Interpretable Machine Learning for Early Neurological Deterioration Prediction in Atrial Fibrillation-Related Stroke

Seong Hwan Kim
Korea University Ansan Hospital
Eun-Tae Jeon
Korea University Ansan Hospital
Sungwook Yu
Korea University Anam Hospital, Korea University College of Medicine
Kyungmi O
Korea University Guro Hospital, Korea University College of Medicine
Chi Kyung Kim
Korea University Guro Hospital, Korea University College of Medicine
Tae-Jin Song
Ewha Womans University
Yong-Jae Kim
Eunpyeong St. Mary's Hospital, The Catholic University of Korea
Sung Hyuk Heo
Kyung Hee University College of Medicine
Kwang-Yeol Park
Chung-Ang University College of Medicine, Chung-Ang University Hospital
Jeong-Min Kim
Chung-Ang University College of Medicine, Chung-Ang University Hospital
Jong-Ho Park
Hanyang University Myongji Hospital Seoul
Jay Chol Choi
Jeju National University
Man-Seok Park
Chonnam National University Hospital
Joon-Tae Kim
Chonnam National University Hospital
Kang-Ho Choi
Chonnam National University Hwasun Hospital
Yang Ha Hwang
Kyungpook National University Hospital
Bum Joon Kim

Asan Medical Center, University of Ulsan College of Medicine
Jong-Won Chung
Samsung Medical Center, Sungkyunkwan University School of Medicine
Oh Young Bang
Samsung Medical Center, Sungkyunkwan University School of Medicine
Gyeongmoon Kim
Samsung Medical Center, Sungkyunkwan University School of Medicine
Woo-Keun Seo
Samsung Medical Center, Sungkyunkwan University School of Medicine
Jin-Man Jung (dr.jinmanjung@gmail.com)
Korea University Zebrafish Translational Medical Research Center

---

Research Article

Keywords: early neurological deterioration (END), machine learning (ML), atrial fibrillation (AF)-related stroke

Posted Date: May 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-446890/v1

Abstract

We aimed to develop a novel prediction model for early neurological deterioration (END) based on an interpretable machine learning (ML) algorithm for atrial fibrillation (AF)-related stroke and to evaluate the prediction accuracy and feature importance of ML models. Data from multi-center prospective stroke registries in South Korea were collected. After stepwise data preprocessing, we utilized logistic regression, support vector machine, extreme gradient boosting, light gradient boosting machine (LightGBM), and multilayer perceptron models. We used the Shapley additive explanations (SHAP) method to evaluate feature importance. Of the 3,623 stroke patients, the 2,363 who had arrived at the hospital within 24 hours of symptom onset and had available information regarding END were included. Of these, 318 (13.5%) had END. The LightGBM model showed the highest area under the receiver operating characteristic curve (0.778, 95% CI, 0.726 - 0.830). The feature importance analysis revealed that fasting glucose level and the National Institute of Health Stroke Scale score were the most influential factors. Among ML algorithms, the LightGBM model was particularly useful for predicting END, as it revealed new and diverse predictors. Additionally, the SHAP method can be adjusted to individualize the features' effects on the predictive power of the model.

Introduction

Early neurological deterioration (END) is a sudden worsening of neurological symptoms during the acute period of stroke. END leads to devastating clinical outcomes despite marked advances in acute stroke management over the past several years. The incidence of END is considerably high, ranging from 5–40%, and is associated with a poor 3-month clinical prognosis and high mortality.[1,2] Therefore, an accurate prediction of END is necessary for early identification and timely management of ischemic stroke. However, because of the complexity and heterogeneity of END, there has been no consensus on the definition. Therefore, various inclusion criteria and study designs have been used, with some studies preferring to define END according to specific stroke subtypes (e.g., cardioembolism), making each of the predictors and recent nomograms difficult to use in real-world clinical practice.[3-6]

Of the etiologies attributed to cardioembolic stroke, atrial fibrillation (AF) is one of the predictors of END.[7,8] Several markers, including clinical, radiological, and laboratory findings, have been associated with END in AF-related stroke.[9-11] However, in those studies, using a single marker had limited predictive power, since the diverse biomarkers and imaging markers relevant to END in AF-related stroke were not considered at the same time.

Continual advancements in machine learning (ML) algorithms have led to their wide application in the medical field, since numerous variables and massive data can be included and analyzed. Contrary to transitional statistical models, ML models are compatible with predicting complex clinical events that can be affected by diverse situations and conditions. Nevertheless, the clinical application of ML models has been limited due to the 'black box problem' of interpretability and explanation.[12] Therefore, it is essential that ML models be interpretable to the current medical fields.[13] The Shapley additive explanations (SHAP) method is a novel, cutting-edge method designed to aid in clinical interpretation and intuitive understanding of feature importance by providing visualizations of the relationship between each feature and the associated predictive power.[14] Therefore, the aim of our study was to develop an interpretable ML model that could predict END using the feature importance technique in AF-related stroke using a real-world multi-center cohort database.

Methods

Study Design and Participants

The dataset from this study are available to be provided from the corresponding author if reasonable request.

This study was based on the Korean Atrial Fibrillation Evaluation Registry in Ischemic Stroke Patients (K-ATTENTION), a real-world cohort composed of prospective stroke registries from 11 tertiary centers in South Korea. K-ATTENTION focused on characteristics, oral anticoagulant use, and outcomes in AF-related stroke patients.[15] Between January 2013 and December 2015, patients who were admitted to one of the participating centers within 7 days of stroke onset were enrolled. Detailed information regarding management and follow-up of the included patients has been provided previously.[15] In our study, only those who arrived at the hospital within 24 h of symptom onset and had information regarding END were included. Using the internet-based clinical recording system, we acquired the following patient information from each center: demographic characteristics, vascular risk factors, brain imaging results, laboratory findings, pre-admission medication histories, stroke severity on admission (according to the National Institutes of Health Stroke Scale [NIHSS] score), and functional status (modified Rankin Score [mRS]). More information on variable acquisition and evaluation is provided in Supplemental Table I and Supplemental Methods I. This study followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.[16] The institutional review boards of Korea University Ansan Hospital (2016AS0051), Korea university Anam hospital, Korea University Guro Hospital, Ewha Womans University School of Medicine, Eunpyeong St. Mary's Hospital, Kyung Hee University College of Medicine, Chung-Ang University Hospital, Hanyang University Myongji Hospital, Jeju National University, Chonnam National University Hospital, Chonnam National University Hwasun Hospital, Kyungpook National University Hospital, Asan Medical Center, and Samsung Medical Center approved the study. The need for informed consent was waived by the ethics committee of all of participating centers due to the retrospective design of the study using anonymous and de-identified information.

Definition of END as the Main Outcome

END was defined as an increase of at least 2 points in the total NIHSS score and at least 1 point on the level of consciousness or motor items score within 72 h of arrival at the hospital.[17]

Data Splitting and Preprocessing

Binary variables with less than 80% missing values and multinomial and numeric variables with less than 60% missing values were included to generate the available dataset.[18] In the first step, 25% of the dataset was randomly separated according to END stratification and used only in the final evaluation of model performance as a test set. The remaining 75% of the dataset was used as a training set for hyperparameter determination and training processes using leave-one-out cross-validation. Isolation forest and multivariate imputation by chained equations were used for outlier detection and imputation. Details of the methods are provided in Supplemental Methods II.

Feature Selection and Feature Importance Analyses

Recursive feature elimination[19] was used to select the top-k ranked features that contributed to the overall model performance of the area under the receiver operating characteristic curve (AUROC). To measure and rank the contribution of each variable, we obtained mean absolute SHAP values[14] with a gradient boosted tree-based model, light gradient boosting machine (LightGBM)[20], which can deal natively with categorical features[21] using leave-one-out cross-validation. The positive SHAP value for each variable indicated that the variable contributed positively to the model's positive prediction, and vice versa. We performed an additional stepwise process to prevent underestimation of relative importance of features due to multicollinearity, the details of which are described in Supplemental Methods III.

Modeling

We selected and tested one conventional statistical model, logistic regression[22], as a baseline comparator, and four popular ML models which were support vector machine[23], extreme gradient boosting[24] (XGBoost), Light GBM, and multilayer perceptron[25] (MLP) with a basic architecture. Detailed instructions of the applied models are provided in Supplemental Methods IV. All the processes were implemented in Python 3.8.2 with TensorFlow-GPU 2.4.0[26] and scikit-learn 0.22.1[27] libraries.

Primary Outcome and Evaluation Criteria

AUROC was chosen as a primary evaluation metric for model performance, and all the cross-validation and early stopping strategies in the modeling process were performed to maximize the AUROC score. The models were evaluated for frequency of confident answers and errors, with a threshold of 0.50.

Statistical Analysis

Categorical variables are presented as number (percentage), and continuous variables are presented as mean ± standard deviation or median (interquartile range), as appropriate. A simple comparison was performed using the χ^2 test for categorical variables and the Kruskal-Wallis test for continuous variables. Data analyses were performed using IBM SPSS version 20 software (IBM Corp. Armonk, NY, USA). The AUROC, with a 95% confidence interval (CI), was calculated using the Delong method and a CI that spanned .50 or more was not considered statistically different from a random performance.[28] To evaluate the calibration error of the models, the Brier score, which is the mean squared difference between the predicted probability and the actual outcome, was calculated,[29] with a lower score indicating better probabilistic prediction accuracy. In addition, the area under the precision-recall curve, accuracy, precision, recall, and F1 score were calculated as secondary outcome metrics. The significance level was set at $p < 0.05$ and Bonferroni correction was used for the multiple comparison of the AUROC between models.

Results

Comparisons of Baseline Characteristics

Figure 1 shows the patient flow chart. A total of 2,363 patients were included in this study, of which 318 (13.5%) had END. Comparisons of baseline clinical characteristics and MRI variables are listed in Supplemental Tables II and III.

Missing Value Imputation

The binary variables with missing values over 80% and the multinomial and numeric variables with missing values over 60% were excluded from the model construction dataset according to the missing data imputation strategy described in a previous study.[30] The variables were as follows: all Holter monitoring parameters, smoking pack-years, alcohol consumption, duration of PR and P-axis wave on electrocardiogram, susceptibility vessel sign (SVS) size, urine albumin, serum free fatty acid level, brain natriuretic peptide (BNP), N-terminal proBNP, and troponin T. The remaining missing values were imputed, with non-categorical missing values imputed using the multivariate imputation by chained equations imputation method, and categorical missing values were replaced with a single constant of -1. Details concerning the number of missing values for each variable are listed in Supplemental Table IV.

Model Performances

A flow diagram of the ML model development process is presented in Supplemental Figure I. The performance of each model is shown in Table 1, and the receiver operating characteristic curve and the precision-recall curve are shown in Figure 2. The LightGBM had the highest AUROC value (0.778 [0.726-0.830]), however, there was no significant difference between ML models. The Light GBM and the MLP had significantly higher AUROC values than logistic regression (p=0.0048 and 0.0017, respectively).

Table 1. Comparison of model performance

	Model	*AUROC [95% CI]*	*AUPRC [95% CI]*	*Accuracy [%]*	*Precision*	*Recall*	*F1 score*	*Brier score*	*p value†*
Conventional statistical model	Logistic regression	0.701 [0.647 - 0.755]	0.241 [0.154 - 0.331]	86.49	0.252	0.5	0.335	0.114	
Machine learning model	SVM	0.723 [0.668 - 0.777]	0.267 [0.176 - 0.359]	86.66	0.279	0.575	0.376	0.109	0.436
	XGBoost	0.771 [0.722 - 0.819]	0.306 [0.204 - 0.407]	86.32	0.289	0.7	0.409	0.105	0.011
	LightGBM	0.778 [0.726 - 0.830]	0.354 [0.239 - 0.465]	86.49	0.331	0.562	0.417	0.102	0.005*
	MLP	0.767 [0.713 - 0.820]	0.340 [0.223 - 0.451]	86.15	0.323	0.65	0.432	0.103	0.002*

*Significant difference at $p < 0.005$.
†Comparison with logistic regression on AUROC.

Abbreviations: AUROC, Area under the receiver operating characteristic curve; AUPRC, Area under the precision recall curve; SVM, support vector machine; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; MLP, Multilayer perceptron

Identification of Important Features

From the recursive feature elimination, a total of 24 features were selected as important features. The SHAP feature importance matrix plots show important features according to the degree of contribution (bar plot, Figure 3A) and the overall correlation and directionality between features and the SHAP value (violin plot, Figure 3B) during model construction. Among them, fasting glucose levels and initial NIHSS score contributed the most to the model. The next highest-ranking features were the initial modified Rankin score (mRS) and initial glucose level. All the other features contributed to the model less. In addition, most of the continuous variables, such as the fasting glucose, initial NIHSS score, initial mRS, initial glucose, QRS axis, alkaline phosphatase, homocysteine, fibrin degradation product, initial diastolic blood pressure, D-dimer, hematocrit, total cholesterol, and T-axis had a tendency to be positively correlated with END. Activated partial thromboplastin time, aspartate aminotransferase, total bilirubin, and low-density lipoprotein (LDL) cholesterol showed complex patterns with mixed positive and negative trends. LA diameter and uric acid levels showed a negative correlation.

SHAP values corresponding to changes in four representative features are presented in partial SHAP dependence plots (Figure 4), and other representative feature plots are listed in Supplemental Figure II. The fasting glucose level and initial NIHSS score showed a positive correlation with the sigmoid or double sigmoid curve. The LA diameter declined in a negative pattern. LDL cholesterol was associated with a U-shaped trend line that initially showed a declining tendency followed by a reversed, increasing trend. The cut-off value for each variable that could predict the positive and/or negative probability of END occurrence is marked on each graph.

Lateralization of ischemic lesions and concomitant intracranial atherosclerosis, SVS sign, hemorrhagic transformation and AF diagnosis time were included as categorical variables. The presence of concomitant intracranial atherosclerosis, SVS sign, and symptomatic ICH among hemorrhagic transformation and AF diagnosed during hospitalization or after discharge is likely to related with END occurrence. Posterior circulation lesions are unlikely to develop END (Supplemental Figure II).

In addition, we acquired information about the importance and contribution for each patient according to specific features selected during modeling. Representative cases are summarized in Supplemental Figure III.

Discussion

In this study, we first demonstrated that integrated ML algorithms can be applied to predict END in AF-related stroke. Among the ML models investigated, the LightGBM had the best performance, with an AUROC value of 0.778. This is a novel method with efficient computational power and wide scalability for processing categorical, multidimensional, and incredibly large datasets,[20] which makes it a suitable ML model for the medical field. In addition, this model was implemented using SHAP, which can visualize the level of contribution and directionality of specific input features using the entire dataset as well as individual patient information.

The highest contributing feature in our study was the fasting glucose level, followed by initial NIHSS score which represents the degree of initial neurological functional deficits. These variables have been consistently reported as risk factors for END in all-type as well as AF-related strokes.[2,10,11] A possible explanation is that the impairment of glucose control causes vascular endothelial dysfunction,[31] post-ischemic inflammatory response, and neuroprotective heat-shock chaperone gene attenuation,[32] which could exacerbate post-stroke brain damage through increasing lactate production and leading to the breakdown of the blood-brain barrier, development of brain edema and hemorrhagic transformation, and enlargement of infarct volume.[18] In fact, symptomatic cerebral hemorrhage of hemorrhagic transformation subtype was positively associated with END in this study. Also homocysteine, which was related to vascular endothelial dysfunction,[3] and fibrin degradation product and D-dimer, which were important hematologic markers related to the coagulation system and thrombosis, were important features like previous studies.[33-35] Other features were SVS presence implying large-size infarction; specific ischemic lesion location limited to anterior or posterior circulation;[36] cardiac electrophysiological, and echocardiographic markers such as QRS axis, T axis and left atrium diameter; alkaline phosphatase[37,38] as surrogate markers of atherosclerosis, systemic inflammation, malnutrition, or metabolic syndrome; and the burden of atherosclerosis, such as concurrent intracranial atherosclerosis.[36,39] Among cholesterol lipoproteins, total cholesterol, and LDL were included as important features in this study, which have been previously reported as important predictors.[40]

Interestingly, the clinical implication of cut-off values in selected features may be applicable to real-world clinical practice. With regard to initial stroke severity measured using the NIHSS, cut-off values in the SHAP partial dependence plot were presented according to the effect direction of END prediction, suggesting that patients with severe stroke (NIHSS $\geq$ 16) tended to develop END, suggesting that awareness and close medical attention are necessary for these patients, and patients with mild to moderate stroke (NIHSS $\leq$ 6) have a lower chance of developing END. Some cut-off values were statistically significantly similar to clinical values. Indeed, the cut-off value for fasting glucose predicting END in our study was 117.6 mg/dL, which corresponds to the current diagnostic criteria for diabetes mellitus ($\geq$126 mg/dL).[41]

The SHAP and its corresponding graphs, which were used to evaluate the effect that continuous variables had on the prediction of END, were characterized by four patterns. First, a positive correlation with or without a sigmoid or double sigmoid shape was observed. The initial glucose level, fasting glucose level, initial NIHSS score, initial mRS score, homocysteine, D-dimer, fibrin degradation product, initial diastolic blood pressure, total cholesterol, QRS-axis, and T-axis corresponded to this pattern. Most of these variables have been reported as predictors of END in previous studies.[2] Second, a U-curve or J-shaped pattern with both cut-off values was observed for aspartate aminotransferase, alkaline phosphatase, total bilirubin, and LDL cholesterol. Under the lower cut-off value of each feature may have been associated with poor nutritional status and over the upper cut-off value may imply comorbid conditions including liver disease and hyperlipidemia. However, it is unavailable to investigate underlying pathomechanism of these phenomena in this study. Third, the following had a negative correlation with END, with a reverse S or J shape: LA diameter and uric acid. In particular, the negative association between LA diameter and END is not consistent with the positive correlation found in a previous report.[42] However, more accurate parameters, such as the LA volume index, have recently been identified as important predictors. Considerable imputation (21.1%) could lead to incorrect directions and biased results. Finally, a bizarre pattern with multi-directionality was observed in activated partial thromboplastin time.

One strength of our study is that our interpretable ML model was constructed using many variables, including demographics and laboratory, radiological, and echocardiographic findings, all of which can be obtained upon arrival at the hospital. Additionally, an interpretable and explainable ML model was created to promote the use of applications for making clinical decisions. Our study demonstrates the potential of interpretable ML methods to predict END and individualize such predictions. Previous studies have focused on each risk factor individually and its pathophysiological interpretation, but there has been a shortage of clinical use of a large combination of variables at once.[3-5] Moreover, no standardized risk stratification scheme for predicting END has been available until now. Therefore, our ML model has the advantage of being able to predict END using diverse variables extracted from real-world clinical situations upon arrival at the hospital.

Our study has several limitations. First, a considerable amount of data was missing due to the multi-center retrospective nature of the study. Although imputation of missing data was performed using the ML technique, the results may be biased and contradict previous findings. In particular, it seemed to occur with some elements (such as left atrial size) that were less important. In addition, laboratory and imaging protocols in each center were not concretely established before data collection. Additionally, Holter and electrocardiography parameters were not standardized; therefore, many variables were excluded. Second, since this was a registry-based study with a retrospective design, the ML model's performance is not sufficient to be an absolute criterion for clinical use. It is necessary to conduct prospective studies, develop a more accurate prediction model, and discover novel biomarkers for a deeper understanding of the pathophysiology, in parallel. Third, the implementation and evaluation of the model was difficult to generalize because of the lack of external validation. Further verification is required through well-designed prospective clinical studies in the future.

In conclusion, ML algorithms, using the LightGBM model in particular, can be used to predict END in AF-related stroke. New and diverse predictors for END were revealed through this ML model, suggesting that the pathophysiology of END development could be a complex mechanism. Further verification through prospective clinical studies is needed.

Declarations

Acknowledgments

None

Author contributions

WKS and JMJ conceived and designed the research. SHK contributed to the drafting of the manuscript, statistical analysis, and interpretation of data. ETJ contributed to the machine learning model construction, interpretation of data and drafting of the manuscript. SWY, KMO, CKK, TJS, TJK, SHH, KYP, JMK, JHP, JCC, MSP, JTK, KHC, YHH, BJK, JWC, OYB, and GMK were involved in the acquisition of data and take responsibility for the integrity of data. All authors approved the final manuscript.

Sources of Funding

The authors disclose receipt of the following financial support for the research, authorship, and publication of this article: the Basic Science Research Program through the National Research Foundation of Korea (NRF)

funded by the Ministry of Science and ICT (NRF-2020R1C1C1009294) and Korea University Grant. The funders had no role in the study design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Declaration of conflicting interests

The authors declare the following potential conflicts of interest with respect to the research, authorship, and publication of this article: J-M Jung has received lecture honoraria from Pfizer, Sanofi-Aventis, Ostuka, Dong-A, and Hanmi Pharmaceutical Co., Ltd; consulting fees from Daewoong Pharmaceutical Co., Ltd. WK Seo received honoraria for lectures from Pfizer, Sanofi-Aventis, Otsuka Korea, Dong-A Pharmaceutical Co., Ltd., Beyer, Daewoong Pharmaceutical Co. Ltd., Daiichi Sankyo Korea Co., Ltd., and Boryung Pharmaceutical Co., Ltd.; a study grant from Daiichi Sankyo Korea Co., Ltd.; and consulting fees from OBELAB Inc. All other authors have no conflict of interest

References

1. Siegler, J. E. & Martin-Schild, S. Early Neurological Deterioration (END) after stroke: the END depends on the definition. *International Journal of Stroke* **6**, 211-212 (2011).
2. Seners, P., Turc, G., Oppenheim, C. & Baron, J.-C. Incidence, causes and predictors of neurological deterioration occurring within 24 h following acute ischaemic stroke: a systematic review with pathophysiological implications. *J Neurol Neurosurg Psychiatry* **86**, 87-94 (2015).
3. Kwon, H.-M., Lee, Y.-S., Bae, H.-J. & Kang, D.-W. Homocysteine as a predictor of early neurological deterioration in acute ischemic stroke. *Stroke* **45**, 871-873 (2014).
4. Sun W. *et al*. Asymmetrical cortical vessel sign on susceptibility-weighted imaging: a novel imaging marker for early neurological deterioration and unfavorable prognosis. *Eur J Neurol* **21**, 1411-1418 (2014).
5. Seo W-K, *et al*. C-reactive protein is a predictor of early neurologic deterioration in acute ischemic stroke. *Journal of stroke and cerebrovascular diseases* **21**, 181-186 (2012).
6. Gong P, *et al*. A Novel Nomogram to Predict Early Neurological Deterioration in Patients with Acute Ischemic Stroke. *Eur J Neurol*. **27,** 1996–2005 (2020).
7. Kwan J, Hand P. Early neurological deterioration in acute stroke: clinical characteristics and impact on outcome. *Journal of the Association of Physicians* **99**, 625-633 (2006).
8. Thanvi B, Treadwell S, Robinson T. Early neurological deterioration in acute ischaemic stroke: predictors, mechanisms and management. *Postgraduate Medical Journal* **84**, 412-417 (2008).
9. Hong H, *et al*. Early neurological outcomes according to CHADS2 score in stroke patients with non-valvular atrial fibrillation. *Eur J Neurol* **19**, 284-290 (2012).
10. Kim J-S, *et al*. Pre-stroke glycemic control is associated with early neurologic deterioration in acute atrial fibrillation-related ischemic stroke. *eNeurologicalSci* **8**, 17-21 (2017).
11. Duan Z, *et al*. Relationship between high-sensitivity C-reactive protein and early neurological deterioration in stroke patients with and without atrial fibrillation. *Heart Lung*. **49** 193–197 (2020).
12. Michael KY, *et al*. Visible machine learning for biomedicine. *Cell* **173**, 1562-1565 (2018).

13. Tjoa E, Guan C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. Preprint at https://arxiv.org/abs/1907.07374 *IEEE Trans Neural Netw Learn Syst*. (2020).
14. Lundberg SM, *et al*. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. **2**, 2522-5839 (2020).
15. Jung J-M, *et al*. Long-Term Outcomes of Real-World Korean Patients with Atrial-Fibrillation-Related Stroke and Severely Decreased Ejection Fraction. *J Clin Neurol*. **15**, 545-554 (2019).
16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BJS*. **102**, 148-158 (2015).
17. Nam KW, *et al*. D-dimer as a predictor of early neurologic deterioration in cryptogenic stroke with active cancer. *Eur J Neurol*. **24**, 205-211 (2017).
18. McBride DW, *et al*. Acute hyperglycemia is associated with immediate brain swelling and hemorrhagic transformation after middle cerebral artery occlusion in rats*. Acta Neurochir Suppl.* **121,** 237–241(2016).
19. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning*. **46**, 389-422 (2002).
20. Ke G, *et al*. LightGBM: A highly efficient gradient boosting decision tree. Paper presented at: *Advances in Neural Information Processing Systems* 3146-3154 (2017).
21. Hyland SL, *et al*. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat Med*. **26**, 364-373 (2020).
22. Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J. Liblinear: A library for large linear classification. *JMLR.* **9**, 1871–1874 (2008).
23. Chang C-C, Lin C-J. Libsvm: A library for support vector machines. *ACM Trans Intel Syst Technol.***2,** 1–27 (2011).
24. Chen, T., & Guestrin, C. Xgboost: A scalable tree boosting system. Paper presented at: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*785-794 (2016).
25. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* **521,** 436–444 (2015).
26. Abadi M, *et al*. Tensorflow: A system for large-scale machine learning. Paper presented at: *12th USENIX symposium on operating systems design and implementation (OSDI 16*) 265-283 (2016).
27. Pedregosa F, *et al*. Scikit-learn: Machine learning in Python. *JMLR.* **12**, 2825-2830 (2011).
28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics. **44,*** 837–845(1988).
29. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol.* **63**, 938-939 (2010).
30. Li X, *et al*. Integrated machine learning approaches for predicting ischemic stroke and thromboembolism in atrial fibrillation. In *AMIA Annual Symposium Proceedings* 799 (2016).
31. Jamwal S, Sharma S. Vascular endothelium dysfunction: a conservative target in metabolic disorders. *Inflamm Res.* **67**, 391-405 (2018)..
32. Tureyen K, Bowen K, Liang J, Dempsey RJ, Vemuganti R. Exacerbated brain damage, edema and inflammation in type-2 diabetic mice subjected to focal ischemia. *J Neurochem.* **116**, 499-507 (2011).
33. Barber M, Langhorne P, Rumley A, Lowe GD, Stott DJ. Hemostatic function and progressing ischemic stroke: D-dimer predicts early clinical progression. *Stroke* **35**, 1421-1425 (2004).

34. Martin AJ, Price CI. A systematic review and meta-analysis of molecular biomarkers associated with early neurological deterioration following acute stroke. *Cerebrovascular Diseases* **46**, 230-241 (2018).
35. Donkel, S. J., Benaddi, B., Dippel, D. W., Ten Cate, H., & de Maat, M. P., Prognostic hemostasis biomarkers in acute ischemic stroke: a systematic review. *Arteriosclerosis, thrombosis, and vascular biology.* **39**, 360-372(2019).
36. Kim J-T, *et al*. MRI findings may predict early neurologic deterioration in acute minor stroke or transient ischemic attack due to intracranial atherosclerosis. *Eur Neurol.* **64**, 95-100 (2010).
37. Kim J, *et al*. Serum alkaline phosphatase and phosphate in cerebral atherosclerosis and functional outcomes after cerebral infarction. *Stroke.* **44**, 3547-3549 (2013).
38. Uehara T, Yoshida K, Terasawa H, Shimizu H, Kita Y. Increased serum alkaline phosphatase and early neurological deterioration in patients with atherothrombotic brain infarction attributable to intracranial atherosclerosis. *eNeurologicalSci.* **20**, 100253 (2020).
39. Lee S-J, Lee D-G. Distribution of atherosclerotic stenosis determining early neurologic deterioration in acute ischemic stroke. *PLoS One.* **12**, e0185314 (2017).
40. Geng H-H, *et al*. Early neurological deterioration during the acute phase as a predictor of long-term outcome after first-ever ischemic stroke. *Medicine (Baltimore).* **96,** e9068(2017).
41. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes-2020*. *Diabetes care*. **43**, S14–S31(2020).
42. Samai AA, Albright K, Navalkele D, Alemayehu C, Martin-Schild S. Abstract WP272: Left Atrial Enlargement is Associated With Neuroworsening and Worse Short-Term Outcomes in Acute Ischemic Stroke. *Stroke* **50**, AWP272-AWP272 (2019).

Figures

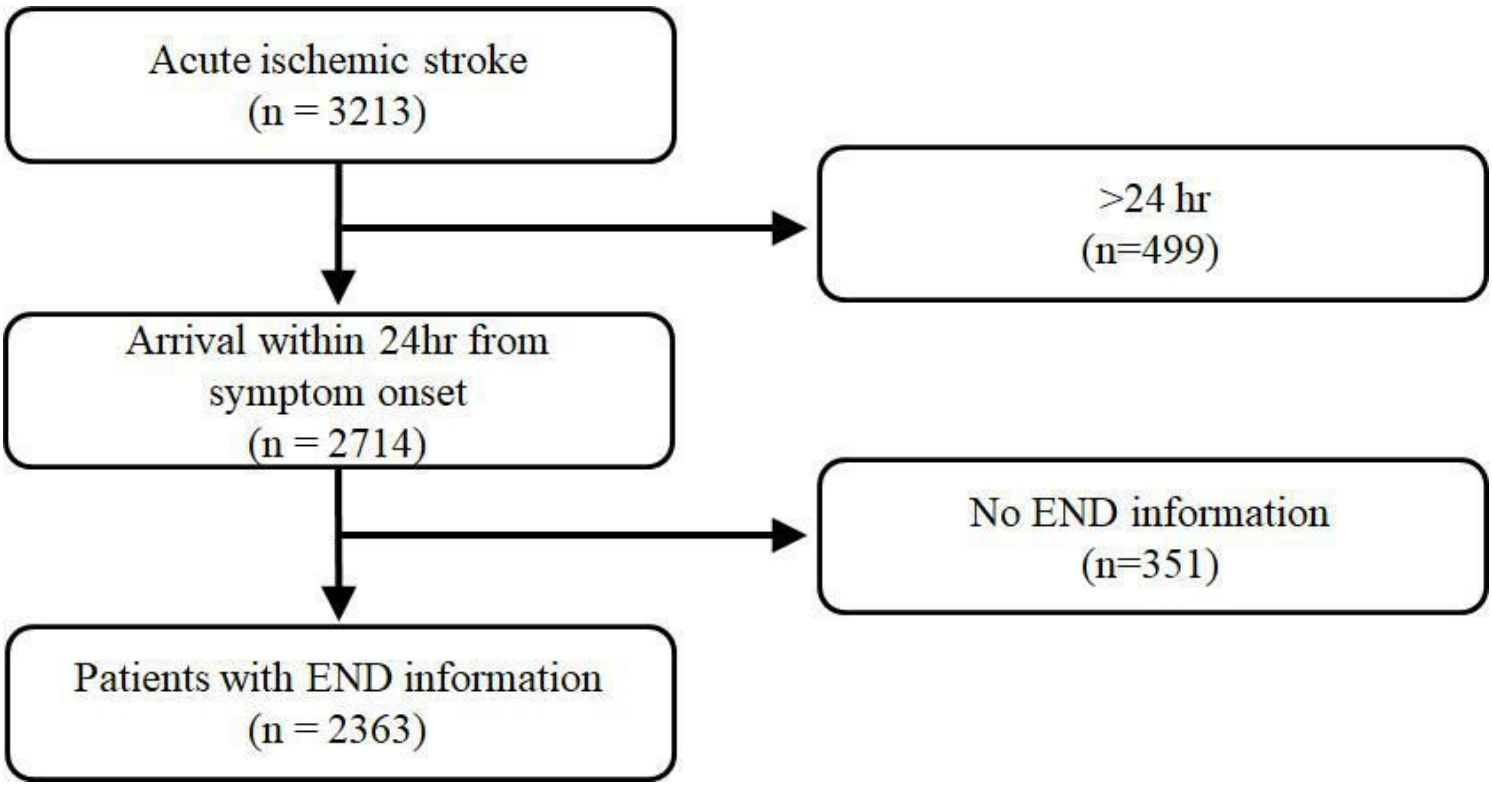

Figure 1

Flowchart of included patients.

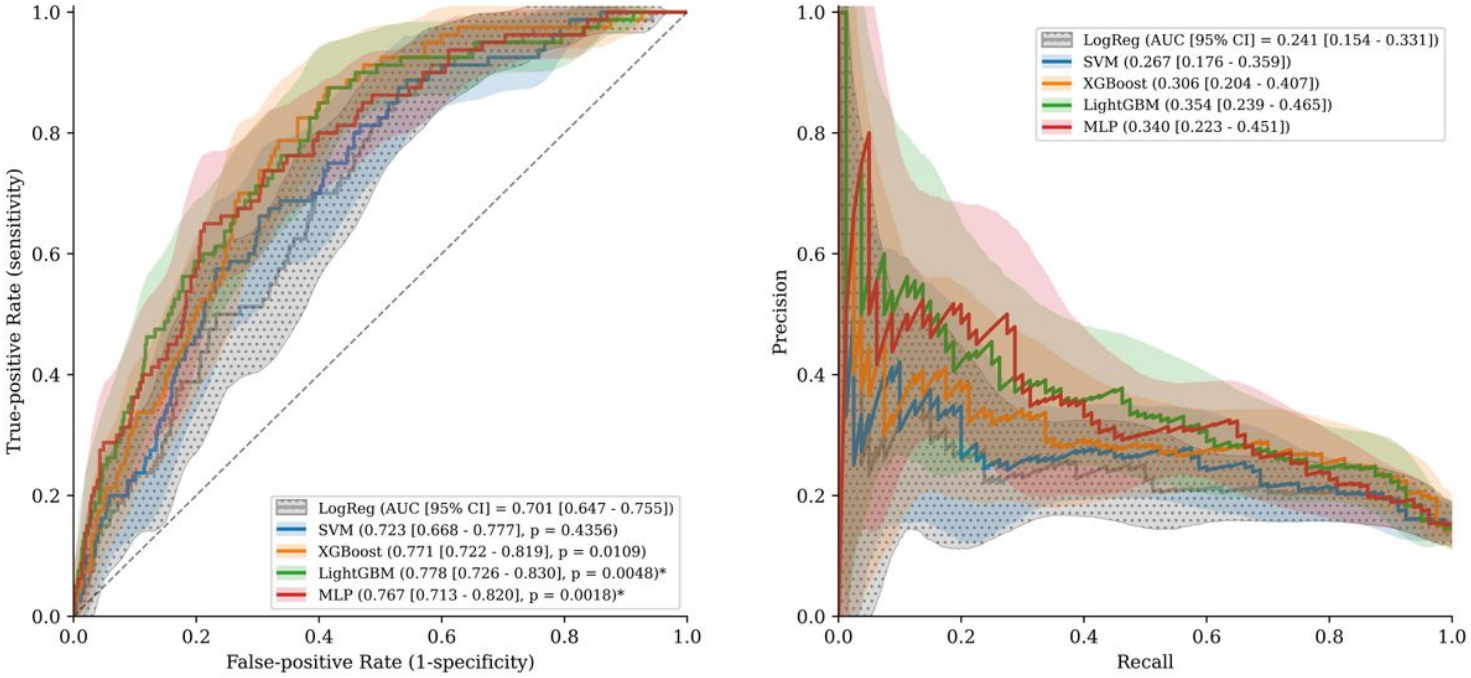

Figure 2

Receiver operating characteristic curve and precision recall curve between machine learning models. The shaded area indicates the 95% confidence interval. Abbreviations: AUC, area under the curve; CI, confidence interval; LogReg, Logistic regression; SVM, Support vector machine; XGBoost, Extreme gradient boosting; LightGBM, light gradient boosting machine; MLP, Multilayer perceptron

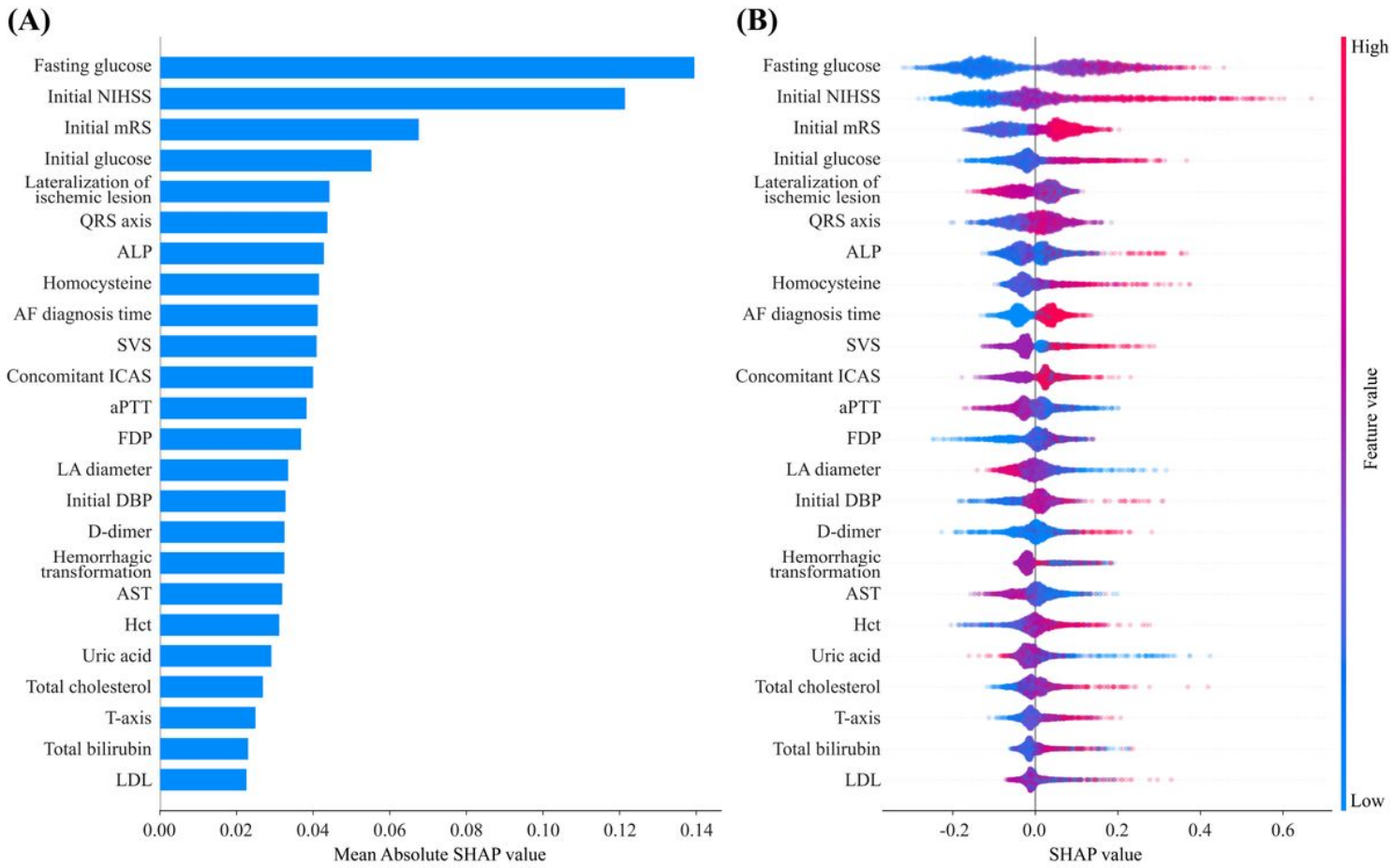

Figure 3

Matrix plots of top 24 important features. Bar plot (A) and violin plot (B). In the bar plot, the SHAP value implies the degree of contribution of a specific feature. The higher the SHAP value, the larger the model contribution of a specific feature. In the violin plot, each dot represents one patient and accumulates vertically to depict the density. The color reflects the high and low values of each feature, with the red color indicating a higher value and the blue color indicating a lower value. The X-axis of the graph represents the SHAP value, and a positive SHAP value indicates that it contributes positively to predicting the model, and that the probability of END occurring is high, and vice versa. Abbreviations: NIHSS, National Institute of Health Stroke Scale; mRS, modified Rankin scale; ALP, alkaline phosphatase; AF, atrial fibrillation; SVS, susceptibility vessel sign; ICAS, intracranial atherosclerosis; aPTT, activated partial thromboplastin time; FDP, fibrin degradation product; LA, left atrium; DBP, diastolic blood pressure; AST, aspartate aminotransferase; Hct, hematocrit; LDL, low-density lipoprotein

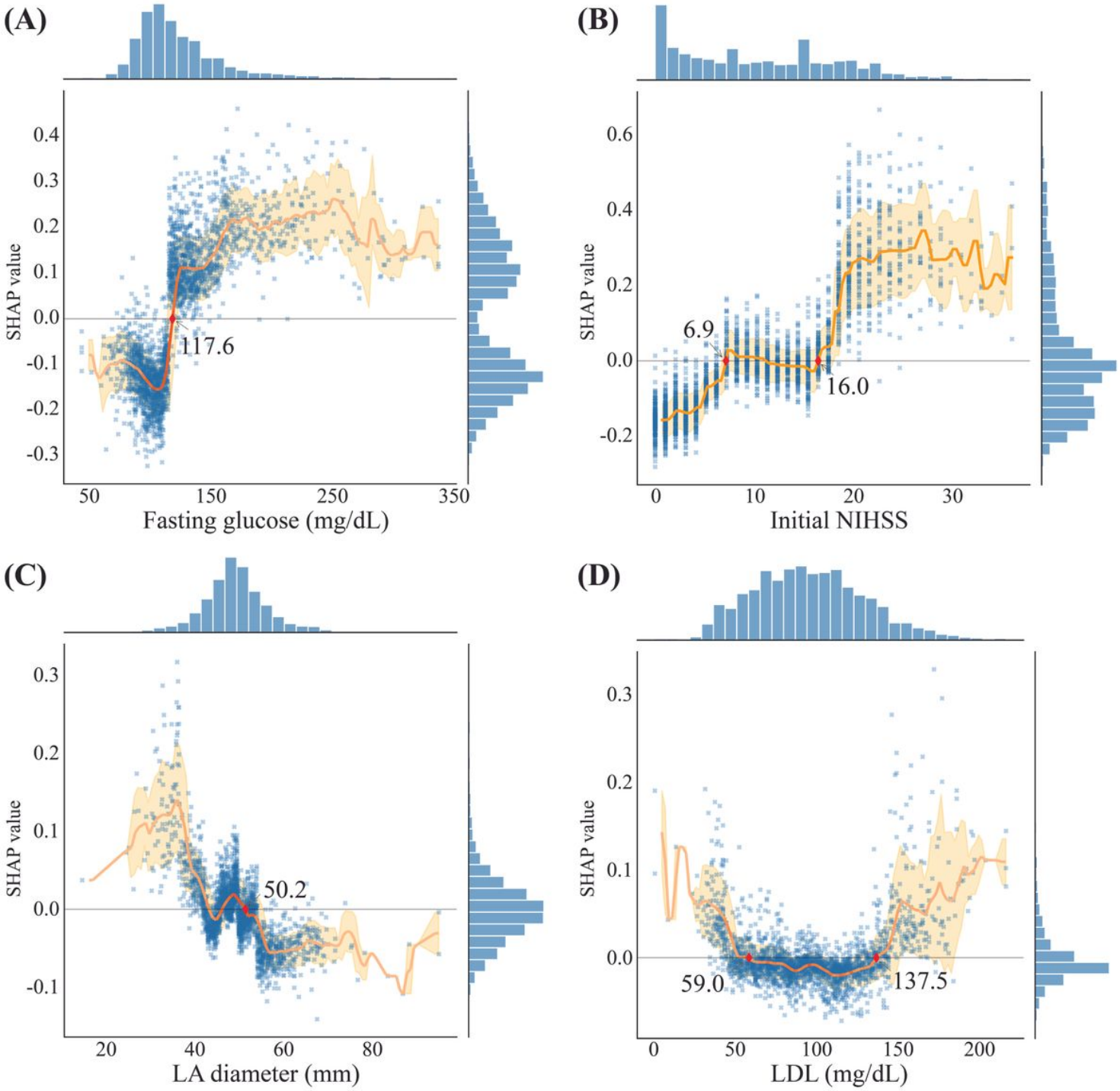

Figure 4

Partial SHAP dependence plot of the four representative features. Trend graph of each feature and SHAP value. Each dot represents one case. The bar graph at the upper right margins depicts the density. The yellow line represents the average value trend line. Red dots represent the cut-off values of the trend line. Abbreviations: NIHSS, National Institute of Health Stroke Scale; LA, left atrium; LDL, low-density lipoprotein

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.pdf