



Machine Learning-Based Model for Predicting Incidence and Severity of Acute Ischemic Stroke in Anterior Circulation Large Vessel Occlusion

Junzhao Cui^{1†}, Jingyi Yang^{2†}, Kun Zhang¹, Guodong Xu³, Ruijie Zhao⁴, Xipeng Li⁴, Luji Liu¹, Yipu Zhu¹, Lixia Zhou⁵, Ping Yu¹, Lei Xu¹, Tong Li¹, Jing Tian¹, Pandi Zhao¹, Si Yuan¹, Qisong Wang¹, Li Guo¹ and Xiaoyun Liu^{1,6*}

¹ Department of Neurology, The Second Hospital of Hebei Medical University, Shijiazhuang, China, ² Department of Information Center, The Second Hospital of Hebei Medical University, Shijiazhuang, China, ³ Department of Neurology, Hebei Province People's Hospital, Shijiazhuang, China, ⁴ Department of Neurology, Xingtai People's Hospital, Xingtai, China, ⁵ Department of Medical Iconography, The Second Hospital of Hebei Medical University, Shijiazhuang, China, ⁶ Neuroscience Research Center, Medicine and Health Institute, Hebei Medical University, Shijiazhuang, China

OPEN ACCESS

Edited by:

Jiang Li,
Geisinger Medical Center,
United States

Reviewed by:

Durgesh Prasad Chaudhary,
Geisinger Health System,
United States
Akram Mohammed,
University of Tennessee Health
Science Center (UTHSC),
United States

*Correspondence:

Xiaoyun Liu
audrey-l@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Stroke,
a section of the journal
Frontiers in Neurology

Received: 29 July 2021

Accepted: 29 October 2021

Published: 02 December 2021

Citation:

Cui J, Yang J, Zhang K, Xu G, Zhao R,
Li X, Liu L, Zhu Y, Zhou L, Yu P, Xu L,
Li T, Tian J, Zhao P, Yuan S, Wang Q,
Guo L and Liu X (2021) Machine
Learning-Based Model for Predicting
Incidence and Severity of Acute
Ischemic Stroke in Anterior Circulation
Large Vessel Occlusion.
Front. Neurol. 12:749599.
doi: 10.3389/fneur.2021.749599

Objectives: Patients with anterior circulation large vessel occlusion are at high risk of acute ischemic stroke, which could be disabling or fatal. In this study, we applied machine learning to develop and validate two prediction models for acute ischemic stroke (Model 1) and severity of neurological impairment (Model 2), both caused by anterior circulation large vessel occlusion (AC-LVO), based on medical history and neuroimaging data of patients on admission.

Methods: A total of 1,100 patients with AC-LVO from the Second Hospital of Hebei Medical University in North China were enrolled, of which 713 patients presented with acute ischemic stroke (AIS) related to AC-LVO and 387 presented with the non-acute ischemic cerebrovascular event. Among patients with the non-acute ischemic cerebrovascular events, 173 with prior stroke or TIA were excluded. Finally, 927 patients with AC-LVO were entered into the derivation cohort. In the external validation cohort, 150 patients with AC-LVO from the Hebei Province People's Hospital, including 99 patients with AIS related to AC-LVO and 51 asymptomatic AC-LVO patients, were retrospectively reviewed. We developed four machine learning models [logistic regression (LR), regularized LR (RLR), support vector machine (SVM), and random forest (RF)], whose performance was internally validated using 5-fold cross-validation. The performance of each machine learning model for the area under the receiver operating characteristic curve (ROC-AUC) was compared and the variables of each algorithm were ranked.

Results: In model 1, among the included patients with AC-LVO, 713 (76.9%) and 99 (66%) suffered an acute ischemic stroke in the derivation and external validation cohorts, respectively. The ROC-AUC of LR, RLR and SVM were significantly higher than that of the RF in the external validation cohorts [0.66 (95% CI 0.57–0.74) for LR, 0.66 (95% CI 0.57–0.74) for RLR, 0.55 (95% CI 0.45–0.64) for RF and 0.67 (95% CI 0.58–0.76) for SVM]. In model 2, 254 (53.9%) and 31 (37.8%) patients suffered disabling ischemic stroke

in the derivation and external validation cohorts, respectively. There was no difference in AUC among the four machine learning algorithms in the external validation cohorts.

Conclusions: Machine learning methods with multiple clinical variables have the ability to predict acute ischemic stroke and the severity of neurological impairment in patients with AC-LVO.

Keywords: anterior circulation large vessel occlusion, acute ischemic stroke, machine learning, prediction model, neurological impairment

INTRODUCTION

Acute ischemic stroke caused by large vessel occlusion accounts for more than 40% of cases, ~80% of which occurs in the anterior circulation (1). Compared to non-large vessel occlusion (LVO) acute ischemic stroke (AIS), patients with anterior circulation large vessel occlusion (AC-LVO) stroke are considered to be at greater risk of mortality or disability before endovascular treatment (2). They tend to improve significantly after mechanical thrombectomy (3, 4). Previously reported prediction models for AC-LVO stroke such as prehospital scales (Prehospital Acute Stroke Severity scale, PASS; Cincinnati Prehospital Stroke Severity Scale, CPSSS; stroke Vision Aphasia Neglect, VAN; Rapid Arterial Occlusion Evaluation scale RACE and Field Assessment Stroke Tri-age for Emergency Destination, FAST-ED) (5–9) that are based on NIHSS, and the recently proposed model by Philipp Hendrix et al., which combines past medical history and neurologic examination (10), have focused on the identification of large vessel occlusion in patients with AIS. The main clinical purpose of the prediction scores is to identify which patients with AIS have LVO so that they can be referred to capable centers for endovascular treatment (EVT). However, accurate prediction of AIS in patients with AC-LVO remains a challenge.

Anterior circulation-LVO stroke can be further divided based on pathogenesis and severity of clinical consequences, into non-disabling and disabling stroke with the latter frequently resulting in post-stroke dependence. Nevertheless, no previous studies have predicted the risk of disabling ischemic stroke in patients with AC-LVO, which may be useful in treatment decisions and prevention.

In this study, we developed and validated two models based on machine learning algorithms with clinical variables, to predict acute ischemic stroke (Model 1) and severity of neurological impairment (Model 2) in patients with AC-LVO.

METHODS

Patient Cohorts

A total of 1,100 patients with AC-LVO admitted between June 2016 and April 2018 at the Second Hospital of Hebei Medical University, North China, were registered in the derivation cohort; 927 of them who presented with AIS related with AC-LVO and asymptomatic AC-LVO were retrospectively reviewed. In addition, 471 patients with first-ever ischemic stroke (including disabling and non-disabling stroke) were selected. For the

external validation, we collected data of patients with AC-LVO from Hebei Province People's Hospital, China between September 2016 and April 2021.

Anterior circulation-LVO was defined as complete occlusion of at least one intracranial internal carotid artery (ICA) or middle cerebral artery (MCA) visualized on computed tomography angiography (CTA) or magnetic resonance angiography (MRA). ICA occlusion refers to the complete occlusion of the C1–C7 segment of the internal carotid artery based on CTA or MRA. MCA occlusion refers to the occlusion of the MCA involving at least the M1 segment (for more details please see in **Supplementary Figure I**). Asymptomatic AC-LVO was defined as the absence of a transient ischemic attack (TIA), amaurosis fugax, and ischemic stroke attributed to anterior circulation large vessel (11, 12). In accordance with previous studies, disabling and non-disabling ischemic strokes were defined by the initial clinician as National Institutes of Health Stroke Scale (NIHSS) > 5 and ≤ 5 on admission, respectively (13).

Data Collection and Variable Selection

Patient characteristics that were collected on admission for the development of Models 1 and 2 include (1) demographic data of the patients such as the age, sex, body mass index (BMI), current smoking and drinking status, comorbidity (hypertension, coronary atherosclerotic heart disease, atrial fibrillation, diabetes mellitus, and hyperlipidemia), history of transient ischemic attack (TIA); (2) clinical variables such as serum apolipoprotein B (Apo B) and homocysteine on arrival; (3) imaging variables such as occluded vessels (unilateral MCA, unilateral ICA, and multiple arteries), posterior circulation large vessel severe stenosis ($\geq 70\%$) /occlusion, anterior cerebral artery (ACA) occlusion, and Alberta Stroke Program Early CT Score (ASPECTS). Data on 14 variables were included in Model 1, and on 12 in Model 2. Specifically, normal blood flow status of the vertebrobasilar arteries *via* the posterior communicating artery plays a major role in primary collateral compensation after anterior circulation large vessel occlusion. Therefore, posterior circulation large vessel stenosis/occlusion was introduced into Model 1. Posterior circulation large vessel refers to the intracranial vertebral artery, basilar artery, or segment P1 of the posterior cerebral artery.

Data Pre-processing

Processing of the data was performed using Python. First, records containing outliers, which were identified by boxplot, were excluded. Furthermore, the median imputation method was used to impute missing values in derivation cohorts. Finally, the

categorical variables were converted into numerical values with dummy encoding, and the continuous features were standardized by removing the mean and scaling to unit variance.

Prediction Models With Machine Learning

Machine learning is a discipline that constructs models base on data, which is a part of artificial intelligence. Machine learning extracts the characteristics and abstracts the model of the data, discovers the information in the data, and then analyzes and predicts it. First, an algorithm and some parameters of the model which were supplied with training data were selected arbitrarily. During training procedures, the model automatically adjusts some trainable parameters stage by stage to achieve better performance optimization. After the training, all the model parameters are fixed. Importantly, the true effectiveness of the model was evaluated using test data that were completely separate from the training data.

We selected logistic regression without regularization (LR), regularized logistic regression (RLR), random forest (RF), and support vector machine (SVM) as machine learning algorithms that are commonly used.

Logistic regression, a classic classification algorithm in machine learning, was regularized using a combination of L1 and L2 loss in this study. Here, the target was determined by Y:

$$Y = \{ \text{"disabling ischemic stroke,"} \\ \text{"non-disabling ischemic stroke"} \}$$

$$Z = W^T X + b$$

$$y = \frac{1}{1 + e^{-Z}} = \frac{1}{1 + e^{-W^T X + b}}$$

We selected binary cross-entropy loss as the cost function, where y is the ground truth, \hat{y} is the predicted score of the model, and R represents the regularization. The loss functions L1 and L2 are defined as follows:

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}^{(i)}, y^{(i)})$$

$$= \frac{1}{m} \sum_{i=1}^m \left(-y^{(i)} \log \hat{y}^{(i)} - (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}) \right) + \frac{\lambda}{2} R$$

$$R_{L1} = \sum_{j=1}^m |W_j| \dots R_{L2} = \sum_{j=1}^m [W_j^2]$$

In the training process of the model, standardization of numerical variables was carried out to accelerate the convergence process and speed of the model.

Random forest is an extended variant of bagging, which uses a decision tree as the base learner and introduces the selection of random attributes in the training process of the decision tree (14). The main parameters that can affect the model performance in RF include the number of trees in the forest, maximum depth of the tree, minimum number of samples required to split an internal node, minimum number of samples required to be at a leaf node, and function to measure the quality of a split. In this study, the values in the dataset were discretized, and

the parameters were optimized with a grid search during the training process.

An SVM classifies data by calculating the maximum-margin hyperplane, which adds a regularization term in the solving process to optimize the structural risk. The strength of SVM is that it can process complex datasets with many variables or dimensions (15). The validity of SVM depends mainly on the selection of the kernel function, parameters of the kernel, and soft margin parameter C. Otherwise, in this study, each combination of parameter selections was checked using cross-validation, and only parameters with optimal accuracy were selected.

Moreover, LR, RLR, RF, and SVM can estimate the contribution of each feature to the model by calculating the absolute value of the standardized regression coefficient, information gain / Gini coefficient, and weight coefficient.

Model Derivation and Internal Validation

In this study, for model derivation, we adopted 5-fold cross-validation, which is a standard way of optimizing the model with inner test data and has been used in a previous study (16). During modeling, the grid search algorithm which is a greedy algorithm was combined to tune and optimize the model hyperparameters. For each group of hyperparameters, we selected 5-fold cross-validation to determine the optimal ones, after which we calculated the means of sensitivity, specificity, accuracy, and AUC to evaluate the performance of each model (Figure 1).

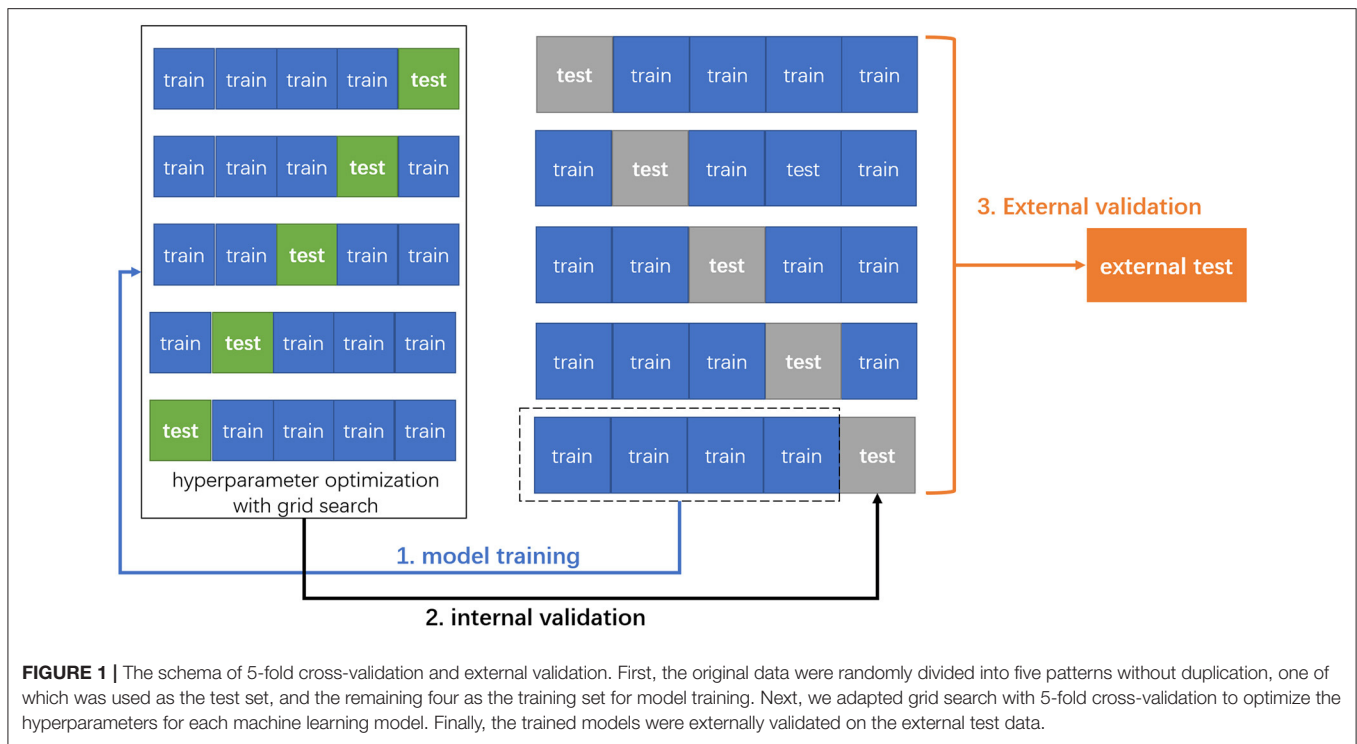
The derivation and validation models were conducted using Python 3.6. The model algorithms, cross-validation, and grid search were based on the Scikit-Learn library of Python in the PyCharm. Matplotlib 3.3.3, NumPy 1.19.5, pandas version 1.1.5, and Scikit-Learn toolkit version 0.21.0 were used to train the machine learning models.

External Validation

After internal training and testing, the performance of the model was evaluated using external validation data. Subsequently, the AUCs were compared among machine learning algorithms.

Statistical Analysis

Clinical variables are presented as mean \pm SD or median with interquartile range, depending on the distribution of the variables. To compare the group differences, continuous variables were compared using the Student's *t*-test or Mann-Whitney *U* test, and categorical variables were compared with the χ^2 test or Fisher's exact test. These two prediction models were discriminated against using AUC. Calculation of AUC, sensitivity, specificity, precision, negative predictive value (NPV), and accuracy criteria were performed with R statistical software version 3.0.2. The area under the precision-recall curve (PRC) and F1-score were calculated with MedCalc. For the derivation and validation cohorts, a comparison of AUC among the machine learning methods was performed using the DeLong test with Bonferroni correction. Two-sided $P < 0.05$ were considered statistically significant.



RESULTS

Baseline Characteristics

Figures 2A,B illustrate the flow diagram of the enrolled patients. For the derivation cohort, 1,100 patients with AC-LVO were hospitalized at the study institution. After excluding 173 patients with prior stroke or TIA as non-acute ischemic cerebrovascular events, 713 patients with AIS related with AC-LAO and 214 with asymptomatic AC-LAO were finally included in the analysis for model 1. Among the 214 patients with asymptomatic AC-LVO, 119 (56%) were hospitalized for head discomfort such as heaviness of the head and fullness in the head. The other reasons for hospitalization in patients with asymptomatic AC-LVO included coronary artery disease, subarachnoid hemorrhage, migraine, cerebral large artery disease detected by routine physical examination, unruptured intracranial aneurysms, diabetic peripheral vascular disease, central nervous system infection, lower extremity atherosclerotic occlusive disease, intracranial space-occupying lesions, epilepsy, Parkinson's disease, cerebral atrophy, subclavian artery steal blood syndrome, cough syncope, and cardiac syncope. The general screening of large artery disease was performed with transcranial Doppler and carotid artery ultrasound in these patients. Further computed tomography angiography (CTA) or magnetic resonance angiography (MRA) examinations were conducted and AC-LVO was identified. Among the 713 patients with AIS, 242 with prior stroke were excluded, and 471 patients with first-ever ischemic stroke (254 with disabling and 217 with non-disabling strokes) were included in the analysis for Model 2. For the external validation cohort, 150 eligible patients with AC-LVO were included in Model 1. Of the 99

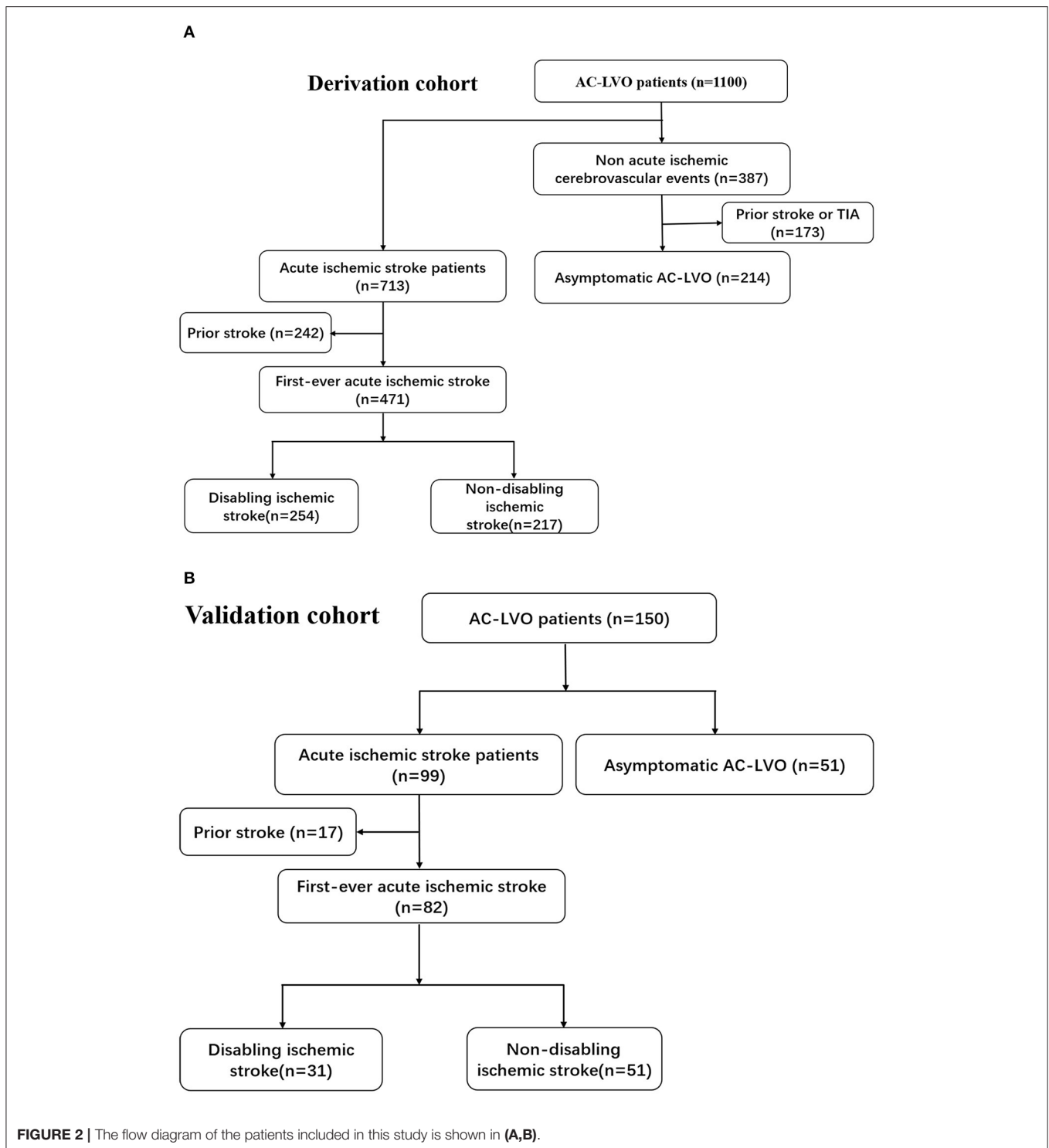
patients with AIS, 82 who presented with the first episode were included in the analysis for model 2. The baseline characteristics of the included patients are presented in Tables 1, 2, and Supplementary Tables I–IV.

Comparison Between the Models in the Derivation Cohort

The performance metrics of each approach for Models 1 and 2 in the derivation cohort are shown in Tables 3, 4, respectively. The receiver operating characteristic (ROC) curve (indicating the predictive performance of our LR/RLR/RF/SVM model) for each algorithm in the two models and the comparison among these machine learning algorithms are shown in Figures 4A,C. In model 1, the AUCs of RF and SVM were significantly higher than those of the LR and RLR, when using the DeLong test with Bonferroni correction (RF vs. LR, $P < 0.0001$; RF vs. RLR, $P < 0.0001$; SVM vs. LR, $P < 0.0001$; SVM vs. RLR, $P < 0.0001$; Figure 3A). Similar results were obtained for accuracy and F1-score. In model 2, while the differences in AUCs among the four machine learning algorithms were not significant (Figure 3C), the RF showed the most perfect classification accuracy (71.8%) compared to that of the other machine learning approaches.

Comparison Between the Models in the External Validation Cohort

The ROC curves for Models 1 and 2 in the external validation cohort are shown in Figures 4B,D. In Model 1, RF exhibited the worst performance among the machine learning models (Table 5). The AUCs in LR, RLR and SVM were significantly



higher than that in RF, when using the Delong test with Bonferroni correction (LR vs. RF, $P = 0.0048$; RLR vs. RF, $P = 0.0048$, SVM vs. RF, $P = 0.0006$; **Figure 3B**). In Model 2, there was no difference in AUCs among the four machine learning algorithms (**Figure 3D**). The AUC of each algorithm was as follows: LR 0.68 (95% CI 0.56–0.8), RLR 0.76 (95% CI 0.66–0.87), RF 0.71 (95% CI 0.59–0.83) and SVM 0.77 (95% CI 0.66–0.87) (**Table 6**).

Important Variables of the Machine Learning Models

After calculating the importance of each feature, the top five selected variables of Models 1 and 2 were ranked by their discriminative performance (**Figures 5, 6**). For LR and RLR, the absolute value of the standardized regression coefficient was calculated in both models. For RF, the important features for information gain and Gini coefficient were ranked in Models 1

TABLE 1 | Baseline characteristics of patients with anterior circulation large vessel occlusion.

	Derivation cohort (<i>n</i> = 927)	Validation cohort (<i>n</i> = 150)
Mean ± SD age, y	59.3 ± 12.4	62.0 ± 10.9
Males <i>n</i> (%)	626 (67.5)	108 (72.0)
Current smoking <i>n</i> (%)	268 (28.9)	57 (38.0)
Drinking <i>n</i> (%)	267 (28.8)	26 (17.3)
Hypertension <i>n</i> (%)	602 (64.9)	107 (71.3)
Coronary atherosclerotic heart disease <i>n</i> (%)	143 (15.4)	22 (14.7)
Atrial fibrillation <i>n</i> (%)	52 (5.6)	11 (7.3)
Diabetes <i>n</i> (%)	225 (24.3)	52 (34.7)
Hyperlipidaemia <i>n</i> (%)	292 (31.5)	14 (9.3)
Occluded vessels <i>n</i> (%)	442 (47.7)	51 (34.0)
Unilateral MCA		
Unilateral ICA	261 (28.2)	69 (46.0)
Multiple artery	224 (24.2)	30 (20.0)
BMI ≥ 24 <i>n</i> (%)	795 (85.8)	108 (72.0)
Posterior circulation large vessel severe stenosis /occlusion <i>n</i> (%)	176 (19.0)	9 (6.0)
ApoB (g/L)	1.0 ± 0.3	0.7 ± 0.2
Homocysteine (μmol/L)	19.0 ± 11.2	15.9 ± 9.6
Acute ischemic stroke <i>n</i> (%)	713 (76.9)	99 (66.0)

ApoB, apolipoprotein B; BMI, body mass index; ICA, internal carotid artery; IQR, interquartile range; MCA, middle cerebral artery; SD, standard deviation.

TABLE 2 | Baseline characteristics of patients with first-ever acute ischemic stroke (AIS) caused by anterior circulation large vessel occlusion.

	Derivation cohort (<i>n</i> = 471)	Validation cohort (<i>n</i> = 82)
Mean ± SD age, y	58.6 ± 12.7	61.0 ± 13.0
Males <i>n</i> (%)	321 (68.2)	59 (72.0)
Current smoking <i>n</i> (%)	147 (31.2)	35 (42.7)
Hypertension <i>n</i> (%)	290 (61.6)	51 (62.2)
Diabetes <i>n</i> (%)	105 (22.3)	25 (30.5)
Coronary atherosclerotic heart disease <i>n</i> (%)	69 (14.6)	7 (8.5)
Prior TIA <i>n</i> (%)	30 (6.4)	0 (0)
Hyperlipidaemia <i>n</i> (%)	158 (33.5)	8 (9.8)
BMI ≥ 24 <i>n</i> (%)	410 (87.0)	60 (73.2)
Median ASPECTS (IQR)	6 (3–8)	8 (7–9)
Occluded vessels <i>n</i> (%)	236 (50.1)	31 (37.8)
Unilateral MCA		
Unilateral ICA	133 (28.2)	36 (43.9)
Multiple artery	102 (21.7)	15 (18.3)
Anterior cerebral artery occlusion <i>n</i> (%)	87 (18.5)	3 (3.7)
Disabling ischemic stroke (NIHSS > 5) <i>n</i> (%)	254 (53.9)	31 (37.8)

ASPECTS, Alberta Stroke Program Early CT Score; BMI, body mass index; ICA, internal carotid artery; IQR, interquartile range; MCA, middle cerebral artery; NIHSS, National Institutes of Health Stroke Scale; SD, standard deviation; TIA, transient ischemic attack.

and 2 respectively. For SVM, the absolute value of the weight was used to rank the variables only in model 2 due to the introduction of the kernel function in Model 1. The absolute values of the important metrics for the features were normalized, ensuring the comparability in feature importance ranking. In Model 1, homocysteine, occluded vessels and BMI appeared together in the top five rankings of all machine learning algorithms. In addition, coronary atherosclerotic heart disease was an important feature in both LR and RLR. Age and Apo B appeared to be important variables in RF. In Model 2, ASPECT, age and BMI were common variables for all machine learning algorithms. Prior TIA was included in LR, RLR, and RF. Hypertension, current smoking, and gender appeared in RLR, RF, and SVM, respectively. Furthermore, occluded vessels coexisted in LR and SVM.

DISCUSSION

This study demonstrated that the use of a machine learning approach can predict the risk of AIS and severity of ischemic stroke in AC-LVO from clinical data. To the best of our knowledge, this is the first report on an attempt to predict AIS and severity of neurological impairment in patients with AC-LVO using the machine learning approach. The machine learning algorithm can eliminate linearity and has various ways of overcoming the imperfections of the polyfactorial models such as overfitting of models and collinearity of variables, which may lead

to a series of problems when it comes to variable selection (17). In the two prediction models in this study, 14 and 12 common variables were collected, respectively, bypassing the traditional method of variable selection.

Contrary to the findings in the derivation cohort of model 1 that RF showed significantly better predictive performance than LR and RLR, in the validation cohort, RF had the worst performance among the machine learning models. The decision trees of RF forced interactions between the features, which might make the result rather inferior if the majority of the features have no or very weak interactions. Therefore, we suspect that the RF was not able to carry on an accurate classified forecast owing to extremely weak interactions between the variables in our dataset. Moreover, the small data sets with 150 cases in the validation cohort may be another reason for the poor performance of the RF. In model 2, although the LR showed a predictive property similar to those of the other three algorithms both in the validation cohort and derivation cohort, the RLR exhibited a higher AUC compared with LR in the validation cohort, this was as a result of the poor generalization performance of LR compared with other algorithms. Accordingly, LR with L2 regularization was implemented in this study to avoid overfitting and improve the generalization performance and robustness of the model; thus, a more optimal result was obtained with an AUC of 0.76.

As shown in **Figures 5, 6**, the important features were not entirely consistent in the machine learning algorithms in

TABLE 3 | Scores for each algorithm of model 1 in derivation cohort.

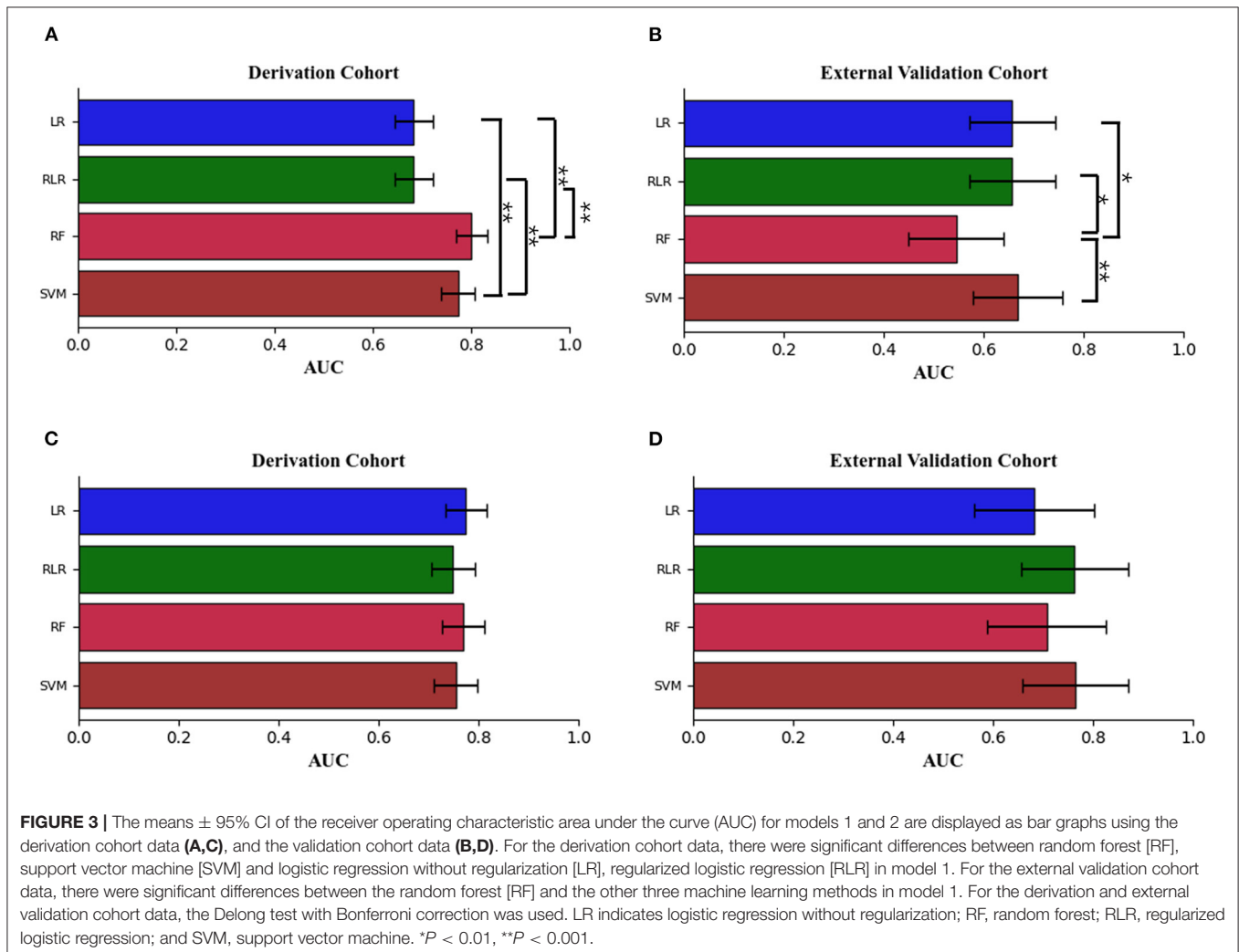
Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	F1 _{max}
LR	0.68 (0.64–0.72)	0.88	57.9	71.5	87.1	33.8	61.1	0.87
RLR	0.68 (0.64–0.72)	0.88	57.9	71.5	87.1	33.8	61.1	0.87
RF	0.80 (0.77–0.83)	0.93	69.1	79.9	92.0	43.7	71.6	0.89
SVM	0.77 (0.74–0.81)	0.92	76.9	68.2	89.0	46.9	74.9	0.88

AUC, receiver operator characteristic area under the curve; F1_{max}, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.

TABLE 4 | Scores for each algorithm of model 2 in derivation cohort.

Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	F1 _{max}
LR	0.78 (0.74–0.81)	0.78	63.0	77.9	76.9	64.3	69.9	0.76
RLR	0.75 (0.71–0.79)	0.77	61.0	78.3	76.7	63.2	69.0	0.74
RF	0.77 (0.73–0.81)	0.78	67.7	76.5	77.1	66.9	71.8	0.75
SVM	0.76 (0.71–0.80)	0.78	63.4	77.0	76.3	64.2	69.6	0.74

AUC, receiver operator characteristic area under the curve; F1_{max}, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.



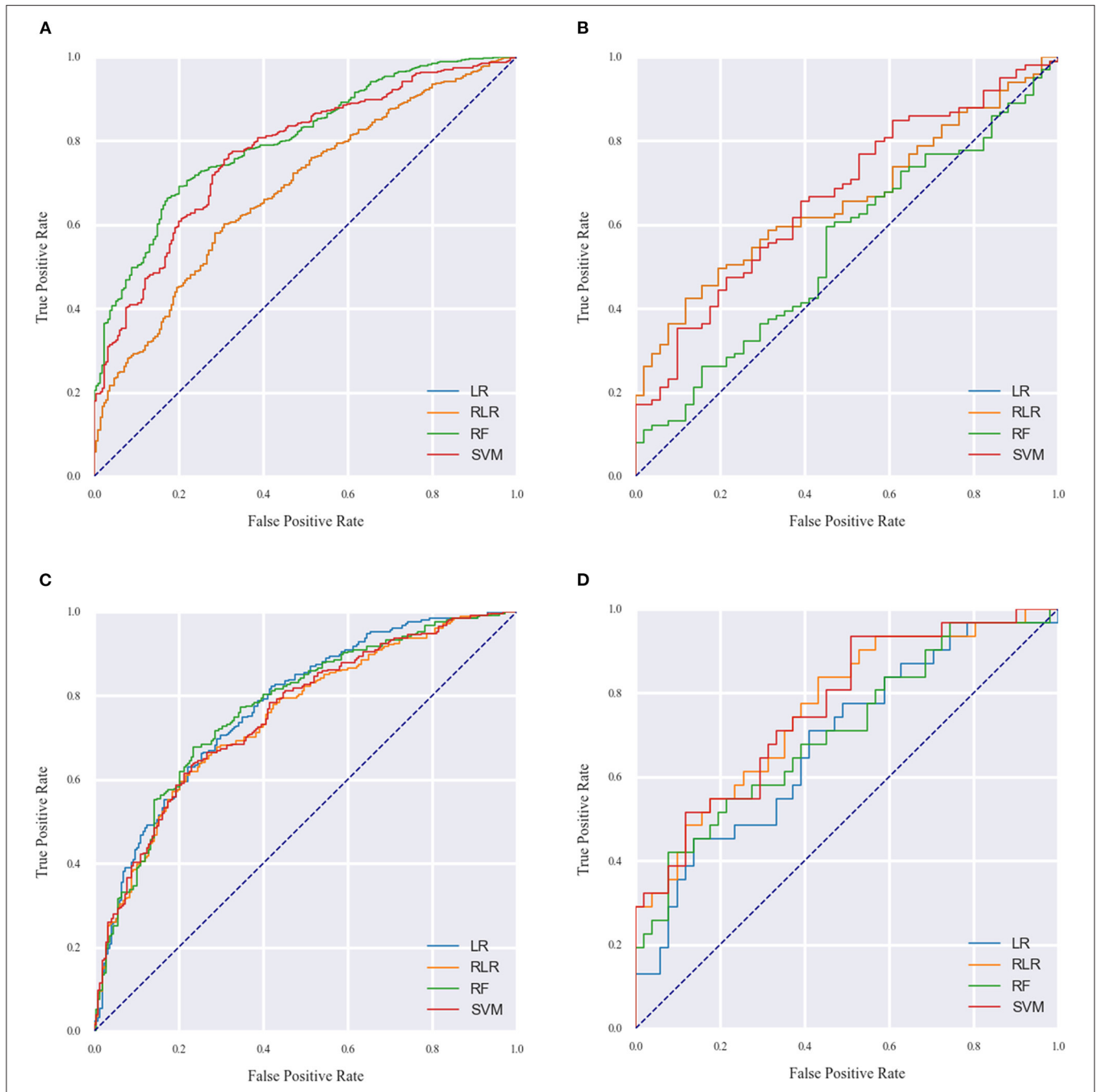


FIGURE 4 | The AUC of the machine learning models for model 1 (A,B) and model 2 (C,D) on the derivation and external validation cohort data. LR indicates logistic regression without regularization; RF, random forest; RLR, regularized logistic regression, SVM, support vector machine.

model 1 and model 2. As important variables of model 1, homocysteine, BMI, and occluded vessels (unilateral MCA) appeared in all three algorithms, and atrial fibrillation and coronary atherosclerotic heart disease were detected in both LR and RLR. Elevated blood homocysteine concentration increases the risk of ischemic stroke by inducing oxidative damage to

vascular endothelial cells and enhancing platelet adhesion to endothelial cells, especially in large vessel strokes (18–22). The results of our study are in accordance with the aforementioned studies, suggesting that elevated homocysteine levels may be a significant marker for predicting ischemic stroke in AC-LVO.

TABLE 5 | Scores for each algorithm of model 1 in external validation cohort.

Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	F1 _{max}
LR	0.66 (0.57–0.74)	0.82	42.4	88.2	87.5	44.1	58.0	0.80
RLR	0.66 (0.57–0.74)	0.82	42.4	88.2	87.5	44.1	58.0	0.80
RF	0.55 (0.45–0.64)	0.72	59.6	54.9	72.0	41.2	58.0	0.80
SVM	0.67 (0.58–0.76)	0.81	65.7	60.8	76.5	47.7	64.0	0.80

AUC, receiver operator characteristic area under the curve; F1_{max}, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.

TABLE 6 | Scores for each algorithm of model 2 in external validation cohort.

Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	F1 _{max}
LR	0.68 (0.56–0.80)	0.60	45.2	86.3	66.7	72.1	70.7	0.60
RLR	0.76 (0.66–0.87)	0.71	83.9	56.9	54.2	85.3	67.1	0.66
RF	0.71 (0.59–0.83)	0.65	41.9	92.2	76.5	72.3	73.2	0.61
SVM	0.77 (0.66–0.87)	0.71	93.5	49.0	52.7	92.6	65.9	0.67

AUC, receiver operator characteristic area under the curve; F1_{max}, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.

Regarding the association between BMI and ischemic stroke, a previous meta-analysis revealed a J-shaped dose-response relationship between being overweight or obese and an increased risk of incident ischemic stroke (23). However, few studies have focused on the relationship between BMI and risks of ischemic stroke subtype (24, 25). Our study showed a robust positive association between overweight/obesity and AC-LVO AIS. Possible explanations for our findings include insulin resistance, endothelial dysfunction, and inflammation, which have been considered to influence the relationship between obesity and atherosclerosis (26). Moreover, our findings further revealed that a high BMI (≥ 24 kg/m²) shows a greater predisposes to disabling than non-disabling ischemic stroke with AC-LVO, emphasizing the importance of weight control and aerobic fitness.

The compensation of the collateral pathway in MCA occlusion mainly depends on the pia meningeal branch from the anterior cerebral artery and the posterior cerebral artery with worse compensatory ability than the circle of Willis, which means it would result in hemodynamic failure and is more prone to decompensation (27). Our study delves deeper into this field and demonstrates that unilateral MCA occlusion plays a crucial role in the occurrence of ischemic stroke. Furthermore, we found that stroke severity at admission was greater in the multiple AC-LAO patients than in unilateral MCA occlusion or unilateral ICA occlusion patients. This is consistent with a previously published study of patients with AC-LVO AIS, which showed that high NIHSS was associated with multiple AC-LAO (28).

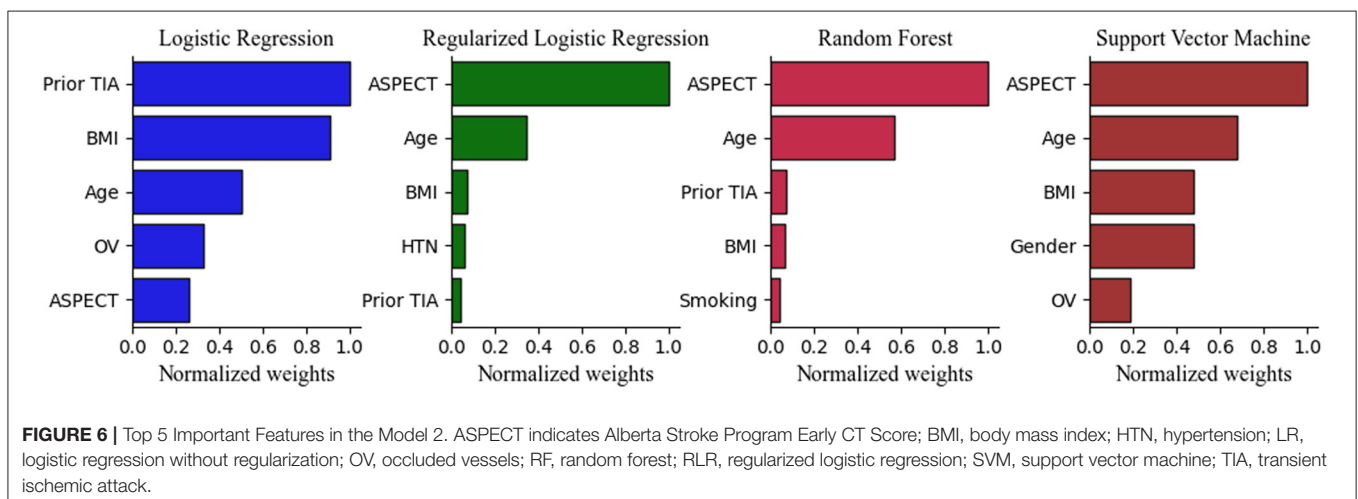
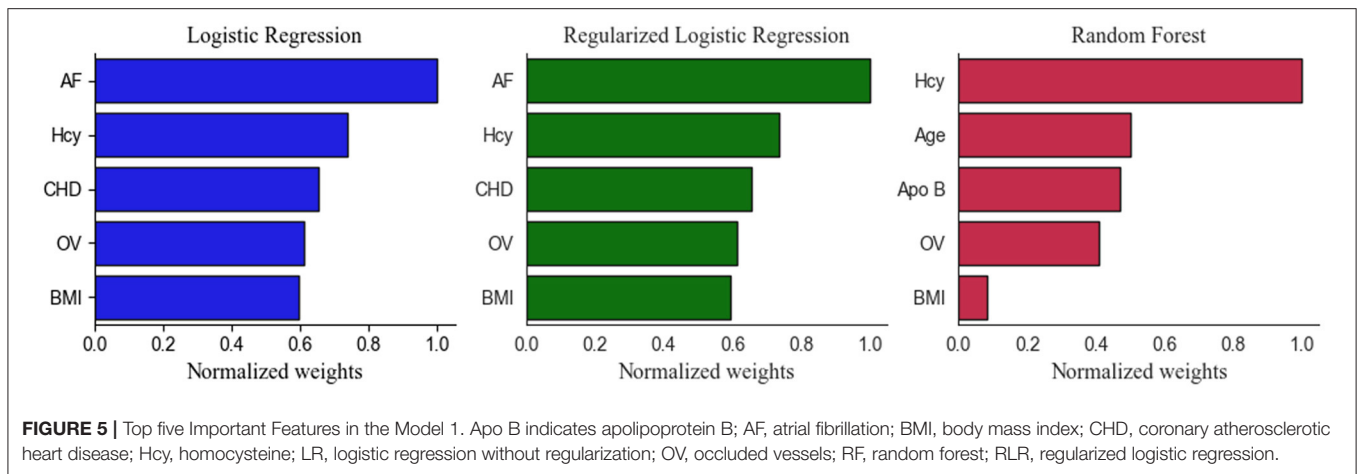
Cardioembolism might be responsible for large vessel occlusion, in which atrial fibrillation accounts for ~50% (29, 30). Atrial fibrillation is strongly associated with a high occurrence rate of LVO, suggesting that it may be a potential risk factor for LVO (31). Otherwise, large emboli that block intracranial vessels usually originate from the left atrial appendage in

patients with symptomatic carotid stenosis or atrial fibrillation (32). Similarly, in our analysis, atrial fibrillation showed a robust association with AC-LVO AIS, further suggesting that knowledge of the potential complications of atrial fibrillation is likely to motivate both patient and clinician to comply with standard treatment.

Large-artery atherosclerotic stroke is associated with a high risk of coronary atherosclerotic heart disease (33). Nevertheless, our results indicate that coronary atherosclerotic heart disease is associated with a low risk of AIS in AC-LVO patients. One explanation for this finding might be that coronary atherosclerosis is significantly correlated with stenosis of the extracranial carotid; therefore, the development of intracranial anterior circulation large vessel occlusion may be independent of coronary atherosclerotic heart disease (34). Furthermore, antiplatelet and statin therapy in coronary atherosclerotic heart disease may reduce the risk of ischemic stroke in AC-LVO.

Apolipoprotein B is the primary apolipoprotein component of chylomicrons and low-density lipoproteins (35). In this study, we found that elevated serum levels of Apo B were associated with an increased risk of ischemic stroke in AC-LVO. Additionally, a Mendelian randomization study reported a positive correlation of Apo B with large artery stroke and small vessel stroke (36). Therefore, we advocate Apo B as a marker of routine serum lipid examination.

In our study, age emerged as an important predictor in both models, as well as in a previously developed model for predicting the clinical outcome of AIS with LVO (17). In general, our results indicate that the prevalence of ischemic stroke and disability increases with age in patients with AC-LVO. In addition, our data also suggested that ASPECT was the common element included in all machine learning methods. Studies have demonstrated that diffusion-weighted imaging (DWI) ASPECTS which represents infarct volume,



is a significant independent predictor of functional outcome in AC-LVO strokes (37). Correspondingly, patients presenting with ASPECTS ≥ 7 are correlated with favorable outcomes following intravascular or thrombolytic therapy (38, 39). Our study further supports the association between ASPECT and the severity of neurological defects in first-ever ischemic stroke with AC-LVO. Consequently, a lower score of ASPECTS suggests less preserved brain parenchyma and predicts severe neurological impairment in patients with first-ever AC-LVO ischemic strokes.

It is well established that TIA increases the risk of ischemic stroke. In the present study, we found that prior TIA decreased ischemic stroke severity at admission, which is similar to the results of Marc Gotkine et al. showing that previous TIA was independently associated with lower severity of the ischemic stroke and a better short-term outcome (40). Prior TIA may have a neuroprotective effect on the subsequent ischemic stroke.

The chief strength of this study is the development and external validation of a new scoring tool, which predicts the risk of ischemic stroke and the severity of ischemia in AC-LVO based on machine learning approaches. Nevertheless, this study has several limitations. Foremost, few neuroimaging features were

taken into consideration, excluding others such as the collateral flow status which might improve the predictive performance of the models. Further evaluation of the level of collateral circulation is necessary. Second, the sample size for this study was small which might have been due to the stringent inclusion criteria for patients with AC-LVO. As a result, the performance advantages of machine learning models may not have been fully realized. Finally, this was a retrospective study; the performance of the model needs to be tested in a prospective population in future studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of the Second Hospital of Hebei Medical University on June 29, 2021 (approval No. 2021-R435). Written informed consent for participation was not

required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JC and JY drafted and revised the manuscript, participated in the study conception and design, performed the statistical analyses, and analyzed and interpreted the data. XLiu participated in the conception and design of the study, data interpretation, and made a major contribution to manuscript revision. JY assisted in designing the machine learning model and in data analysis. KZ, GX, RZ, XLi, LL, YZ, LZ, PY, LX, TL, JT, PZ, SY, QW, and LG participated in the design of the study and contributed to manuscript revision. All authors contributed to the article and approved the submitted version.

REFERENCES

- Rennert RC, Wali AR, Steinberg JA, Santiago-Dieppa DR, Olson SE, Pannell JS et al. Epidemiology, natural history, and clinical presentation of large vessel ischemic stroke. *Neurosurgery*. (2019) 85:S4–S8. doi: 10.1093/neuros/nyz042
- Smith WS, Lev MH, English JD, Camargo EC, Chou M, Johnston SC et al. Significance of large vessel intracranial occlusion causing acute ischemic stroke and TIA. *Stroke*. (2009) 40:3834–40. doi: 10.1161/STROKEAHA.109.561787
- Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. (2016) 387:1723–31. doi: 10.1016/S0140-6736(16)00163-X
- Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. (2015) 372:1019–30. doi: 10.1056/NEJMoa1414905
- Hastrup S, Damgaard D, Johnsen SP, Andersen G. Prehospital acute stroke severity scale to predict large artery occlusion: design and comparison with other scales. *Stroke*. (2016) 47:1772–6. doi: 10.1161/STROKEAHA.115.012482
- Katz BS, McMullan JT, Sucharew H, Adeoye O, Broderick JP. Design and validation of a prehospital scale to predict stroke severity: Cincinnati Prehospital Stroke Severity Scale. *Stroke*. (2015) 46:1508–12. doi: 10.1161/STROKEAHA.115.008804
- Teleb MS, Ver Hage A, Carter J, Jayaraman MV, McTaggart RA. Stroke vision, aphasia, neglect (VAN) assessment—a novel emergent large vessel occlusion screening tool: pilot study and comparison with current clinical severity indices. *J Neurointerv Surg*. (2017) 9:122–6. doi: 10.1136/neurintsurg-2015-012131
- Pérez de la Ossa N, Carrera D, Gorchs M, Querol M, Millán M, Gomis M et al. Design and validation of a prehospital stroke scale to predict large arterial occlusion: the rapid arterial occlusion evaluation scale. *Stroke*. (2014) 45:87–91. doi: 10.1161/STROKEAHA.113.003071
- Lima FO, Silva GS, Furie KL, Frankel MR, Lev MH, Camargo EC et al. Field assessment stroke triage for emergency destination: a simple and accurate prehospital scale to detect large vessel occlusion strokes. *Stroke*. (2016) 47:1997–2002. doi: 10.1161/STROKEAHA.116.013301
- Hendrix P, Sofoluke N, Adams MD, Kunaprayoon S, Z and R, Kolinovsky AN et al. Risk factors for acute ischemic stroke caused by anterior large vessel occlusion. *Stroke*. (2019) 50:1074–80. doi: 10.1161/STROKEAHA.118.023917
- Zhao W, Meng R, Ma C, Hou B, Jiao L, Zhu F et al. Safety and efficacy of remote ischemic preconditioning in patients with severe carotid artery stenosis before carotid artery stenting: a proof-of-concept, randomized controlled trial. *Circulation*. (2017) 135:1325–35. doi: 10.1161/CIRCULATIONAHA.116.024807
- Miyazawa N, Hashizume K, Uchida M, Nukui H. Long-term follow-up of asymptomatic patients with major artery occlusion: rate of symptomatic change and evaluation of cerebral hemodynamics. *AJNR Am J Neuroradiol*. (2001) 22:243–7.
- Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American heart association/American stroke association. *Stroke*. (2019) 50:e344–418. doi: 10.1161/STR.0000000000000211
- Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *J Med Syst*. (2002) 26:445–63. doi: 10.1023/A:1016409317640
- Loosli G, Canu S, Ong CS. Learning SVM in Krein Spaces. *IEEE Trans Pattern Anal Mach Intell*. (2016) 38:1204–16. doi: 10.1109/TPAMI.2015.2477830
- Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T et al. Deep learning-derived high-level neuroimaging features predict clinical outcomes for large vessel occlusion. *Stroke*. (2020) 51:1484–92. doi: 10.1161/STROKEAHA.119.028101
- Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T et al. Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning. *Stroke*. (2019) 50:2379–88. doi: 10.1161/STROKEAHA.119.025411
- Stamler JS, Osborne JA, Jaraki O, Rabbani LE, Mullins M, Singel D et al. Adverse vascular effects of homocysteine are modulated by endothelium-derived relaxing factor and related oxides of nitrogen. *J Clin Invest*. (1993) 91:308–18. doi: 10.1172/JCI116187
- Okamura T, Kitamura A, Moriyama Y, Imano H, Sato S, Terao A et al. Plasma level of homocysteine is correlated to extracranial carotid-artery atherosclerosis in non-hypertensive Japanese. *J Cardiovasc Risk*. (1999) 6:371–7. doi: 10.1177/204748739900600603
- Dardik R, Varon D, Tamarin I, Zivelin A, Salomon O, Shenkman B et al. Homocysteine and oxidized low density lipoprotein enhanced platelet adhesion to endothelial cells under flow conditions: distinct mechanisms of thrombogenic modulation. *Thromb Haemost*. (2000) 83:338–44. doi: 10.1055/s-0037-1613809
- Tsai JC, Perrella MA, Yoshizumi M, Hsieh CM, Haber E, Schlegel R et al. Promotion of vascular smooth muscle cell growth by homocysteine: a link to atherosclerosis. *Proc Natl Acad Sci U S A*. (1994) 91:6369–73. doi: 10.1073/pnas.91.14.6369
- Shi Z, Guan Y, Huo YR, Liu S, Zhang M, Lu H et al. Elevated total homocysteine levels in acute ischemic stroke are associated with long-term mortality. *Stroke*. (2015) 46:2419–25. doi: 10.1161/STROKEAHA.115.009136

FUNDING

This study was supported by a grant from XLiu from the National Natural Science Foundation of China (81571160).

ACKNOWLEDGMENTS

We wish to acknowledge the contributions of patients and their family members who provide clinical information.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2021.749599/full#supplementary-material>

23. Liu X, Zhang D, Liu Y, Sun X, Hou Y, Wang B. et al. A J-shaped relation of BMI and stroke: Systematic review and dose-response meta-analysis of 443 million participants. *Nutr Metab Cardiovasc Dis.* (2018) 28:1092–9. doi: 10.1016/j.numecd.2018.07.004
24. Li Y, Yatsuya H, Iso H, Yamagishi K, Saito I, Kokubo Y et al. Body mass index and risks of incident ischemic stroke subtypes: the japan public health center-based prospective (JPHC) study. *J Epidemiol.* (2019) 29:325–33. doi: 10.2188/jea.JE20170298
25. Yatsuya H, Yamagishi K, North KE, Brancati FL, Stevens J, Folsom AR, et al. Study investigators. Associations of obesity measures with subtypes of ischemic stroke in the ARIC Study. *J Epidemiol.* (2010) 20:347–54. doi: 10.2188/jea.JE20090186
26. Lovren F, Teoh H, Verma S. Obesity and atherosclerosis: mechanistic insights. *Can J Cardiol.* (2015) 31:177–83. doi: 10.1016/j.cjca.2014.11.031
27. Ogasawara K, Ogawa A, Yoshimoto T. Cerebrovascular reactivity to acetazolamide and outcome in patients with symptomatic internal carotid or middle cerebral artery occlusion: a xenon-133 single-photon emission computed tomography study. *Stroke.* (2002) 33:1857–62. doi: 10.1161/01.STR.0000019511.81583.A8
28. Zhang K, Li T, Tian J, Li P, Fu B, Yang X et al. Subtypes of anterior circulation large artery occlusions with acute brain ischemic stroke. *Sci Rep.* (2020) 10:3442. doi: 10.1038/s41598-020-60399-3
29. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB et al. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. *Circulation.* (2013) 127:e6–e245.
30. Freeman WD, Aguilar MI. Stroke prevention in atrial fibrillation and other major cardiac sources of embolism. *Neurol Clin.* (2008) 26:1129–60. doi: 10.1016/j.ncl.2008.07.001
31. Inoue M, Noda R, Yamaguchi S, Tamai Y, Miyahara M, Yanagisawa S et al. Specific factors to predict large-vessel occlusion in acute stroke patients. *J Stroke Cerebrovasc Dis.* (2018) 27:886–91. doi: 10.1016/j.jstrokecerebrovasdis.2017.10.021
32. Pagola J, Juega J, Francisco-Pascual J, Bustamante A, Penalba A, Pala E et al. Large vessel occlusion is independently associated with atrial fibrillation detection. *Eur J Neurol.* (2020) 27:1618–24. doi: 10.1111/ene.14281
33. Adams RJ, Chimowitz MI, Alpert JS, Awad IA, Cerqueria MD, Fayad P et al. American Heart Association/American Stroke Association. Coronary risk evaluation in patients with transient ischemic attack and ischemic stroke: a scientific statement for healthcare professionals from the Stroke Council and the Council on Clinical Cardiology of the American Heart Association/American Stroke Association. *Stroke.* (2003) 34:2310–22. doi: 10.1161/01.STR.0000090125.28466.E2
34. Conforto AB, Leite Cda C, Nomura CH, Bor-Seng-Shu E, Santos RD. Is there a consistent association between coronary heart disease and ischemic stroke caused by intracranial atherosclerosis? *Arq Neuropsiquiatr.* (2013) 71:320–6. doi: 10.1590/0004-282X20130028
35. Au A, Griffiths LR, Irene L, Kooi CW, Wei LK. The impact of APOA5, APOB, APOC3 and ABCA1 gene polymorphisms on ischemic stroke: Evidence from a meta-analysis. *Atherosclerosis.* (2017) 265:60–70. doi: 10.1016/j.atherosclerosis.2017.08.003
36. Yuan S, Tang B, Zheng J, Larsson SC. Circulating Lipoprotein Lipids, Apolipoproteins and Ischemic Stroke. *Ann Neurol.* (2020) 88:1229–36. doi: 10.1002/ana.25916
37. Rangaraju S, Streib C, Aghaebrahim A, Jadhav A, Frankel M, Jovin TG. Relationship Between Lesion Topology and Clinical Outcome in Anterior Circulation Large Vessel Occlusions. *Stroke.* (2015) 46:1787–92. doi: 10.1161/STROKEAHA.115.009908
38. Demchuk AM, Hill MD, Barber PA, Silver B, Patel SC, Levine SR. NINDS rtPA Stroke Study Group, NIH. Importance of early ischemic computed tomography changes using ASPECTS in NINDS rtPA Stroke Study. *Stroke.* (2005) 36:2110–5. doi: 10.1161/01.STR.0000181116.15426.58
39. Hill MD, Rowley HA, Adler F, Eliasziw M, Furlan A, Higashida RT et al. Selection of acute ischemic stroke patients for intra-arterial thrombolysis with pro-urokinase by using ASPECTS. *Stroke.* (2003) 34:1925–31. doi: 10.1161/01.STR.0000082483.37127.D0
40. Wang WW, Chen DZ, Zhao M, Yang XF, Gong DR. Prior transient ischemic attacks may have a neuroprotective effect in patients with ischemic stroke. *Arch Med Sci.* (2017) 13:1057–61. doi: 10.5114/aoms.2016.63744

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cui, Yang, Zhang, Xu, Zhao, Li, Liu, Zhu, Zhou, Yu, Xu, Li, Tian, Zhao, Yuan, Wang, Guo and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.