Personalized risk prediction of symptomatic intracerebral hemorrhage after stroke thrombolysis using a machine-learning model

Feng Wang (iD), **Yuanhanqing Huang, Yong Xia, Wei Zhang, Kun Fang, Xiaoyu Zhou, Xiaofei Yu, Xin Cheng, Gang Li** (iD)**, Xiaoping Wang, Guojun Luo, Danhong Wu, Xueyuan Liu, Bruce C.V. Campbell, Qiang Dong** and **Yuwu Zhao**

Abstract

Background: Personalized prediction of the risk of symptomatic intracerebral hemorrhage (sICH) after stroke thrombolysis is clinically useful. Machine-learning-based modeling may provide the personalized prediction of the risk of sICH after stroke thrombolysis.

Methods: We identified 2578 thrombolysis-treated ischemic stroke patients between January 2013 and December 2016 from a multicenter database, where 70% were used to train models and the remaining 30% were used as the nominal test sets. Another 136 consecutive tissue plasminogen-activated-treated patients between January 2017 and December 2017 from our institute were enrolled as the independent test sets for clinical usability evaluation. Five machine-learning models were developed to predict the risk of sICH after stroke thrombolysis, and the receiving operating characteristic (ROC) was used to compare the prediction performance.

Results: In total, 2237 cases were included in our study, of which 102 had sICH transformation (4.56%). Finally, the three-layer neuro network was selected with the best performance on nominal test sets (AUC = 0.82). The probability of the model score was further categorized into three risk ranks (18.97%, 5.63%, and 0.81%) according to the risk distribution. Implementing our system in clinical practice was associated with reduced computed tomography (CT)-to-treatment time (CTT; 41 min *versus* 52 min, $p < 0.001$). All sICH patients were correctly predicted to be within the high-sICH risk rank.

Conclusions: The machine-learning-based modeling is feasible for providing personalized risk prediction of sICH after stroke thrombolysis, and is able to reduce the CTT. More data are needed to further optimize the model and improve the accuracy of prediction.

Keywords: ischemic stroke, machine learning, prediction, tPA, symptomatic intracerebral hemorrhage, thrombolysis

Received: 11 August 2019; revised manuscript accepted: 4 January 2020.

Correspondence to:
Yuwu Zhao
Department of Neurology,
Shanghai Jiao Tong
University Affiliated Sixth
People's Hospital, No.
600, Yishan Road, Xuhui
District, Shanghai, 200233,
China
zhaoyuwu2005@126.com

Qiang Dong
Department of Neurology,
Huashan Hospital,
Fudan University, No.12,
Wulumuqi Zhong Road,
Jingan District, Shanghai,
200040, China
qiang_dong163@163.com

Feng Wang
Department of Neurology,
Shanghai Jiao Tong
University Affiliated
Sixth People's Hospital,
Shanghai, China

Yuanhanqing Huang
Wei Zhang
College of Electronics and
Information Engineering,
Tongji University,
Shanghai, China

Yong Xia
IBM, Shanghai, China

Kun Fang
Xin Cheng
Department of Neurology,
Huashan Hospital, Fudan
University, Shanghai,
China

Xiaoyu Zhou
Xueyuan Liu
Department of Neurology,
Shanghai Tenth People's
Hospital, Tongji University,
Shanghai, China

Xiaofei Yu
Department of Neurology,
Shuguang Hospital
Affiliated to Shanghai
University of Traditional
Chinese Medicine,
Shanghai, China

Gang Li
Department of Neurology,
East Hospital, Tongji
University, Shanghai,
China

Introduction

The efficacy of intravenous thrombolysis in acute ischemic stroke (AIS) has been established for over 2 decades,[1] yet <5% AIS patients worldwide receive this therapy.[2] Concern regarding symptomatic intracerebral hemorrhage (sICH) after thrombolysis is one of the factors limiting implementation.[3] Multiple risk factors for sICH have been identified.[4,5] Moreover, a variety of sICH risk scales[6–8] has been developed in the past decade. However, these scales have not identified a subgroup with sufficient risk of sICH to justify withholding thrombolysis.

Machine-learning technologies are well suited to balancing the contributions of multiple variables, and have been applied in various medical fields with great success.[9–11] However, there has been

Xiaoping Wang
Department of Neurology, Shanghai TongRen Hospital, Tongji University, Shanghai, China

Guojun Luo
Department of Neurology, Jinshan Branch of Shanghai Sixth People's Hospital, Shanghai, China

Danhong Wu
Department of Neurology, Shanghai Fifth People's Hospital, Fudan University, Shanghai, China

Bruce C.V. Campbell
Departments of Medicine and Neurology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Parkville, Australia

little application in stroke. Two preliminary analyses[12,13] were based on single-center data with limited numbers and did not apply methods to deal with missing data or to prioritize the relevant data features for optimized prediction.

In contrast to previous studies, we aimed to explore a better way of applying machine learning to treatment of cerebrovascular disease through developing and validating a simple machine-learning system that could assist clinicians to predict the risk of thrombolysis-related sICH in ischemic stroke patients.

Methods

Study population

We identified 2578 consecutive intravenous (IV) tissue plasminogen-activated (tPA)-treated AIS patients between January 2013 and December 2016 from the multicenter Shanghai Stroke Service System (4S) database (indicated as multicenter training and testing sets). Another 136 consecutive tPA-treated patients between January 2017 and December 2017 from Shanghai Sixth People's Hospital were also included as independent testing sets. The 4S database[14] underpins a quality-improvement project for stroke care throughout Shanghai (population more than 20 million). Clinical information is automatically extracted from electronic medical records and uploaded to the database with checks for data validity, and cases of eight different general hospitals were included. The general characteristics of patients are collected, including age, sex, weight, past history, smoking status, alcohol consumption, medication history, admission blood pressure, admission blood glucose, baseline National Institute of Health Stroke Scale (NIHSS) score, onset-to-treatment time (OTT) and door-to-needle time (DNT). The Clinical Research Institute of Huashan Hospital served as the data analysis center and had local institutional review board approval to conduct the study.

Machine-learning process

Definition of sICH: considering the actual treatment process of tPA-treated patients, thrombolysis-related sICH was defined as neurological worsening $\geqslant 4$ NIHSS points within 24h of treatment that was attributed to hemorrhagic transformation of the infarct evident on CT brain; slightly different to the ECASS II study.[5] CT scan images were confirmed by investigators at each center.

Data preprocessing: imputation of missing feature values, normalization, and imbalanced processing (detailed in supplementary material) had been sequentially applied in all data sets. The missing features were imputed using the missing-indicator method.[15] For the categorized features, the one hot encoding was used to cover all the possibilities, and for the continuous type of features, Z score normalization was applied. Oversampling and cost-sensitive adaptation were used because of the imbalanced distribution of sICH to no-sICH cases (102: 2135).[16]

Feature selection: we used the wrapper method,[16] correlation-based feature selection,[17] and conservative-mean (CM) method[12] to analyze the correlation of features with sICH. All cases were randomly partitioned into 70% training set and 30% testing set with the same percentage of sICH in each set. Feature selection was only conducted in the training set with a 10-fold validation method.

Modeling: logistic regression, neural network, support-vector machine (SVM),[18] random forest and adaptive boosting (AdaBoost) were developed on multicenter testing sets and compared to test their performance on multicenter testing sets and independent testing sets with area under the receiver-operating-characteristic curve (AUC).

Interpretation of output: the native output of the model is a float value. To allow clinical application, we converted these outputs to three ranks of sICH possibilities. Unsupervised equal frequency method and supervised method were used to rank the outputs. Wilcoxon rank-sum test was used to evaluate the performance of output interpretation.

Details of the machine-learning processing pipeline are provided in the online supplement.

Traditional statistical analyses

Analyses were performed using IBM SPSS V17.0 statistical package (Shanghai, China). All
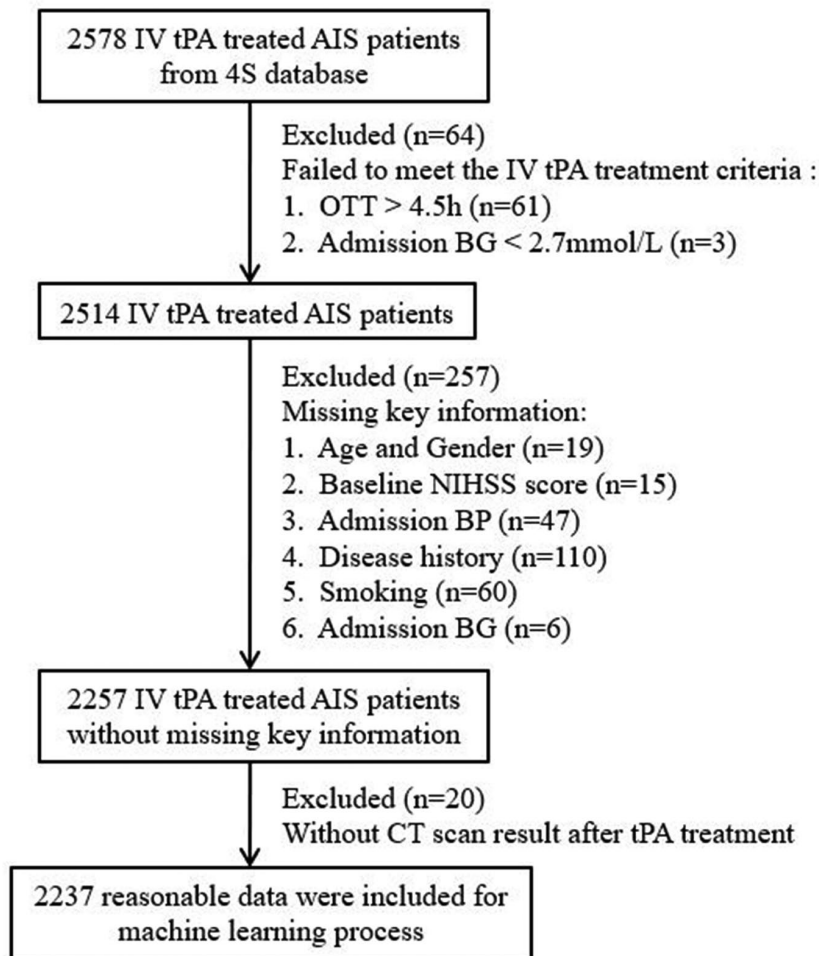
Figure 1. Flowchart of the data included in machine-learning process.
AIS, acute ischemic stroke; BG, blood glucose; BP, blood pressure; CT, computed tomography; IV, intravenous; NIHSS, National Institutes of Health Stroke Scale; OTT, onset-to-treatment time; tPA, tissue plasminogen activator.

continuous variables were first tested for normality of distribution. Variables with normal distribution were expressed as mean ± standard; others were expressed as median ± interquartile range. Differences between groups were analyzed using the *t* test or Mann–Whitney U test as appropriate to distribution. Categorical variables were expressed as number (percentage) and Fisher's exact test was used for comparison between groups. All *p* values were two tailed, and $p < 0.05$ was considered statistically significant.

Results

A total of 2237 of 2578 thrombolysis patients in a multicenter database were included for model development and test, where 61 cases had OTT longer than 4.5 h, 3 cases had blood glucose less than 2.7 mmol/l, 257 cases had more than 5% features missing, and 20 cases had no 24 h CT scan after thrombolysis (Figure 1). For the independent test sets, all the 136 consecutive tPA-treated patients between January 2017 and December 2017 from our institute were included.

Among the included 2237 patients in the multi-center data sets, there were 102 sICH patients (4.6%). The baseline characteristics are shown in Table 1. The sICH rate ranged from 1.01% to 6.95% among different hospitals in our database. Increased age, atrial fibrillation, elevated blood glucose, higher baseline NIHSS score and prolonged DNT were significantly associated with sICH ($p < 0.05$).

Table 1. Baseline characteristics of the study population.

Characteristics	No-sICH group (*n* = 2135)	sICH group (*n* = 102)	*p* value (two sided)
Male, *n* (%)	1378 (64.54)	60 (58.82)	0.246
Age, years (mean ± SD)	66.32 ± 12.67	69.54 ± 11.89	0.012
Medical history, *n* (%)			
Hypertension	1361 (63.75)	71 (69.61)	0.247
Diabetes mellitus	506 (23.70)	32 (31.37)	0.096
Atrial fibrillation	436 (20.42)	39 (38.24)	<0.001
Previous stroke	360 (16.86)	15 (14.71)	0.592
Myocardial infarction	119 (9.50)	5 (12.20)	0.585
Dyslipidemia	122 (12.02)	1 (2.70)	0.113
Prior medication, *n* (%)			
Oral anticoagulants	13 (1.06)	0 (0.00)	1.000
Any antithrombotic	220 (15.40)	9 (16.67)	0.847
Admission information			
Smoking, *n* (%)	735 (34.43)	35 (34.31)	1.000
Alcohol consumption, *n* (%)	270 (17.75)	12 (20.00)	0.609
Blood glucose, mmol/l (mean ± SD)	7.81 ± 3.30	8.44 ± 3.45	0.005
Diastolic pressure (mmHg, mean ± SD)	83.16 ± 13.49	83.73 ± 14.85	0.859
Systolic pressure (mmHg, mean ± SD)	149.90 ± 23.12	151.90 ± 26.32	0.443
Baseline NIHSS, median (IQR)	7 (4–12)	15 (10–19)	<0.001
OTT (min, mean ± SD)	161.70 ± 52.02	166.07 ± 50.20	0.383
DNT (min, mean ± SD)	70.51 ± 35.29	79.81 ± 37.21	0.044

p value refers to the difference of participants' characteristics between sICH group and no-sICH group.
DNT, door-to-needle time; IQR, interquartile range; NIHSS, National Institutes of Health Stroke Scale; OTT, onset-to-treatment time; SD, standard deviation; sICH, symptomatic intracerebral hemorrhage.

Model development and evaluation using multicenter sets

To discover the well-performed model with proper feature inputs, we used different feature-selection methods and model architectures to select the best-performed model.

Without feature selection, the AUCs of five machine-learning technologies (logistic regression, neural network, SVM, random forest and AdaBoost) were 0.69, 0.73, 0.58, 0.76, and 0.75 in the multicenter test sets, respectively.

Among the used feature-selection methods, in the logistic regression model [Figure 2(a)] and SVM model [Figure 2(b)], AUC value rose to 0.76 with a correlation-based feature selection (CFS) method and 0.79 with a CM method. In the neural network model [Figure 2(c)], the AUC value increased 20% with the CM method. In our
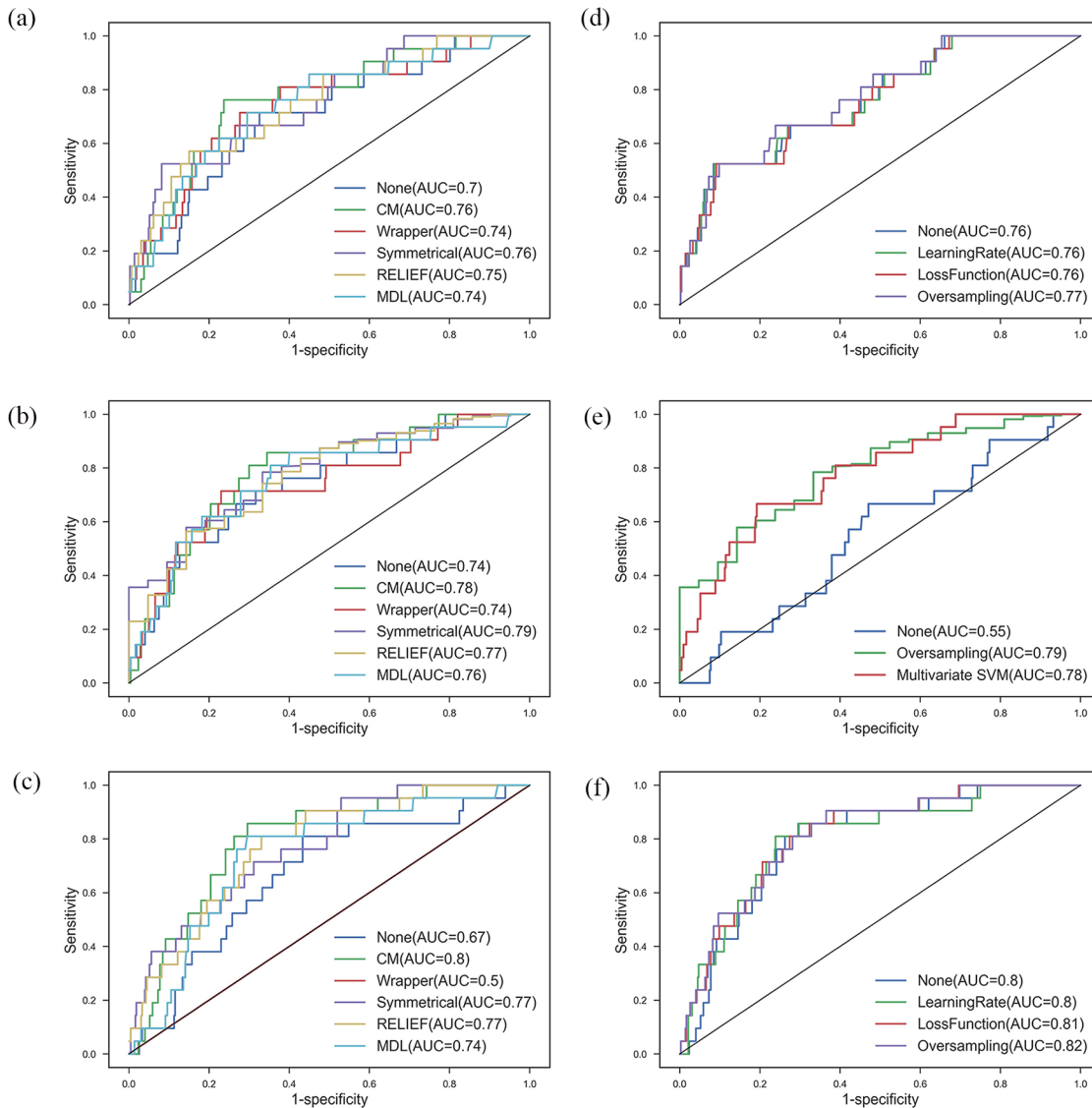
Figure 2. Results of feature selection and imbalanced data-processing methods.
(a) ROC curves for logistic regression; (b) ROC curves for SVM with linear kernel. Here, oversampling was conducted; (c) ROC curves for a three-layer neural network. None: without feature selection, (d) ROC curve for different imbalanced data-processing methods in logistic regression. Here, the CFS method with symmetrical uncertainty was used (it was also used in the following SVM); (e) ROC curve for the different imbalanced data processing method in SVM; (f) ROC curve for the different imbalanced data processing methods in perceptron with the CM method. None: without feature selection for (a), (b), (c), and without imbalanced data processing for (d), (e), (f).
AUC, area under the (ROC) curve; CFS, correlation-based feature selection; CM, conservative mean; learning rate, cost-sensitive learning rate; MDL, minimum description length; RELIEF, relevant features; ROC, receiver operating characteristics; SVM, support-vector machine; symmetrical, symmetrical uncertainty.

model, age, atrial fibrillation, glucose level, NIHSS score, and DNT were selected as the most important input factors in the feature-selection step, consistent with the results of traditional statistical analyses.

The ratio of patients with and without sICH was approximately 1: 21. We compared the

effectiveness of three methods that aim to address this imbalanced distribution. The AUC of the logistic regression model [Figure 2(d)] did not change significantly using the three imbalanced data-processing methods. However, the AUC of the SVM model [Figure 2(e)] rose to 0.79 with an oversampling method, and 0.78 with multivariate SVM. The neural network model [Figure 2(f)]
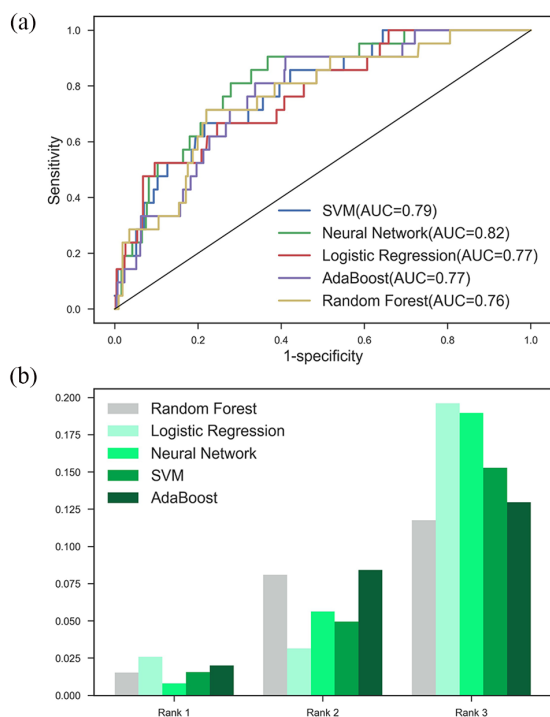
Figure 3. Evaluation of different classifiers in predicting sICH and ranking results.
(a) ROC curve for different classifiers, with appropriate modification and improvement made to cope with the inferior quality of some features, strong relevance between different features and imbalanced data; and (b) the bar chart records the ratio of sICH in each rank of different machine-learning algorithms with different colors. Rank 1–3 suggests an ascending trend in the risk of sICH.
AdaBoost, adaptive boosting; AUC, area under the (ROC) curve; ROC, receiver operating characteristics; sICH, symptomatic intracerebral hemorrhage; SVM, support-vector machine.

AUC did not change significantly with the three imbalanced data-processing methods.

All of the classifiers achieved AUCs greater than 0.70. Random forest and AdaBoost are ensemble learning approaches suited to large datasets and had AUCs of 0.76 and 0.77, respectively, in our study [Figure 3(a)]. A three-layer neural network using feature selection and oversampling had the highest AUC of 0.82, followed by SVM and logistic regression using corresponding data preprocessing methods with an AUC of 0.79 and 0.77, respectively [Figure 3(a)]. Four-layer and five-layer neural networks were also constructed and produced results similar to the three-layer neural network.

We set three ranks based on the output from the neural network model, representing an sICH risk of 18.97%, 5.63% and 0.81% [Figure 3(b)]. In Wilcoxon rank-sum test, using a three-layer neural network and supervised discretization, we demonstrated a significant correlation between the actual sICH status after thrombolysis and the model-derived ranking of sICH rate ($Z = 4.670$, $p < 0.001$; Table 2).

Clinical usability evaluation in independent test sets

This sICH after stroke thrombolysis-risk prediction model was tested in independent test sets to evaluate its clinical usability. Patient demographics were shown in Table 3. In 2017, the proportion of

Table 2. Results of Wilcoxon rank-sum test of different machine-learning models and partition methods.

Model	Random forest	Logistic regression	Neural network	SVM	AdaBoost
Unsupervised					
Z value	3.293	3.385	4.123	3.752	3.391
p value (two tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
Supervised					
Z value	3.404	3.108	4.670	3.852	3.301
p value (two tailed)	<0.001	<0.001	<0.001	<0.001	<0.001

AdaBoost, adaptive boosting; unsupervised, unsupervised discretization methods; supervised, supervised discretization methods; SVM, support-vector machine; *Z* value, value of *Z* statistic used to compute the approximate *p* value of the test; *p* value, the difference of participants' characteristics between the sICH group and no-sICH group (Wilcoxon rank-sum test).

Table 3. Patient characteristics from a single center.

Characteristics	Patients in 2016 (*n* = 120)	Patients in 2017 (*n* = 136)	*p* value (two sided)
Male, *n* (%)	82 (68.3)	94 (69.1)	0.894
Age (year, mean ± SD)	63.47 ± 10.82	66.07 ± 12.32	0.247
Medical history, *n* (%)			
Hypertension	77 (64.2)	85 (61.0)	0.689
Diabetes mellitus	26 (21.7)	29 (21.3)	1.000
Atrial fibrillation	20 (16.7)	24 (17.6)	0.869
Previous stroke/TIA	15 (12.5)	22 (16.2)	0.477
Ischemic heart disease	14 (11.7)	5 (3.7)	0.585
Prior medication, *n* (%)			
Oral anticoagulants	2 (1.7)	1 (0.7)	0.601
Any antithrombotic	12 (10.0)	30 (22.1)	0.011
Statins	7 (5.8)	21 (15.4)	0.016
Admission information			
Smoking, *n* (%)	22 (18.3)	33 (24.3)	0.287
Alcohol consumption, *n* (%)	10 (8.3)	16 (11.8)	0.412
Blood glucose (mmol/l, mean ± SD)	7.78 ± 3.50[1]	7.75 ± 3.67	0.614
Diastolic pressure (mmHg, mean ± SD)	82.46 ± 12.18	80.47 ± 15.31	0.116
Systolic pressure (mmHg, mean ± SD)	150.34 ± 19.28	147.68 ± 20.55	0.389
Baseline NIHSS, median (IQR)	7 (4–12)	6 (2–14)	0.695
Dose of Alteplase, mg/kg, median (IQR)	0.88 (0.82–0.9)	0.89 (0.84–0.9)	0.194
OTT, min, median (IQR)	173 (132–230)	155 (124–191)	0.011
DNT, min, median (IQR)	70 (59–92)[3]	64 (56–83)[3]	0.062
OTD, min, median (IQR)	90 (60–147)[3]	89 (55–122)[4]	0.207
DTC, min, median (IQR)	25 (18–36)[6]	28 (22–36)[11]	0.025
OTC, min, median (IQR)	115 (83–165)[3]	112 (87–153)[3]	0.563
CTT, min, median (IQR)	42 (34–56)	34 (26–45)	<0.001

The superscript numbers in the table represent the number of missing data.
p value refers to the difference of participants' characteristics between the two groups.
CTT, CT-to-treatment time; DNT, door-to-needle time; DTC, door-to-CT time; IQR, interquartile range; NIHSS, National Institutes of Health Stroke Scale; OTC, onset-to-CT time; OTD, onset-to-door time; OTT, onset-to-treatment time; SD, standard deviation; sICH, symptomatic intracerebral hemorrhage; TIA, transischemic attack.

oral antiplatelet and statin use was significantly higher than in 2016 ($p < 0.05$). There was a trend toward reduced OTT (median 155 min *versus* 173 min, $p < 0.05$), largely driven by a significantly reduced CT-to-treatment time (median 34 min *versus* 42 min, $p < 0.001$).

There were eight sICH patients in 2017 (5.88%). This model was used to predict sICH after CT scan in the emergency room in 129 patients, without interfering with the normal treatment process and decision making. There were 22 cases in rank 3 (possible sICH rate 18.97%), 42 cases were classified as rank 2 (possible sICH rate 5.63%), and the remaining 65 cases were in rank 1 (possible sICH rate 0.81%). Of the 22 patients in the highest risk category, 4 developed sICH (actual sICH rate 18.18%). Another 4 sICH cases were in rank 2 (actual sICH rate 9.52%). About 50% of patients were classified as very low risk of sICH and none developed sICH.

Discussion

In this study, we identified clinical and laboratory characteristics that were readily available before thrombolysis and validated a semiautomated post-thrombolysis sICH prediction system by leveraging machine-learning technologies. All the data used to derive the machine-learning process came from a real-world multicenter patient database. As far as we knew, this is the first study using multicenter data to develop and evaluate the models and then test its clinical usability in independent sets.

Since sICH after thrombolysis is a complex phenomenon, the machine-learning model processes weighs multiple parameters, which are routinely assessed before treatment in potential thrombolysis candidates, and so are immediately available for input into the model.

Clearly, certain parameters are more strongly associated with sICH than others. Traditional statistical analyses showed that patients having older age, atrial fibrillation, higher glucose level, higher NIHSS score, and longer DNT were more likely to have sICH in our database. By machine-learning process, age, atrial fibrillation, glucose level, NIHSS score, and DNT were selected as the most important input features in the feature-selection step, consistent with the results of traditional statistical analyses. However, there are some differences in independent risk factors related to sICH from different trials[4–8,19] previously published. So, we have not discarded other features that currently seem to be less related to sICH in our model, hoping the selection of sICH-related factors will be clarified with more data included, and to achieve personalized sICH risk prediction in tPA-treated stroke patients.

Specific methods were applied to solve missing data and imbalanced data problems. The high rate (>30%) of missing data may generate bias and affect the prediction performance of the sICH risk prediction model. We have applied various imputation techniques as indicated in supplementary material and marked the missing data, but it is still not possible to guess the missing data based on the other information in the datasets. This might be the reason we did not achieve higher prediction accuracy. A large population with fewer missing features is expected to be established to further improve the accuracy of sICH risk prediction using machine-learning algorithms. Without data preprocessing, the AUCs of machine-learning models are not ideal enough. The three-layer neural network model combined with feature selection and imbalanced data processing was chosen for clinical implementation for the time being. As the data increase, other machine-learning technologies might perform better in the future. The model can be adjusted from time to time with increasing data. The most important point of this study is the idea of applying machine-learning model to predict personalized risk of sICH after stroke thrombolysis.

In general, the clinically useful sICH risk prediction system must be simple to use, given the time-critical nature of thrombolysis, where the clinician can put all the information (basic information about the patient, clinical information, IV tPA eligibility checklist) into the system any time before thrombolysis. The average time to input the data and obtain risk-score prediction was within 3–4 min.

Although IV tPA is recommended in Chinese AIS guidelines, informed consent from the patient or family usually costs time. Our study found that the CTT was significantly shortened after implementation of the prediction system. This is an indication that machine-learning decision support may assist clinicians to make faster thrombolysis decisions in future. Approximately 50% of

patients were classified in the low-risk group and none of these developed sICH in the prospective cohort which may offer some reassurance to the clinician considering thrombolysis. However, it is important to emphasize that the use of our system should not affect clinical decision making until further validation with a larger number of sICH cases is undertaken to improve the efficacy of the system.

One of the limitations of our study is the smaller population of sICH-after-thrombolysis patients due to the low incidence of sICH, which might mean that when applying this model to a new territorial population, it should be recalibrated or even retrained. But, at least it means that the machine-learning model is feasible for making personalized risk prediction. Its generalization might be further improved using federal training policy, by exploring a more territorial population. Another limitation of our study is lacking the imaging features of emergency-room CT scan.[20,21] One of the biggest challenges is that of time-critical processing and interpretation processing. Recent progress on deep-learning techniques may provide the opportunity to further integrate the clinical aspects with imaging features to make automatic, time-critical processing to predict risk of sICH after stroke thrombolysis. However, it takes many more cases for training and evaluation.

In conclusion, we have demonstrated the feasibility and proof of principle that the machine-learning model can generate a clinical useful prediction of sICH risk using readily available clinical data in a very short time. The system has the potential for continuous improvement, with addition of further sICH data and new parameters, and provides an illustration of how machine learning may benefit clinical practice in the future.

ORCID iDs
Feng Wang https://orcid.org/0000-0003-3882-4008

Gang Li https://orcid.org/0000-0003-1737-0107

Supplemental material
Supplemental material for this article is available online.

References

1. Campbell BCV, Meretoja A, Donnan GA, *et al.* Twenty-year history of the evolution of stroke thrombolysis with intravenous alteplase to reduce long-term disability. *Stroke* 2015; 46: 2341–2346.

2. Reeves MJ, Arora S, Broderick JP, *et al.* Acute stroke care in the US: results from the 4 pilot prototypes of the Paul Coverdell national acute stroke registry. *Stroke* 2005; 36: 1232–1240.

3. Caplan LR. Stroke thrombolysis: slow progress. *Circulation* 2006; 114: 187–190.

4. Tanne D, Kasner SE, Demchuk AM, *et al.* Markers of increased risk of intracerebral hemorrhage after intravenous recombinant tissue plasminogen activator therapy for acute ischemic stroke in clinical practice: the multicenter rt-PA stroke survey. *Circulation* 2002; 105: 1679–1685.

5. Larrue V, Von Kummer RR, Müller A, *et al.* Risk factors for severe hemorrhagic transformation in ischemic stroke patients treated with recombinant tissue plasminogen activator: a secondary analysis of the European-Australasian acute stroke study (ECASS II). *Stroke* 2001; 32: 438–441.

6. Lou M, Safdar A, Mehdiratta M, *et al.* The HAT score: a simple grading scale for predicting hemorrhage after thrombolysis. *Neurology* 2008; 71: 1417–1423.

7. Mazya M, Egido JA, Ford GA, *et al.* Predicting the risk of symptomatic intracerebral hemorrhage in ischemic stroke treated with

intravenous alteplase: safe implementation of treatments in stroke (SITS) symptomatic intracerebral hemorrhage risk score. *Stroke* 2012; 43: 1524–1531.

8. Strbian D, Michel P, Seiffge DJ, *et al.* Symptomatic intracranial hemorrhage after stroke thrombolysis: the SEDAN score. *Ann Neurol* 2012; 71: 634–641.

9. Yu KH, Zhang C, Berry GJ, *et al.* Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016; 7: 12474.

10. Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115.

11. Ghazvinian Zanjani F, Zinger S, Piepers B, *et al.* Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images. *J Med Imaging (Bellingham)* 2019; 6: 1.

12. Khosla A, Cao Y, Lin CCY, *et al.* An integrated machine learning approach to stroke prediction. In: *16th ACM SIGKDD international conference on knowledge discovery and data mining*, Washington, DC, 25–28 July 2010.

13. Dharmasaroja P and Dharmasaroja PA. Prediction of intracerebral hemorrhage following thrombolytic therapy for acute ischemic stroke using multiple artificial neural networks. *Neurol Res* 2012; 34: 120–128.

14. Dong Y, Fang K, Wang X, *et al.* The network of Shanghai stroke service system (4S): a public health-care web-based database using automatic extraction of electronic medical records. *Int J Stroke*. Epub ahead of print 21 March 2018. DOI: 10.1177/174749301876549.

15. Jones MP. Indicator and stratification methods for missing explanatory variables in multiple linear regression. *J Am Stat Assoc* 1996; 91: 222–230.

16. He H and Garcia EA. Learning from imbalanced data. In: *IEEE Transactions on knowledge and data engineering*, IEEE, New York, 2009, pp.1263–1284.

17. Hall MA. Correlation-based feature selection for discrete and numeric class machine learning. In: *Seventeenth international conference on machine learning*, Stanford University, Stanford, CA, 29 June – 2 July 2000, pp.349–366.

18. Tsochantaridis I, Hofmann T, Joachims T, *et al.* Support vector machine learning for interdependent and structured output spaces. In: *Twenty-first international conference on Machine learning*, 4 July 2004, Banff, Alberta, Canada.

19. Gilligan AK, Markus R, Read S, *et al.* Baseline blood pressure but not early computed tomography changes predicts major hemorrhage after streptokinase in acute ischemic stroke. *Stroke* 2002; 33: 2236–2242.

20. Lyden P. Early major ischemic changes on computed tomography should not preclude use of tissue plasminogen activator. *Stroke* 2003; 34: 821–822.

21. Bentley P, Ganesalingam J, Carlton Jones AL, *et al.* Prediction of stroke thrombolysis outcome using CT brain machine learning. *Neuroimage Clin* 2014; 4: 635–640.