therapeutic measures, intervene on preventable risk factors, promote patient recovery, and reduce poor outcomes.

Traditional AIS prognosis studies have frequently used scoring scales to evaluate patient prognosis, such as the ASTRAL (Acute Stroke Registry and Analysis of Lausanne) score, THRIVE (Totaled Health Risks in Vascular Events) score, and IScore (Ischemic Stroke Risk Prediction Score) [10-11]. While these scales have achieved certain effectiveness in AIS prognosis prediction, they have limitations [12]. First, scoring scales often include only simple indicators such as admission status and medical history, without collecting post-treatment data, resulting in insufficient predictive accuracy. For example, the THRIVE score comprises only clinical data collected at admission; although convenient to obtain, it lacks corresponding imaging assessments and laboratory tests, which reduces its predictive performance to some extent. Second, AIS prognosis is influenced by multiple factors, including demographic characteristics, cerebral perfusion status, inflammatory response, and medication effects, which collectively affect patient outcomes. However, scoring scales often cannot explain the interactions between these factors.

3 Overview of ML Methods

ML is a branch of artificial intelligence (AI). AI pioneer Arthur Samuel defined ML as a class of methods that give computers the ability to learn without being explicitly programmed [13]. ML aims to explore the intersection of statistical data analysis and computer algorithms [14]. By learning and extracting patterns from massive datasets, it can produce reliable and reproducible decisions, demonstrating good applicability in tasks related to high-dimensional data such as classification, regression, and clustering [15].

ML applications in medicine mainly include: (1) Disease risk prediction. ML algorithms predict disease risk by learning from patients' clinical test data, helping doctors intervene and treat diseases early, reducing patient risk, lowering healthcare costs, and playing an important role in disease prevention and control [16-18]. (2) Assisted diagnosis. ML can automate the interpretation of clinical test results, saving time and effectively improving diagnostic accuracy. With advantages of convenience and repeatability, it has been widely used for clinical disease diagnosis [19-21]. (3) Prognosis assessment. ML enables early prediction of patient prognosis, facilitating individualized therapeutic interventions, reducing poor outcomes, and improving patient survival rates and quality of life [22-23].

4 Application of ML Methods in AIS Prognosis Prediction

Prognosis prediction research typically includes variable selection, model construction, and model validation. Traditional model construction methods often use regression analysis, such as Logistic regression [6-7]. While simple, interpretable, and easy to use, these methods have limitations. Regression-based prediction models are often constrained by the number of included variables [8], large-scale data processing capabilities, model complexity, and difficulty in capturing nonlinear relationships between prognostic factors [9]. These deficiencies affect variable selection and predictive performance. With the continuous development of big data in medicine, clinical data have become increasingly complex and diverse, prompting the search for new prediction methods to fully utilize clinical data.

ML algorithms can effectively identify features highly correlated with outcomes from numerous candidate features, with many configurable parameters for optimization, demonstrating excellent performance in handling high-dimensional data and identifying complex interactions between variables [24]. Multiple studies have shown that ML methods achieve better predictive performance than traditional models. Extreme Gradient Boosting (XGBoost) is a tree boosting ensemble algorithm that improves data processing through optimized decision tree algorithms, enhances accuracy through regularization and built-in cross-validation, and effectively addresses overfitting [25]. Tong et al. [26] used XGBoost and traditional logistic models to predict AIS patient prognosis, finding that XGBoost had superior predictive performance. Artificial Neural Networks (ANN) are ML algorithms that simulate signal transmission between neurons in the human brain for information analysis and processing, consisting of interconnected processing units including input, hidden, and output layers that analyze intrinsic relationships between data through complex signal pathway structures [27]. Deep Neural Network (DNN) models represent a deeper application of ANN, benefiting from hierarchical structures that allow reuse of features computed in given hidden layers at higher layers, improving function approximation accuracy within fixed parameter budgets and enhancing generalization ability after learning new examples [28]. Heo et al. [29] used DNN to predict AIS patient prognosis and compared it with the ASTRAL score, showing that DNN' s predictive ability was significantly higher than ASTRAL, demonstrating that ML models have superior performance in predicting long-term outcomes of AIS patients. Similarly, other ML models have shown good predictive performance in AIS prognosis prediction. Jiang et al. [30] constructed ML models using multiple clinical and imaging features and compared them with the SPAN-100 (Stroke Prognostication using Age and NIHSS-100) index model, finding that ML models could more accurately predict AIS patient prognosis. The SPAN-

100 index has proven effective in predicting patient prognosis and complication risks after vascular therapy, but its main limitation is inapplicability to younger patients. The ML model in this study overcame this limitation, being applicable to any AIS patient over 18 years old, thus expanding its scope. Additionally, commonly used ML methods include Support Vector Machine (SVM) [31], Random Forest (RF) [32], and deep learning [33-34], all playing important roles in AIS prognosis prediction.

4.1 ML Methods for Predicting AIS Functional Outcomes Between 30% and 70% of stroke survivors live with disability, directly or indirectly affecting the quality of life of AIS patients [35]. Early identification of poor functional outcomes in AIS patients and timely intervention can effectively reduce disability burden and improve quality of life. Researchers have increasingly focused on using ML methods to predict functional prognosis in AIS patients.

Park et al. [36] used multiple ML models to predict three-month functional outcomes after AIS, with the SVM model achieving an AUC of 0.850 and highest F1 score of 0.860, demonstrating good predictive performance. The proposed model was based on initial admission assessments and test results, enabling prognosis prediction shortly after hospitalization. However, it lacked more detailed post-admission clinical data, which may affect model performance. In a retrospective study, Jo et al. [37] used clinical models, convolutional neural network models incorporating imaging features, and ensemble models combining imaging and clinical features to predict three-month functional outcomes in AIS patients. Results showed that ensemble models had superior predictive performance, identifying patient age, NIHSS score, and early neurological deterioration as risk factors affecting AIS functional outcomes. Ozkara et al. [38] constructed ML models based on clinical, laboratory, and imaging data to predict short- and medium-term functional outcomes in AIS patients with proximal middle cerebral artery occlusion, finding that the LightGBM (Light Gradient Boosting Machine) model performed best with an AUC of 0.958. Additionally, studies have found that RF models have better predictive performance than traditional models such as HIAT (Houston Intra-Arterial Therapy) score, THRIVE score, and NADE nomogram in predicting six-month functional outcomes in AIS patients [39]. Lee et al. [40] reported that XGBoost showed optimal performance in predicting three-month functional outcomes after AIS, identifying initial NIHSS score, early neurological deterioration, age, and white blood cell count as key predictive factors. However, most current studies on post-AIS functional outcomes are limited to short-term predictions, and future research should focus more on long-term functional prognosis to identify high-risk patients in a timely manner.

4.2 ML Methods for Predicting AIS Mortality Risk Stroke has become the second leading cause of death in rural areas and the third in urban areas in China, imposing a heavy disease burden on the nation and society [41]. Wang et al. [42] developed multiple ML prediction models for one-year mortality risk in

AIS patients, finding that XGBoost had the best predictive performance, with C-reactive protein (CRP), homocysteine (Hcy) levels, stroke severity, and stroke lesion number being independent risk factors for one-year mortality. Similarly, another study constructing a neural network model based on clinical, imaging, and biomarker data to predict one-year mortality after stroke reported good predictive performance, identifying immune-inflammatory factors (interleukin-6, tumor necrosis factor-α, interleukin-10, transforming growth factor-β, etc.) and coagulation biomarkers (coagulation factor VIII, von Willebrand factor, and fibrinogen, etc.) as predictors of one-year mortality after ischemic stroke, which helps improve model performance [43], providing reference for clinical indicators and treatment planning. Additionally, among AIS subtypes, small vessel occlusion (SVO) carries higher mortality risk when receiving anticoagulation therapy than other subtypes. Zhang et al. [44] constructed ML models with typical blood characteristics for early etiological diagnosis of AIS patients, showing that ML models performed well in AIS subtype diagnosis, with the ML model for SVO subtype diagnosis achieving an AUC of 0.780, helping to identify AIS subtypes early and provide targeted treatment to reduce mortality risk.

4.3 ML Methods for Predicting Stroke Recurrence Risk Stroke is highly recurrent, with a global lifetime risk of approximately 25% [45]. Identifying high-risk AIS recurrence populations and providing timely diagnosis and treatment can reduce disability and mortality rates and improve quality of life. ML methods help predict high-risk AIS recurrence populations. Wang et al. [8] established ML prediction models for one-year stroke recurrence risk in AIS patients, finding that RF models had the best predictive performance with an AUC of 0.946. The study also identified right hemisphere lateralization, Hcy, CRP levels, and stroke severity as independent risk factors for stroke recurrence in AIS patients. Additionally, in a retrospective study [46], researchers used an unsupervised Gaussian mixture model with 92 biomarkers from different sources (including blood components, coagulation function, liver and kidney function, inflammatory factors, etc.) to classify AIS patients into four clinical phenotypes with different characteristics, finding that phenotype 2, characterized by inflammation and renal dysfunction, had the highest stroke recurrence risk. This ML model helps identify patients at high risk of recurrence after AIS and provide individualized treatment plans. Ma et al. [47] constructed RF models to identify AIS patients at high recurrence risk, finding that incorporating plasma phenylacetylglutamine (PAGln) levels improved the RF model' s AUC from 0.949 to 0.980, demonstrating good predictive performance. The study emphasized that ML models incorporating PAGln can more effectively predict stroke recurrence risk and highlighted the importance of this biomarker in predicting stroke recurrence.

4.4 ML Methods for Predicting AIS-Related Complications Various AIS complications such as hemorrhagic transformation (HT) and post-stroke cognitive impairment (PSCI) are closely associated with poor prognosis. Reasonable and timely prediction and assessment of complication risks after AIS can help select appropriate prevention and treatment methods to improve patient outcomes. Multiple ML methods have been used for complication prediction in AIS patients with significant results. Choi et al. [48] used SVM, XGBoost, and ANN to predict HT in AIS patients, with ANN performing best in predicting HT occurrence. However, this study used only clinical data without imaging analysis, representing a limitation. In contrast, Ren et al. [49] developed ML models based on clinical data, radiomics, and combined clinical-radiomics approaches to predict HT risk after intravenous thrombolysis in AIS patients, showing that the combined clinical-radiomics ML model outperformed pure clinical or radiomics models. The study indicated that combining clinical data with radiomics can improve HT risk prediction accuracy after AIS. Additionally, Lee et al. [50] used ML methods to predict PSCI after AIS, finding that XGBoost had the best predictive performance with an AUC of 0.792 and accuracy of 79.6%, facilitating early diagnosis and cognitive improvement measures. Ji et al. [51] used ML algorithms to predict PSCI occurrence within 3-6 months after AIS and identify key associated factors, finding that the Gaussian Naive Bayes model performed best with an AUC of 0.919 in the test set, while identifying age, education level, NIHSS score, brain protein degeneration, Hcy, and CRP as important PSCI predictors. However, current PSCI prediction studies have a common limitation: they only predict short-term PSCI risk within a few months and cannot assess cognitive performance beyond six months. Future research should expand the prediction timeframe to provide reliable long-term cognitive performance predictions.

5 Current Problems and Challenges

Despite the advantages of ML methods in AIS prognosis prediction, they still face numerous problems and challenges in clinical practice and translation.

5.1 Lack of Interpretability of ML Methods In current ML applications, most studies focus primarily on prediction accuracy while paying relatively insufficient attention to interpretation using specific data. Due to the complexity of ML models, the processes used to generate final outputs often lack transparency, making model results difficult to interpret [53-54]. In clinical applications, ML models are typically trained and evaluated in narrow environments for specific diseases, relying on researchers' personal expertise in statistics and ML. To facilitate understanding and use of models by healthcare workers for prediction, further interpretation of highly complex prediction models is necessary [55]. The interpretability of ML is as important as achieving high prediction accuracy [56], and future ML development should focus on improving methods for clinical decision interpretability.

5.2 Data Quality Issues and Standardization Difficulties High-quality, large-scale, and diverse datasets are key to successful ML models [57]. Healthcare data quality issues mainly manifest as data inaccuracy, outliers, incompleteness, isolated data records, and obsolescence [58]. Additionally, data interoperability between medical institutions and systems is low [59]. Currently, there are no standardized protocols or expert consensus for data acquisition, management, and application, and patient compliance, privacy, and research protection policies hinder ensuring data homogeneity and completeness. This requires coordinated cooperation among departments to build a healthcare data quality governance system to improve data quality and promote data sharing.

5.3 Difficulties in Selecting Optimal ML Models Stroke prognosis-related data are often imbalanced, and ML methods with data balancing techniques are effective tools for handling imbalanced data [60]. However, ML model performance partially depends on the compatibility between input data and models. Different types of algorithms incorporate different data and have varying application scenarios, with no unified standard indicating which algorithm has better classification performance. Furthermore, due to individual patient differences and clinical symptoms, different ML algorithms produce varying outcome predictions for stroke, making it challenging to build models that accurately reflect each patient' s unique situation. Future research needs to unify data inclusion standards and standardize application scenarios for ML algorithms to enhance model generalization ability for applicability to different patient populations and healthcare environments.

6 Summary and Outlook

With the advent of the big data era, the scale of available medical data and feature dimensions have increased dramatically, placing higher demands on data analysis methods [52]. However, many prediction studies on stroke recurrence and mortality have not considered survival time or have violated proportional hazards assumptions. ML methods have fewer restrictions on data distribution and unique advantages in handling large data, making ML models a feasible approach for survival analysis research. Additionally, current studies mostly use data from a single time point, ignoring causal relationships between variables and temporal evolution patterns of disease changes. For stroke prognosis prediction research, using ML methods for dynamic prediction has important value.

Prognosis prediction for AIS patients is crucial for aiding clinical decision-making, improving care quality, optimizing resource utilization, and enhancing doctor-patient communication. As ML and AI technologies deepen their applications in healthcare, comprehensive utilization of multimodal data from different sources—such as radiomics, biomarkers, and electronic medical records—can better capture the complex biological features and clinical manifestations of stroke patients, thereby providing more comprehensive prognostic predictions. ML is expected to achieve significant advances in AIS patient prognosis prediction research through multimodal data fusion, improved model interpretability, real-time dynamic prognosis monitoring, and individualized prognosis prediction.

Author Contributions: DU Huijie and LIU Xingyu were responsible for conceptualization and writing; XU Minghuan, YANG Xuezhi, ZHANG Huiqin, and MO Jiali were responsible for literature collection and organization; KUANG Jie was responsible for revision, quality control, and final approval of the manuscript; all authors confirmed the final version.

Conflict of Interest: The authors declare no conflict of interest.

References

[1] GBD Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016[J]. Lancet Neurol, 2019, 18(5): 459-480. DOI: 10.1016/S1474-4422(18)30499-X.

[2] WANG Yongjun, LI Zixiao, GU Hongqiu, et al. China Stroke Report 2020 (Chinese version) (1)[J]. Chin J Stroke, 2022, 17(5): 433-447. DOI: 10.3969/j.issn.1673-5765.2022.05.001.

[3] HANKEY GJ. Stroke[J]. Lancet, 2017, 389(10069): 641-654. DOI: 10.1016/S0140-6736(16)30962-X.

[4] HANDELMAN G S, KOK H K, CHANDRA R V, et al. eDoctor: machine learning and the future of medicine[J]. J Intern Med, 2018, 284(6): 603-619. DOI: 10.1111/joim.12822.

[5] MOONS K G M, ROYSTON P, VERGOUWE Y, et al. Prognosis and prognostic research: what, why, and how?[J]. BMJ, 2009, 338: b375. DOI: 10.1136/bmj.b375.

[6] GONG P Y, LIU Y K, GONG Y C, 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]. J Neuroinflammation, 2021, 18(1): 51. DOI: 10.1186/s12974-021-02503-7.

[7] FAINARDI E, BUSTO G, ROSI A, et al. Tmax volumes predict final infarct size and functional outcome in ischemic stroke patients receiving endovascular treatment[J]. Ann Neurol, 2022, 91(6): 878-888. DOI: 10.1002/ana.26354.

[8] WANG K, SHI Q Q, SUN C, et al. A machine learning model for visualization and dynamic clinical prediction of stroke recurrence in acute ischemic stroke patients: a real-world retrospective study[J]. Front Neurosci, 2023, 17: 1130831. DOI: 10.3389/fnins.2023.1130831.

[9] TONG L, SUN Y, ZHU Y Q, et al. Prognostic estimation for acute ischemic stroke patients undergoing mechanical thrombectomy within an extended therapeutic window using an interpretable machine learning model[J]. Front Neuroinform, 2023, 17: 1273827. DOI: 10.3389/fninf.2023.1273827.

[10] SHEN B, YANG X, SUI R B, et al. The prognostic value of the THRIVE score, the iScore score and the ASTRAL score in Chinese patients with acute ischemic stroke[J]. J Stroke Cerebrovasc Dis, 2018, 27(10): 2877-2886. DOI: 10.1016/j.jstrokecerebrovasdis.2018.06.011.

[11] COORAY C, MAZYA M, BOTTAI M, et al. External validation of the ASTRAL and DRAGON scores for prediction of functional outcome in stroke[J]. Stroke, 2016, 47(6): 1493-1499. DOI: 10.1161/STROKEAHA.116.012802.

[12] DROZDOWSKA B A, SINGH S, QUINN T J. Thinking about the future: a review of prognostic scales used in acute stroke[J]. Front Neurol, 2019, 10: 274. DOI: 10.3389/fneur.2019.00274.

[13] BAŞTANLAR Y, OZUYSAL M. Introduction to machine learning[J]. Methods Mol Biol, 2014, 1107: 105-128. DOI: 10.1007/978-1-62703-748-8_7.

[14] OBERMEYER Z, EMANUEL E J. Predicting the future - big data, machine learning, and clinical medicine[J]. N Engl J Med, 2016, 375(13): 1216-1219. DOI: 10.1056/NEJMp1606181.

[15] JANIESCH C, ZSCHECH P, HEINRICH K. Machine learning and deep learning[J]. Electron Mark, 2021, 31(3): 685-695. DOI: 10.1007/s12525-021-00475-2.

[16] KOBAYASHI M, HUTTIN O, MAGNUSSON M, et al. Machine learning-derived echocardiographic phenotypes predict heart failure incidence in asymptomatic individuals[J]. JACC Cardiovasc Imaging, 2022, 15(2): 193-208. DOI: 10.1016/j.jcmg.2021.07.004.

[17] SHAO H, SHI L Z, FONSECA V. Comment on segar et al. machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: the WATCH-DM risk score. diabetes care 2019;42:2298-2306[J]. Diabetes Care, 2020, 43(2): e25. DOI: 10.2337/dc19-1891.

[18] SEGAR M W, JAEGER B C, PATEL K V, et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis[J]. Circulation, 2021, 143(24): 2370-2383. DOI: 10.1161/CIRCULATIONAHA.120.053134.

[19] KASHOU A H, MEDINA-INOJOSA J R, NOSEWORTHY P A, et al. Artificial intelligence-augmented electrocardiogram detection of left ventricular systolic dysfunction in the general population[J]. Mayo Clin Proc, 2021, 96(10): 2576-2586. DOI: 10.1016/j.mayocp.2021.02.029.

[20] BRATT A, KIM J, POLLIE M, et al. Machine learning derived segmentation of phase velocity encoded cardiovascular magnetic resonance for fully automated aortic flow quantification[J]. J Cardiovasc Magn Reson, 2019, 21(1): 1. DOI: 10.1186/s12968-018-0509-0.

[21] YANG F F, CHEN X T, LIN X X, et al. Automated analysis of Doppler echocardiographic videos as a screening tool for valvular heart diseases[J]. JACC

Cardiovasc Imaging, 2022, 15(4): 551-563. DOI: 10.1016/j.jcmg.2021.08.015.

[22] DEBAUN M R, CHAVEZ G, FITHIAN A, et al. Artificial neural networks predict 30-day mortality after hip fracture: insights from machine learning[J]. J Am Acad Orthop Surg, 2021, 29(22): 977-983. DOI: 10.5435/JAAOS-D-20-00429.

[23] JU C S, ZHOU J D, LEE S R, et al. Derivation of an electronic frailty index for predicting short-term mortality in heart failure: a machine learning approach[J]. ESC Heart Fail, 2021, 8(4): 2837-2845. DOI: 10.1002/ehf2.13358.

[25] WANG X Y, LYU J H, MENG Z H, et al. Small vessel disease burden predicts functional outcomes in patients with acute ischemic stroke using machine learning[J]. CNS Neurosci Ther, 2023, 29(4): 1024-1033. DOI: 10.1111/cns.14071.

[26] TONG L, SUN Y, ZHU Y Q, et al. Prognostic estimation for acute ischemic stroke patients undergoing mechanical thrombectomy within an extended therapeutic window using an interpretable machine learning model[J]. Front Neuroinform, 2023, 17: 1273827. DOI: 10.3389/fninf.2023.1273827.

[27] BI Q, GOODMAN K E, KAMINSKY J, et al. What is machine learning? A primer for the epidemiologist[J]. Am J Epidemiol, 2019, 188(12): 2222-2239. DOI: 10.1093/aje/kwz189.

[28] KRIEGESKORTE N, GOLAN T. Neural network models and deep learning[J]. Curr Biol, 2019, 29(7): R231-R236. DOI: 10.1016/j.cub.2019.02.034.

[29] HEO J, YOON J G, PARK H, et al. Machine learning-based model for prediction of outcomes in acute stroke[J]. Stroke, 2019, 50(5): 1263-1268. DOI: 10.1161/STROKEAHA.118.024293.

[30] JIANG B, ZHU G, XIE Y, et al. Prediction of clinical outcome in patients with large-vessel acute ischemic stroke: performance of machine learning versus SPAN-100[J]. AJNR Am J Neuroradiol, 2021, 42(2): 240-246. DOI: 10.3174/ajnr.A6918.

[31] TEKKEŞIN A İ. Artificial intelligence in healthcare: past, present and future[J]. Anatol J Cardiol, 2019, 22(Suppl 2): 8-9. DOI: 10.14744/AnatolJCardiol.2019.28661.

[32] BADILLO S, BANFAI B, BIRZELE F, et al. An introduction to machine learning[J]. Clin Pharmacol Ther, 2020, 107(4): 871-885. DOI: 10.1002/cpt.1796.

[33] YANN L, YOSHUA B, GEOFFREY H. Deep learning[J]. Nature, 2015, 521(7553): 436-444. DOI: 10.1038/nature14539.

[34] GREENER J G, KANDATHIL S M, MOFFAT L, et al. A guide to machine learning for biologists[J]. Nat Rev Mol Cell Biol, 2022, 23: 40-55. DOI: 10.1038/s41580-021-00407-0.

[35] YANG Y, SHI Y Z, ZHANG N, et al. The disability rate of 5-year post-stroke and its correlation factors: a national survey in China[J]. PLoS One, 2016, 11(11): e0165341. DOI: 10.1371/journal.pone.0165341.

[36] PARK D, JEONG E, KIM H, et al. Machine learning-based three-month outcome prediction in acute ischemic stroke: a single cerebrovascular-specialty hospital study in South Korea[J]. Diagnostics, 2021, 11(10): 1909. DOI: 10.3390/diagnostics11101909.

[37] JO H, KIM C, GWON D, et al. Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: an automated machine learning approach[J]. Sci Rep, 2023, 13(1): 16926. DOI: 10.1038/s41598-023-44201-8.

[38] OZKARA B B, KARABACAK M, HAMAM O, et al. Prediction of functional outcome in stroke patients with proximal middle cerebral artery occlusions using machine learning models[J]. J Clin Med, 2023, 12(3): 839. DOI: 10.3390/jcm12030839.

[39] LI X, PAN X D, JIANG C L, et al. Predicting 6-month unfavorable outcome of acute ischemic stroke using machine learning[J]. Front Neurol, 2020, 11: 539509. DOI: 10.3389/fneur.2020.539509.

[40] LEE J, PARK K M, PARK S. Interpretable machine learning for prediction of clinical outcomes in acute ischemic stroke[J]. Front Neurol, 2023, 14: 1234046. DOI: 10.3389/fneur.2023.1234046.

[41] Report on Stroke Prevention and Treatment in China Writing Group. Summary of "China Stroke Prevention and Treatment Report 2021" [J]. Chin J Cerebrovasc Dis, 2023, 20(11): 783-793.

[42] WANG K, GU L Y, LIU W C, et al. The predictors of death within 1 year in acute ischemic stroke patients based on machine learning[J]. Front Neurol, 2023, 14: 1092534. DOI: 10.3389/fneur.2023.1092534.

[43] LEHMANN A L C F, ALFIERI D F, DE ARAÚJO M C M, et al. Immune-inflammatory, coagulation, adhesion, and imaging biomarkers combined in machine learning models improve the prediction of death 1 year after ischemic stroke[J]. Clin Exp Med, 2022, 22(1): 111-123. DOI: 10.1007/s10238-021-00732-w.

[44] ZHANG J, YUAN T, WEI S X, et al. New strategy for clinical etiologic diagnosis of acute ischemic stroke and blood biomarker discovery based on machine learning[J]. RSC Adv, 2022, 12(23): 14716-14723. DOI: 10.1039/d2ra02022j.

[45] GBD LIFETIME RISK OF STROKE COLLABORATORS, FEIGIN V L, NGUYEN G, et al. Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016[J]. N Engl J Med, 2018, 379(25): 2429-2437. DOI: 10.1056/NEJMoa1804492.

[46] DING L L, MANE R, WU Z Z, et al. Data-driven clustering approach to

identify novel phenotypes using multiple biomarkers in acute ischaemic stroke: a retrospective, multicentre cohort study[J]. EClinicalMedicine, 2022, 53: 101639. DOI: 10.1016/j.eclinm.2022.101639.

[47] MA L, FU G P, LIU R R, et al. Phenylacetyl glutamine: a novel biomarker for stroke recurrence warning[J]. BMC Neurol, 2023, 23(1): 74. DOI: 10.1186/s12883-023-03118-5.

[48] CHOI J M, SEO S Y, KIM P J, et al. Prediction of hemorrhagic transformation after ischemic stroke using machine learning[J]. J Pers Med, 2021, 11(9): 863. DOI: 10.3390/jpm11090863.

[49] REN H H, SONG H J, WANG J J, et al. A clinical-radiomics model based on noncontrast computed tomography to predict hemorrhagic transformation after stroke by machine learning: a multicenter study[J]. Insights Imaging, 2023, 14(1): 52. DOI: 10.1186/s13244-023-01399-5.

[50] LEE M, YEO N Y, AHN H J, et al. Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning[J]. Alzheimers Res Ther, 2023, 15(1): 147. DOI: 10.1186/s13195-023-01289-4.

[51] JI W C, WANG C J, CHEN H Q, et al. Predicting post-stroke cognitive impairment using machine learning: a prospective cohort study[J]. J Stroke Cerebrovasc Dis, 2023, 32(11): 107354. DOI: 10.1016/j.jstrokecerebrovasdis.2023.107354.

[52] RAJPURKAR P, CHEN E, BANERJEE O, et al. AI in health and medicine[J]. Nat Med, 2022, 28(1): 31-38. DOI: 10.1038/s41591-021-01614-0.

[53] BELTRAMIN D, LAMAS E, BOUSQUET C. Ethical issues in the utilization of black boxes for artificial intelligence in medicine[J]. Stud Health Technol Inform, 2022, 295: 249-252. DOI: 10.3233/SHTI220709.

[54] REYES M, MEIER R, PEREIRA S, et al. On the interpretability of artificial intelligence in radiology: challenges and opportunities[J]. Radiol Artif Intell, 2020, 2(3): e190043. DOI: 10.1148/ryai.2020190043.

[55] YANG Fengchun, ZHENG Si, LI Jiao. Application of interpretable machine learning methods in disease prediction: a study on mortality risk in sepsis patients[J]. J Capital Med Univ, 2022, 43(4): 610-617. DOI: 10.3969/j.issn.1006-7795.2022.04.015.

[56] LUNDBERG S M, ERION G, CHEN H, et al. From local explanations to global understanding with explainable AI for trees[J]. Nat Mach Intell, 2020, 2(1): 56-67. DOI: 10.1038/s42256-019-0138-9.

[57] HAGENDORFF T. Linking human and machine behavior: a new approach to evaluate training data quality for beneficial machine learning[J]. Minds Mach, 2021, 31(4): 563-593. DOI: 10.1007/s11023-021-09573-8.

[58] WANG M Y, LI S Y, ZHENG T, et al. Big data health care platform with multisource heterogeneous data integration and massive high-dimensional data

governance for large hospitals: design, development, and application[J]. JMIR Med Inform, 2022, 10(4): e36481. DOI: 10.2196/36481.

[59] GAO Jinghong, LI Mingyuan, WANG Lin, et al. Application and challenges of health medical big data in precision medicine[J]. J Med Inform, 2022, 43(5): 15-20. DOI: 10.3969/j.issn.1673-6036.2022.05.003.

[60] WU Y F, FANG Y. Stroke prediction with machine learning methods among older Chinese[J]. Int J Environ Res Public Health, 2020, 17(6): 1828. DOI: 10.3390/ijerph17061828.

Note: Figure translations are in progress. See original paper for figures.