EBioMedicine 35 (2018) 251–259

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Research paper

Development and validation of a penumbra-based predictive model for thrombolysis outcome in acute ischemic stroke patients

Tian-Yu Tang [a,1], Yun Jiao [a,1], Ying Cui [a], Chu-Hui Zeng [a], Deng-Ling Zhao [a], Yi Zhang [a], Cheng-Yu Peng [a], Xin-Dao Yin [b], Pei-Yi Gao [c], Yun-Jun Yang [d], Sheng-Hong Ju [a,*], Gao-Jun Teng [a,*]

[a] Jiangsu Key Laboratory of Molecular and Functional Imaging, Department of Radiology, Zhongda Hospital, Medical School of Southeast University, 87 Dingjiaqiao Road, Nanjing, Jiangsu 210009, China
[b] Department of Radiology, Nanjing First Hospital, Nanjing Medical University, 68 Changle Road, Nanjing, Jiangsu 210006, China
[c] Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, 6 Tiantanxili, Beijing 100050, China
[d] Department of Radiology, First Affiliated Hospital of Wenzhou Medical University, 2 Fuxuexiang, Wenzhou, Zhejiang 325000, China

ARTICLE INFO

Article history:
Received 17 June 2018
Received in revised form 18 July 2018
Accepted 19 July 2018
Available online 23 August 2018

Keywords:
Penumbra
Predictive model
Clinical outcome
Acute ischemic stroke
Thrombolysis

ABSTRACT

The use of thrombolysis in acute ischemic stroke is restricted to a small proportion of patients because of the rigid 4·5-h window. With advanced imaging-based patient selection strategy, rescuing penumbra is critical to improving clinical outcomes. In this study, we included 155 acute ischemic stroke patients (84 patients in training dataset, age from 43 to 80, 59 males; 71 patients in validation dataset, age from 36 to 80, 45 males) who underwent MR scan within the first 9-h after onset, from 7 independent centers. Based on the mismatch concept, penumbra and core area were identified and quantitatively analyzed. Moreover, predictive models were developed and validated to provide an approach for identifying patients who may benefit from thrombolytic therapy. Predictive models were constructed, and corresponding areas under the curve (AUC) were calculated to explore their performances in predicting clinical outcomes. Additionally, the models were validated using an independent dataset both on Day-7 and Day-90. Significant correlations were detected between the mismatch ratio and clinical assessments in both the training and validation datasets. Treatment option, baseline systolic blood pressure, National Institutes of Health Stroke Scale score, mismatch ratio, and three regional radiological parameters were selected as biomarkers in the combined model to predict clinical outcomes of acute ischemic stroke patients. With the external validation, this predictive model reached AUCs of 0·863 as short-term validation and 0·778 as long-term validation. This model has the potential to provide quantitative biomarkers that aid patient selection for thrombolysis either within or beyond the current time window.

1. Introduction

Stroke is the second most common cause of death and the leading cause of adult disability worldwide [1] and patients with acute ischemic stroke (AIS) may benefit from thrombolysis up to 4·5 h after onset [2], while a treatment beyond this time window is associated with an increased risk of mortality [3,4]. The time window of 4·5 h is derived from large-sample randomized clinical trials [5,6]. Using imaging guidance to confirm the eligibility of AIS patients for thrombolytic therapy, despite the 4·5-hour time window, has become clinically pertinent. It is of great value to popularize imaging-guided decision-making for the improved management of AIS progress, especially in primary healthcare centers.

Penumbra quantification is critical for improving clinical outcomes and is a promising strategy for extending the treatment time window [5,7]. As a dynamic process, the penumbra can persist for as long as 48 h after symptom onset [4]. Both CT and MR are valuable imaging modalities for assessing the penumbra [8]. Although CT is faster and more widely used, MR is potentially superior to CT because of its higher resolution, rapid identification of acute infarction, and sensitivity to intracranial hemorrhage [9,10]. Several clinical trials, including DEFUSE 2 and EPITHET, have used the MR mismatch criteria to guide patient selection for thrombolysis [11,12]. However, the clinical conditions and radiological signatures of patients who are most likely to benefit from thrombolysis have not been clarified [13]. In addition, most current MR studies rely on qualitative imaging which might lead to weak interpretation of results [14]. Furthermore, based on suggestions from previous studies that the location, volume ratio, and perfusion/diffusion features of the

* Corresponding authors at: Jiangsu Key Laboratory of Molecular and Functional Imaging, Department of Radiology, Zhongda Hospital, Medical School, Southeast University, 87 Dingjiaqiao Road, Nanjing 210009, China.
E-mail addresses: jsh0836@hotmail.com (S.-H. Ju), gjteng@vip.sina.com (G.-J. Teng).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2018.07.028

Research in context

Evidence before this study

Thrombolysis is the sole effective treatment with a Level Ib evidence for the early treatment of acute ischemic stroke (AIS) that can improve clinical outcomes. Unfortunately, the narrow 4·5-h window with thrombolysis restricts the eligibility of patients and the clinical significance of the treatment. The concept of the penumbra describes dynamic hypo-perfused brain tissue with the capacity to recover if perfusion is promoted. Penumbral quantification would contribute to extending the current time window and including more treatable AIS patients. However, because the methodology for quantifying the penumbra is inadequate, an imaging-based selection for thrombolysis beyond 4·5 h has not been established.

Added value of this study

By including global and regional features of the impaired brain region, we developed a multivariable quantification method for penumbral assessment that reflects more information than assessments from previous studies. The predictive model, which combined both clinical and penumbral signatures, demonstrated a sufficiently high discrimination for clinical outcomes. Through combing clinical factors and quantification of impaired brain regions after AIS, the model enabled the identification of patients who may benefit from thrombolysis.

Implications of all the available evidence

Our model provides an approach via the penumbral quantification for identifying patients who may benefit from thrombolytic therapy either within or beyond the 4·5-h window. The current results suggest the necessity of combining clinical and radiological characteristics, as opposed to using the time lag from stroke onset alone, for the selection. The application of this model to a validated cohort in this study showed promising results toward selecting patients who may qualify for thrombolysis and benefit from treatment.

penumbra are all critical predictors of clinical outcome [15–17]. Therefore, an overall assessment of penumbral profiles might be helpful in determining whether thrombolytic treatment is effective either within or beyond the 4·5-h window.

The goal of this study was to develop and validate an MR-based model for clinical outcomes. We first introduced a short-term clinical assessment of AIS patients to generate clinical labels. Second, quantitative analyses of the global and regional parameters of the ischemic penumbra and core area were performed. The potential baseline clinical and radiological signatures of AIS patients were screened using machine learning methods based on a short-term clinical label. Finally, predictive models were validated in an independent dataset, with both short-term and long-term clinical labels. We hypothesized that these models would help to select patients eligible for thrombolysis and to develop personalized therapy during AIS management.

2. Materials and methods

2.1. Study design

In this study, we developed predictive models for AIS patients including two independent datasets which were collected from seven independent hospitals, as listed in Supplementary Table S1. Firstly, the training dataset was retrospectively collected in three hospitals from June 2012 to October 2016. After the training dataset were collected, we used a prospective dataset, which were collected from September 2008 to July 2010 [18], to validate our results. The study was approved by the Ethics Committees of all listed hospitals and informed consent was obtained. The study design is shown in Fig. 1.

The enrollment criteria were as follows: 1) diagnosis of early stroke; 2) age between 18 and 80 years; 3) absence of contraindications to MRI; 4) absence of brain tumor or pregnancy; 5) absence of intracranial hemorrhage on CT/MRI at baseline; 6) both diffusion-weighted imaging (DWI) and perfusion-weighted imaging (PWI) acquired within 9-hours after AIS onset; 7) intravenous recombinant tissue plasminogen activator (IV-rtPA) treatments or conventional medical treatments after MR scanning; and 8) completion of National Institutes of Health Stroke Scale (NIHSS) and modified Rankin Scale (mRS) assessments on Day-7 after onset for the training dataset; or completion of NIHSS on Day-7 and mRS assessments on both Day-7 and Day-90 after onset for the validation dataset.

The exclusion criteria were as follows: 1) poor image quality; 2) perfusion imaging failure; and 3) non-detection of lesion on DWI or PWI.

Certain clinical assessments were performed by experienced neuroradiologists to record the symptoms at onset. Specific information was obtained from each patient at baseline, namely, age, sex, time lag between onset and baseline MRI scan (T-M), baseline systolic blood pressure (SBP), and diastolic blood pressure (DBP). Additionally, other potential risk factors such as hypertension, diabetes, hyperlipidemia, atrial fibrillation, and previous stroke or transient ischemic attack (TIA) were collected. The radiological analyses had not yet been performed at the time of hospital admission and the decision-making regarding treatment was based on conventional clinical information.

2.2. Clinical assessments

NIHSS and mRS scores on Day-7 were recorded and combined for an early clinical outcome assessment. A combined score (MN) of mRS and NIHSS is defined below (Eq. (1.1)) and a favorable clinical outcome (FCO) was defined as MN ≥ 4. Unfavorable clinical outcome (UFCO) was defined as MN < 4. Furthermore, a long-term clinical label was evaluated on Day-90. Specifically, FCO was defined as mRS < 2 and UFCO was defined as mRS ≥ 2.

$$MN = \Delta NIHSS + (3 - mRS) * 4 \quad (1.1)$$

2.3. MR image processing and penumbral quantification

DWI and PWI sequences were acquired for all participants. The detailed acquisition parameters are listed in Supplementary Table S2. Pre-processing and statistical analyses of MRI data were carried out using tools from the FMRIB Software Library (www.fmrib.ox.ac.uk/fsl) [19,20]. In addition, MATLAB scripts developed in house were used in the preprocessing.

The DWI images were preprocessed as follows: 1) skull stripping on b_0 imaging; 2) calculation of the ADC maps; and 3) estimation of a rigid transformation from the individual space to the MNI152 standard brain (FNIRT in FSL, 2 mm isotropic). The PWI images were preprocessed as follows: 1) detection of the baseline image; 2) alignment of each volume to the baseline image; 3) smoothing of the 4D images using a 6-mm FWHM Gaussian kernel; 4) estimation of a rigid transformation from the perfusion baseline images to the DWI image space; and 5) generation of a linear transformation between the individual space to the MNI152 template (FNIRT in FSL, 2 mm isotropic). The perfusion image calculation was based on a standard singular value decomposition model [21], including cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT) and Tmax.

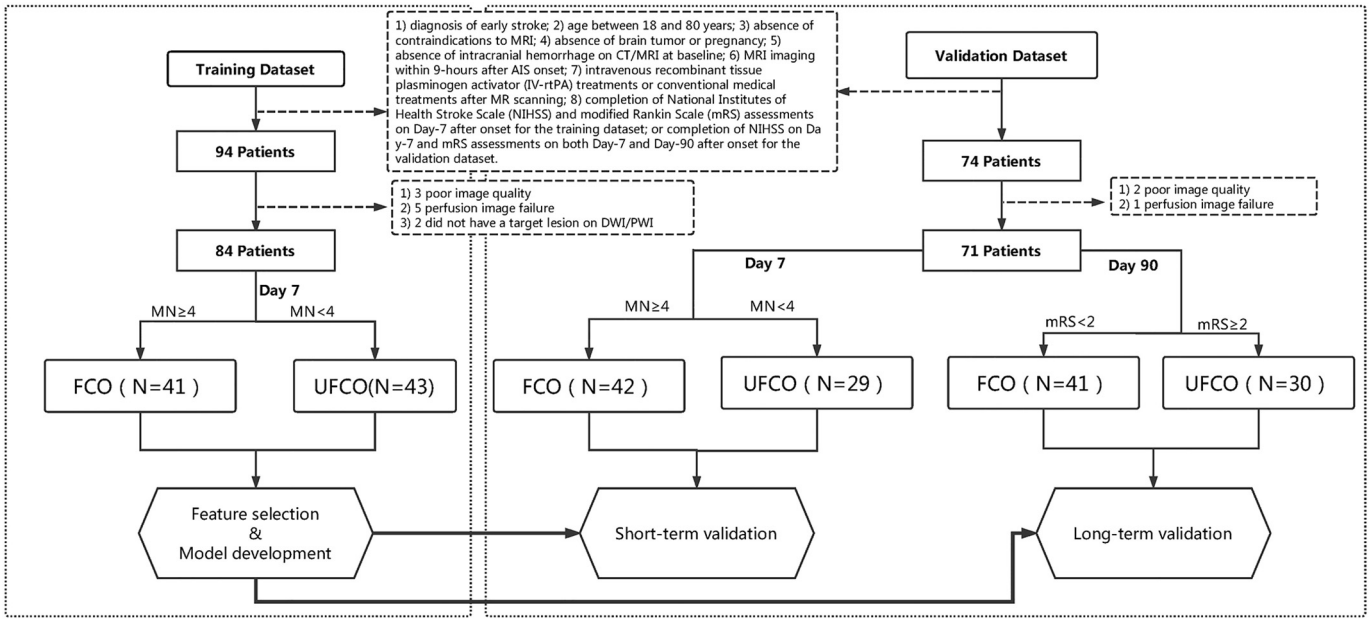

Fig. 1. Study design. FCO = favorable clinical outcome; UFCO = unfavorable clinical outcome; DWI = diffusion-weighted imaging; PWI = perfusion-weighted imaging; MN is defined in Eq. (1.1); mRS = modified Rankin scale.

Before applying mismatch, all maps were transferred and resampled to b_0 (DWI space), resampling with a 2-mm voxel size. The quantitative penumbral analysis was conducted with the following steps. Firstly, the brain mask (see Fig. 3b) was generated and applied to all calculated maps in the DWI space. Second, penumbral segmentation was performed based on a mismatch model [22,23]. Thirdly, with a clustering filter, any cluster <9 voxels (72 mm^3) was excluded from the final results. Finally, the penumbral identification was reconstructed into 3-dimensional visualization [24] and the results are shown in Fig. S1. In comparison with the traditional 2D demonstration of penumbral identification, a 3D reconstruction model clearly showed a jagged interlocking pattern of the penumbra. After image preprocessing and penumbral identification, the images were reviewed by two radiologists (Y.C. and D.L.Z.).

Meanwhile, the cortex labels, as shown in Fig. S2, were transferred to the native space to identify the positions of the ischemic penumbra and core area based on the invert transfer matrix (See Fig. 3b, labeled maps). The labeled atlas was based on the MNI space (MNI-maxprob-thr0-2 mm.nii.gz). After merging the thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and accumbens into one area, the gray matter of the brain was divided into six separate regions and labeled accordingly, as shown in Fig. S2. Along with the white matter area, the whole brain was divided into seven local regions for further analysis. In addition, a group of three parameters (volume ratio, ADC and CBF) within each separate region was defined as the regional features, and the penumbra and core area were separately calculated for each group. The value of the intact brain region was replaced with the mean value of the region. For regional quantification, 42 features from the seven separate regions were extracted and are listed in Table S3.

MIS was defined to quantify the ischemic penumbra and defected area of AIS patients (Eq. (1.2)). The volumes of the core area were defined by DWI defected area volume, and the volumes of penumbra were defined by PWI defected with intact DWI maps. V_{PWI} represented the volume of PWI defected areas, V_{PWI} represented the volume of DWI defected areas and V_{MIX} represented the volume of both the DWI and PWI defected areas.

$$MIS = (V_{PWI} - V_{DWI})/(V_{PWI} + V_{DWI} - V_{mix}) \quad (1.2)$$

Global ischemic penumbra and core variables including core region volume, penumbral region volume, ADC/CBF of the core region, ADC/CBF value of the penumbra region were quantitatively calculated, as well as 42 regional features of the penumbra and core area (see Supplementary Table S3). The value of the intact brain region was replaced with the mean value of the region.

2.4. Statistical analyses

2.4.1. Clinical measurements

SPSS Statistics software (Version 21.0, IBM) was used for the statistical analyses. Two-sample *t*-tests and Pearson's correlation analyses were employed and the reported significance levels were all two-sided, with statistical significance set at 0·05. Data are presented as mean [standard deviation (SD)], median [interquartile range (IQR)], and number (percentage), as appropriate.

2.4.2. Feature selection

Based on the short-term clinical labels, the clinical and radiological features were evaluated with a feature selection procedure. In our study, the Least Absolute Shrinkage and Selection Operator method (LASSO) logistic regression model [25] was conducted in the training dataset to select the most predictive features for clinical outcome.

To compare the predictive capabilities of the clinical features, MIS and regional features (see Supplementary Table S3), three different models were evaluated as listed below.

Clinical model: The input of this model contained 12 potential variables that were purely clinical, namely, sex, age, baseline SBP, baseline DBP, hypertension, atrial fibrillation, diabetes, hyperlipidemia, previous stroke or TIA, T-M, baseline NIHSS, and treatment option.

Mismatch model: The input of this model covered all 12 potential variables listed in Clinical model plus MIS.

Combined model: The input of this model contained 61 potential variables covering clinical information, global and regional information, including the 12 clinical variables listed in Clinical model, seven global ischemic penumbra and core variables (MIS, core region volume, penumbral region volume, ADC/CBF of the core region, ADC/CBF value of the penumbra region), and 42 regional features of the penumbra and core area (see Supplementary Table S3).

2.4.3. Development and validation of the predictive models

Based on the short-term clinical label, the clinical model, mismatch model, and combined model were independently validated. Moreover, the models were further tested with a long-term clinical label which was defined based on mRS scores assessed on Day-90.

The predictive models were developed and validated using WEKA software (http://www.cs.waikato.ac.nz/ml/weka/). Logistic model tree (LMT) models were developed based on the selected features in the training dataset [26]. The performances of the predictive models were evaluated with receiver operating characteristic (ROC) analysis, and pairwise ROC comparisons between our models were tested using Delong method [27]. Model calibration was assessed with the Hosmer–Lemeshow goodness-of-fit test [28]. To estimate the clinical utility of our models, decision curve analysis was performed by calculating the net benefits for a range of threshold probabilities in the validation dataset.

To avoid algorithm bias, we applied four machine learning methods and used reclassification methods to evaluate the models, the detailed information was list in Supplementary S1. The packages used in R for the discrimination and calibration of the predictive models were also list in Supplementary S1.

3. Results

3.1. Clinical assessment

As shown in Figs. 1, 5 patients were excluded because of poor image quality, 6 patients were removed because of the failure to acquire a perfusion image, and another 2 patients were excluded because of the absence of lesions on the DWI/PWI maps. Finally, a total of 155 patients were included in the final data analyses, with 84 in the training dataset (age from 43 to 80, 59 males) and the remaining 71 in the validation dataset (age from 36 to 80, 45 males). The mRS scores of the AIS patients are shown in Fig. 2. For the short-term clinical label, there were 41 patients (48·8%) in the FCO group in the training dataset and 42 patients (59·2%) in the FCO group in the validation dataset. On Day-90, there were 41 patients (57·7%) in the FCO group.

The characteristics of the AIS patients at baseline are shown in Table 1. Significant differences were evident for SBP and IV-rtPA involvement in both training and validation datasets. Age, Sex, baseline NIHSS score and hyperlipidemia involvement were detected to be significantly different between the FCO group and UFCO group in the training dataset while no difference was found in the validation dataset. In the validation dataset, there were more hypertension patients in the UFCO group than in the FCO group.

3.2. Penumbral quantification

The quantification results of the penumbra and core area are shown in Fig. 3. As shown in Fig. 3b, the penumbra and core area were labeled green and red. The probability of core and penumbral area in all the AIS patients in the training and validation datasets is shown in Supplementary Fig. S3. The relationship between MIS and the clinical assessments MN was investigated on Day-7 are shown in Fig. 4a. On Day-7, there were significant correlations between MIS and clinical assessments such as NIHSS score, mRS score, MN, and improvement of NIHSS score between baseline and Day-7 in both datasets (see Table 2).

3.3. Feature selection and model development

In clinical model, five features survived in the feature selection, namely, sex, baseline SBP, T-M, baseline NIHSS, and the treatment option. In MIS model, four features survived, namely, baseline SBP, baseline NIHSS, treatment option, and MIS. In combined model, seven features survived, namely, baseline SBP, baseline NIHSS, treatment option, MIS, ratio of the core area in the temporal lobe (R.C.T), ADC value in the white matter area of the penumbra (ADC.C.W) and CBF value in the occipital lobe of the penumbra (CBF.P.O).

As shown in Table 3, significant differences were detected in MIS, R.C.T, and CBF.P.O in both datasets, whereas, ADC.C.W was significantly different in the training dataset alone. The predictive models were developed based on LMT and are outlined below. The scatter plots of MN and the output of the combined model of all of our patients are shown in Fig. 4b.

The model calculates the predicted probability of the models as follows:

$$P_{Clinical} = 4 \cdot 44 - Sex^{*}0 \cdot 50 - SBP^{*}0 \cdot 02 - T - M^{*}0 \cdot 02 - N_0{}^{*}0 \cdot 12 + Treatment^{*}0 \cdot 80 \quad (1.3)$$

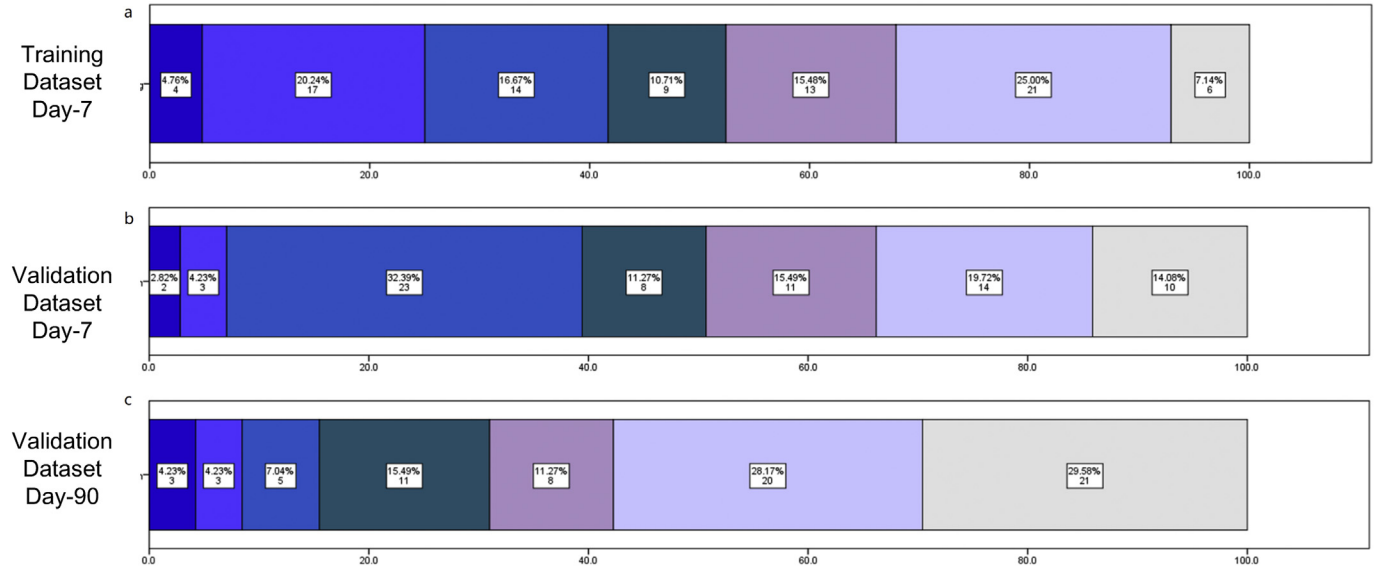

Fig. 2. Scores on the modified Rankin Scale (mRS) on Day-7 in the training dataset and on Day-7 and Day-90 in the validation dataset. 0, no symptoms; 1, no clinically relevant disability; 2, slight disability (able to look after own affairs without assistance but not to a full extent); 3, moderate disability (requires some help but able to walk unassisted); 4, moderately severe disability (requires assistance and unable to walk unassisted); 5, severe disability (requires constant nursing care); 6, dead.

Table 1
Characteristics of the AIS patients at baseline in training and validation datasets.

	Training dataset			Validation dataset		
	UFCO ($N = 43$)	FCO ($N = 41$)	*P*-value	UFCO ($N = 29$)	FCO ($N = 42$)	*P*-value
General clinical information						
Sex(Male)	26 (60·5%)	33 (80·5%)	0·044	21 (72·4%)	24 (57·1%)	0·187
Age,years, median (Range)	67 (43–80)	66 (47–80)	0·041	62 (36–80)	61·5 (40–77)	0·797
Risk factors						
Hypertension	28 (65·1%)	28 (68·3%)	0·761	19 (65·5%)	17 (40·5%)	0·038
Atrial fibrillation	10 (23·3%)	10 (24·4%)	0·904	2 (7·0%)	8 (19·0%)	0·123
Diabetes	14 (32·6%)	9 (22·0%)	0·280	3 (10·0%)	3 (7·0%)	0·650
Hyperlipidemia	3 (7·0%)	10 (24·4%)	0·030	3 (10·0%)	7 (17·0%)	0·443
Previous stroke or TIA	6 (14·0%)	10 (24·4%)	0·231	2 (7·0%)	4 (10·0%)	0·693
Baseline information						
T-M (hours)	4·39 ± 1·99	3·60 ± 1·63	0·051	4·54 ± 1·73	4·50 ± 2·00	0·931
SBP (mmHg)	156·65 ± 16·36	145·90 ± 21·04	0·011	159·55 ± 21·38	143·43 ± 23·17	0·004
DBP (mmHg)	88·35 ± 13·75	85·15 ± 13·56	0·286	93·41 ± 16·44	84·93 ± 15·07	0·031
NIHSS score, median (Range)	10 (2–24)	6 (1–14)	<0·001	10 (4–21)	8·5 (4–21)	0·104
IV–rtPA	16 (37·2%)	25 (61·0%)	0·030	10 (34·0%)	26 (62·0%)	0·023

Note: Data are presented as mean ± SD or n (%). SBP = systolic blood pressure; DBP = diastolic blood pressure; T-M = time from stroke onset to MRI scan before treatment; IV-rtPA = intravenous tissue plasminogen activator; FCO = favorable clinical outcome; UFCO = unfavorable clinical outcome. FCO and UFCO were defined based on MN on Day-7; MN is defined in Eq. (1.1).

$$P_{MIS} = 2 \cdot 40 - SBP^{*}0 \cdot 02 - N_0{}^{*}0 \cdot 12 + Treatment^{*}0 \cdot 74 + MIS^{*}1 \cdot 15 \quad (1.4)$$

$$P_{Combined} = 0 \cdot 69 - SBP^{*}0 \cdot 01 - N_0{}^{*}0 \cdot 09 + Treatment^{*}0 \cdot 54 + MIS^{*}1 \cdot 15 + R.C.T^{*}2 \cdot 61 + ADC.C.W^{*}0 \cdot 99 + CBF.P.O^{*}2 \cdot 73 \quad (1.5)$$

In Eq. (1.3), Sex is defined as 1 with male patients, while 2 with female patients. In Eq. (1.3)–(1.5), Treatment is defined as 1 with IV-rtPA while Treatment is defined as 0 with conventional medical treatment. In Eq. (1.3)–(1.5), N_0 is the NIHSS score after onset.

3.4. Model validation

The ROC curves of the predictive models are shown in Fig. 5. Based on the short-term validation, the AUCs of the clinical model, MIS model, and combined model were 0·743 (95% CI: 0·629–0·856), 0·854 (95% CI: 0·767–0·941), and 0·863 (95% CI: 0·774–0·951), respectively. With pairwise ROC comparisons, there are significant differences between clinical model vs mismatch model ($P = 0·002$) and

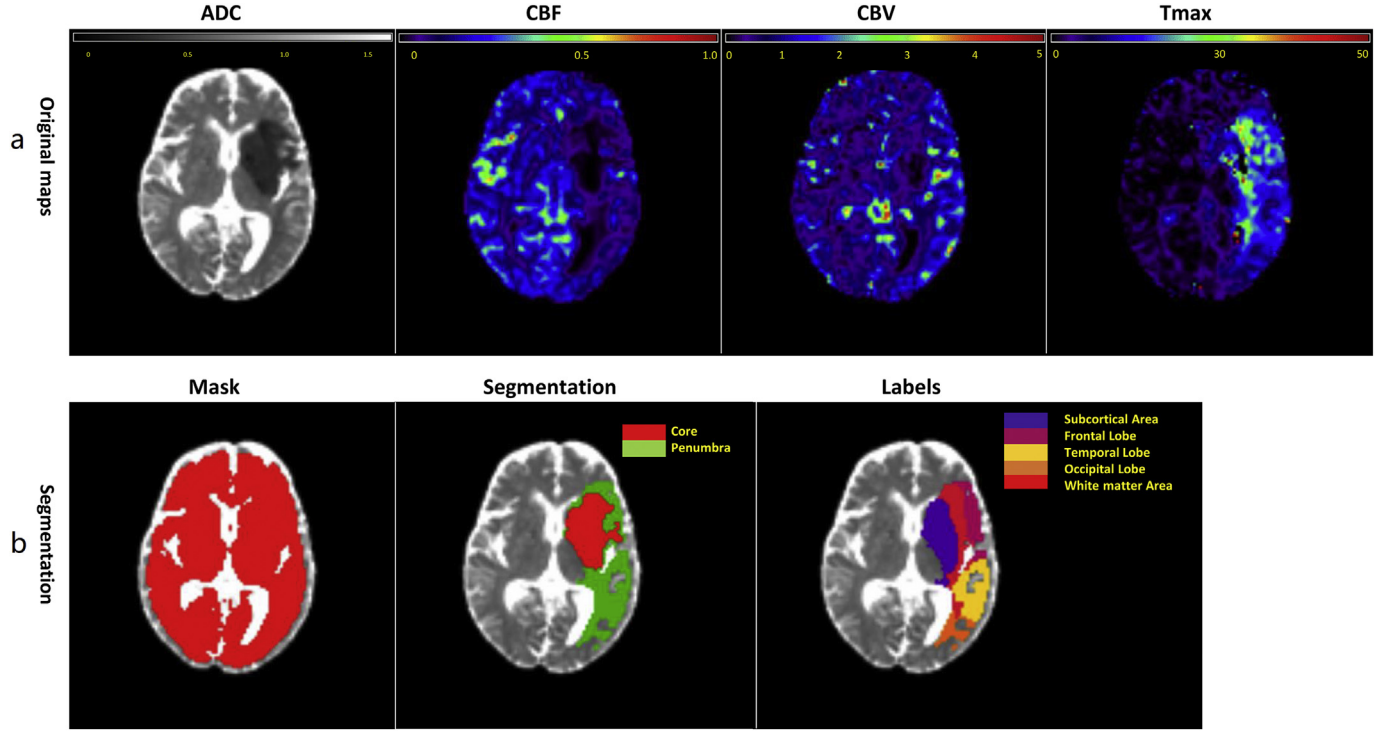

Fig. 3. Segmentation of the penumbra and core area. (a) Original maps. ADC map, CBF map, CBV map and Tmax map; (b) mask, segmentation and labeled results. The penumbra is shown in green and the core area is shown in red. ADC = Apparent diffusion coefficient; CBF = cerebral blood flow; CBV = cerebral blood volume; DWI = diffusion-weighted imaging; PWI = perfusion-weighted imaging.

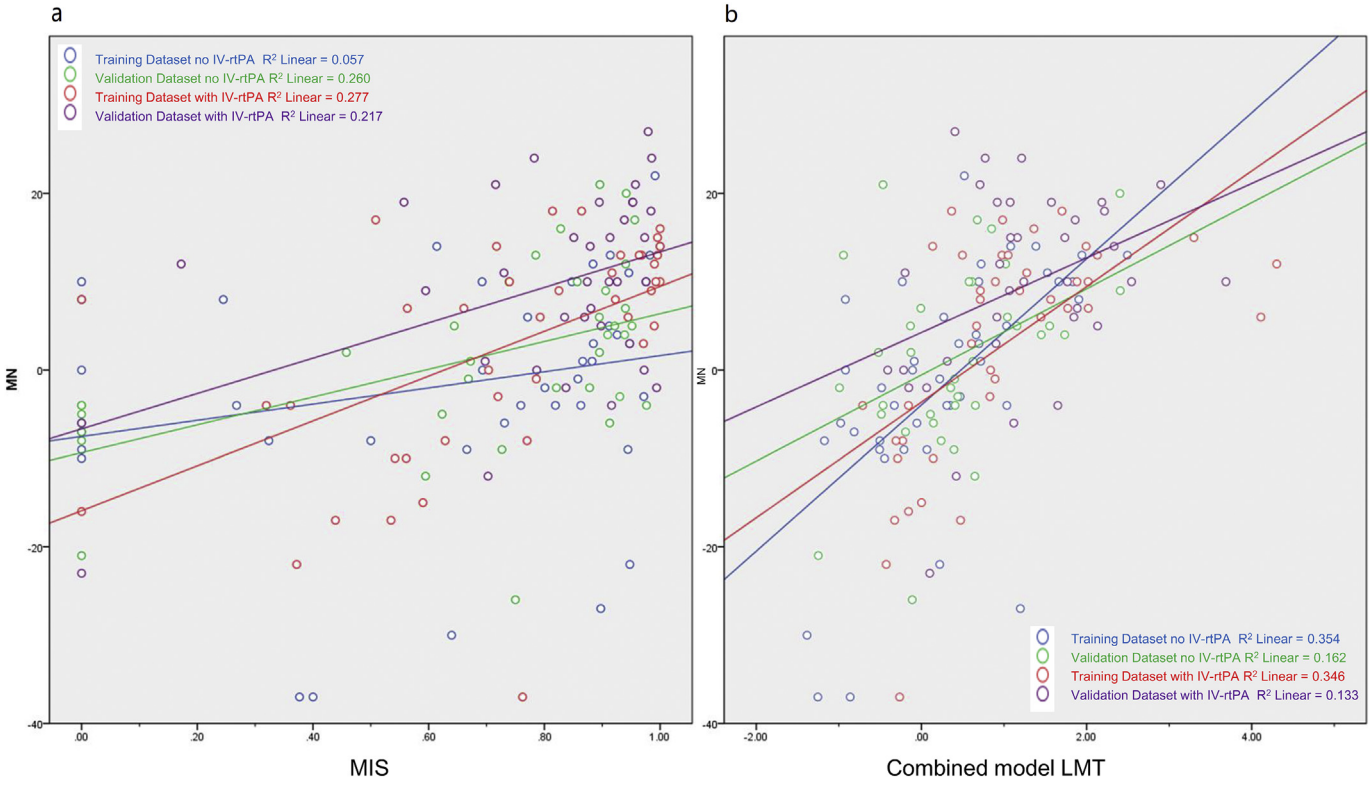

Fig. 4. Scatter plots of the clinical assessments and radiological measurements. (a) Scatter plots of MIS and MN; (b) Scatter plots of the outputs of the combined LMT model and MN; MN is defined in Eq. (1.1); MIS represents quantification of the ratio of ischemic penumbra to the defected area, which is defined in Eq. (1.2); IV–rtPA = Intravenous tissue plasminogen activator; LMT = logistic model tree.

clinical model vs. combined model ($P = 0{\cdot}003$). There is no significant difference between our two imaging based models: mismatch model and combined model ($P = 0{\cdot}749$). In addition, we analyzed the predict performance of the combined model of short-term validation in patients with or without IV-rtPA treatments. The AUC in the patients with IV-rtPA is 0·869 (95% CI: 0·739–1·000), while in patients without IV-rtPA is 0·799 (95% CI: 0·636–0·963). The Hosmer–Lemeshow test of model calibration with a non-significant *p* value showed favorable calibration of our models, for short-term models: $P_{clinical\ model} = 0{\cdot}778$; $P_{mismatch\ model} = 0{\cdot}202$; $P_{combined\ model} = 0{\cdot}450$. The decision curves of our models in the validation datasets were shown in Fig. S4, indicating the clinical usefulness of our models.

Furthermore, based on the long-term validation, the AUCs of the clinical model, MIS model, and combined model were 0·697 (95% CI: 0·573–0·821), 0·773 (95% CI: 0·662–0·884) and 0·778 (95% CI: 0·668–0·888). In the long-term models, there is significant difference between clinical model vs mismatch model ($P = 0.030$). There is no significant difference between clinical model vs. combined model ($P = 0{\cdot}058$) and mismatch model vs. combined model ($P = 0{\cdot}870$). The Hosmer–Lemeshow test of model calibration with a non-significant p value showed favorable calibration of our models, for long-term models: $P_{clinical\ model} = 0{\cdot}959$; $P_{mismatch\ model} = 0{\cdot}649$; $P_{combined\ model} = 0{\cdot}437$.

4. Discussion

In this study, predictive models for individual clinical outcome after AIS onset was developed and validated with both short-term and long-term clinical labels. The combined model included both clinical characteristics and advanced MR penumbra profiles derived from machine learning methods. The predictive power of the model was validated using an independent dataset and reached AUCs of 0·863 in the short-term validation and 0·778 in the long-term validation, indicating a favorable discriminative ability.

4.1. Clinical labels

To identify the potential features at baseline and to develop predictive models for further clinical outcomes of AIS patients, reasonable clinical labels were needed. It has been shown that the early mRS can serve

Table 2
Pearson's correlation between the MIS and clinical assessments in the training and validation datasets.

		T-M	N_0	N_7	mRS_7	mRS_{90}	MN	ΔNIHSS
Training dataset	Pearson's correlation	−0·308	−0·091	−0·278	−0·360	–	0·370	0·313
	P-value	0·004	0·410	0·011	0·001	–	0·001	0·004
	N	84	84	84	84	–	84	84
Validation dataset	Pearson's correlation	−0·173	0·179	−0·363	−0·402	−0·299	0·513	0·528
	P-value	0·150	0·135	0·002	0·001	0·011	<0·001	<0·001
	N	71	71	71	71	71	71	71

Note: MIS represents quantification of the ratio of ischemic penumbra to the defected area, which is defined in Eq. (1.2); T-M = time lag between onset and the baseline MRI scan; NIHSS = National Institutes of Health Stroke Scale; N_0 = baseline NIHSS score; N_7 = NIHSS score on Day-7; mRS = modified Rankin Scale; mRS_7 = mRS score on Day-7; mRS_{90} = mRS score on Day-90; ΔNIHSS = the improvement of NIHSS score between baseline and Day-7.

Table 3
Radiological characteristics of AIS patients with FCO versus UFCO in the training and validation datasets.

	Training dataset			Validation dataset		
	FCO (N = 41)	UFCO (N = 43)	*P*-value	FCO (N = 42)	UFCO (N = 29)	*P*-value
Radiological characteristics						
MIS	0·800 ± 0·280	0·584 ± 0·309	<0·001	0·858 ± 0·150	0·578 ± 0·384	0·001
R.C.T	0·146 ± 0·263	0·040 ± 0·064	0·033	0·198 ± 0·216	0·113 ± 0·138	0·049
ADC.C·W	0·615 ± 0·219	0·485 ± 0·152	0·001	0·490 ± 0·179	0·513 ± 0·165	0·574
CBF.P·O	0·162 ± 0·107	0·114 ± 0·066	0·006	0·196 ± 0·099	0·144 ± 0·070	0·012

Note: Data are presented as mean ± SD. MIS represents quantification of the ratio of ischemic penumbra to the defected area, which is defined in Eq. (1.2); AIS = acute ischemic stroke; FCO = favorable clinical outcome; UFCO = unfavorable clinical outcome; FCO and UFCO were defined by MN on Day-7; MIS = mismatch ratio; R.C.T = ratio of the core area in the temporal lobe; ADC.C.W = ADC value in the white matter area of the core area; CBF.P.O = CBF value in the occipital lobe of the penumbra.

as a good proxy for Day-90 outcomes [29,30], suggesting that short-term and long-term clinical assessments had a similar treatment effect magnitude. A combined score was used to generate an early clinical label to discriminate potential clinical and radiological features using machine learning methods. With external validation, this combined score has demonstrated a reasonable ability to distinguish the relative features of AIS patients at baseline.

4.2. Penumbral quantification

Several early studies reported that DWI/PWI mismatch was correlated with clinical assessments [31,32]. Consistently, our study found that MIS was significantly correlated with clinical outcomes, suggesting that the penumbra, the main target of salvageable brain tissue, is associated with FCOs with adequate medical care [4]. With penumbral imaging, two clinical trials based on diffusion-perfusion were performed to select patients for thrombolysis beyond the current time window, but failed to reach the endpoints [11,33]. Although the exact reasons for the failure remain unknown, the inadequacy of the penumbral identification methodology may have led to the negative results. Our study aimed to develop a quantitative penumbral method to evaluate the extent of the global penumbral existence and local tissue characteristics. Compared with global signatures, the regional features of the penumbra and core area might capture more information regarding tissue impairment after AIS [16,17] and may help predict clinical outcomes.

4.3. Predictive model performance

Three different models were developed in our study, the combined model included both clinical and radiological signatures had the best ability to predict the 3-month clinical outcomes of AIS patients. These results might suggest that with machine learning methods, the regional features of the penumbra and core area are integrated with the clinical factors and treatment option to affect clinical outcome. A recent study developed a multivariable model, including clinical information and CT assessments, to predict clinical outcomes [34]. Consistent with our results, SBP, baseline NIHSS, and IV-rtPA treatment were confirmed to be independent predictors of clinical outcomes. In contrast, rather than via visual assessment, the features in our model were derived from machine learning techniques, which included more objective radiological signatures. These favorable results demonstrate that with machine learning techniques, essential clinical and radiological features can be revealed and predictions of clinical outcomes for AIS patients can be more accurately identified than with conventional methods. This interpretation could be supported by the external validation in the independent dataset in our study, which also demonstrated a favorable prediction performance based on the combined model with both short-term and long-term clinical labels.

4.4. Clinical Implications

Although penumbral imaging is already widely acknowledged in clinical practice, penumbral imaging for selecting candidates for thrombolysis has not been validated [35]. Several clinical trials, including EPITHET, DIAS II, and MR RESCUE, have failed to confirm the clinical validity of penumbral imaging [11,33,36], largely because of improper penumbral definition. In the current study, our model enabled the prediction of favorable clinical outcome for each individual up to 9 h after stroke onset. By providing a promising strategy in detail to guide current patient selection in AIS management, the application of our models will be particularly meaningful to hospitals that are not competent enough to be qualified for endovascular treatment.

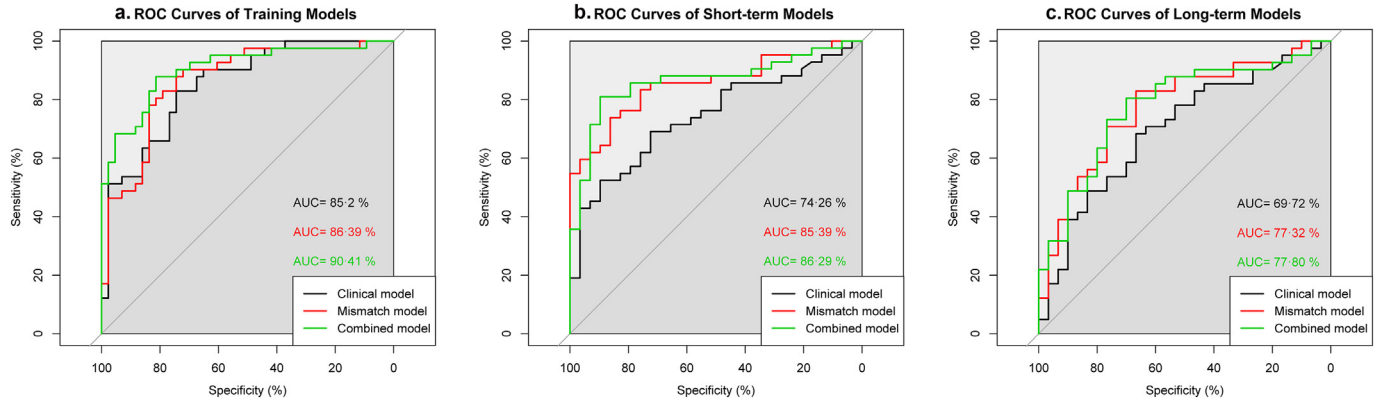

Fig. 5. ROC curves of predictive models. (a) ROC curves of the models in the training dataset; (b) ROC curves of models in the validation dataset, with short-term labels; (c) ROC curves models in the validation dataset, with long-term labels. Black lines indicate the ROCs of the clinical models. Red lines indicate the ROCs of the mismatch models. Green lines indicate the ROCs of the combined models; ROC = Receiver operating characteristics; AUC = area under the curve.

4.5. Limitations

The main limitation of this study is that the treatment choice was not randomized and there might have been selection bias and information bias. With an independent validation, these biases might have been controlled. However, the observational nature of this study may have still limited the interpretation of current results. Further validations in predesigned clinical trials are recommended. Second, the sample size was relatively small and imaging protocols for current data varied from one center to another, so the homogeneity of the sample may have been affected. The application of the models to a relatively larger cohort with a prospective design would generate more convincible evidence to guide current patient selection in AIS management.

5. Conclusions

In summary, the proposed imaging-based models contain both clinical and imaging signatures, showing sufficiently high discrimination for clinical outcomes of AIS patients. Despite patient heterogeneity, the model still acquired relatively high AUCs, indicating the robustness and reliability of our methodology. This model may be clinically useful because the model enables patient selection for thrombolysis either within or beyond the 4·5-h time window, especially in hospitals that are not qualified for endovascular treatments.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.07.028.

Acknowledgements

The authors would like to acknowledge Pei-Yi Gao from the Beijing Tiantan Hospital for providing the validation dataset collected from the following institutes: the Beijing Tiantan Hospital, the First Affiliated Hospital of Wenzhou Medical University, the Guangdong Hospital of Traditional Chinese Medicine, the Shanghai Pudong New Area People's Hospital and the Tianjin Huanhu Hospital.

Funding sources

This work was supported by the Major State Basic Research Development Program of China (973 Program; 2013CB733803), the National Science Fund for Distinguished Young Scholars (81525014), the National Key Research and Development Program of China (2017YFA0104302), the Jiangsu Provincial Special Program of Medical Science (BL2013029), and the Key Research and Development Program of Jiangsu Province (BE2016782).

Declaration of interests

The authors declare no competing interests

Author contributions

G.J.T. was study chair and principal investigator; T.Y.T. drafted the initial manuscript, which was reviewed by all the other authors. T.Y.T., Y.J., S.H.J., and G.J.T. designed and carried out the study. T.Y.T. and Y.J. analyzed the MR data. Y.C., C.H.Z., S.H.J., and G.J.T. provided extensive critical insights and revisions of all drafts of the manuscript. D.L.Z., Z. Y., C.Y. P., Y.J.Y, X.D.Y. and P.Y.G. enrolled patients. Y.C. and D.L.Z. reviewed the image processing results. All authors contributed to the final version of the manuscript.

T.Y.T. and Y.J. contributed equally to this study.

References

[1] GBD Mortality. Causes of Death C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016; 388(10053):1459–544.
[2] Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke 2018;49(3):e46-110.
[3] Lees KR, Bluhmki E, von Kummer R, Brott TG, Toni D, Grotta JC, et al. Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. Lancet 2010;375(9727): 1695–703.
[4] Davis S, Donnan GA. Time is Penumbra: imaging, selection and outcome. The Johann jacob wepfer award 2014. Cerebrovasc Dis 2014;38(1):59–72.
[5] Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. N Engl J Med 2008; 359(13):1317–29.
[6] Wahlgren N, Ahmed N, Davalos A, Hacke W, Millan M, Muir K, et al. Thrombolysis with alteplase 3-4.5 h after acute ischaemic stroke (SITS-ISTR): an observational study. Lancet 2008;372(9646):1303–9.
[7] Zhang B, Sun XJ, Ju CH. Thrombolysis with alteplase 4.5-6 hours after acute ischemic stroke. Eur Neurol 2011;65(3):170–4.
[8] Kawano H, Bivard A, Lin L, Ma H, Cheng X, Aviv R, et al. Perfusion computed tomography in patients with stroke thrombolysis. Brain 2017;140(3):684–91.
[9] Kim BJ, Kang HG, Kim HJ, Ahn SH, Kim NY, Warach S, et al. Magnetic resonance imaging in acute ischemic stroke treatment. J Stroke 2014;16(3):131–45.
[10] Sabarudin A, Subramaniam C, Sun Z. Cerebral CT angiography and CT perfusion in acute stroke detection: a systematic review of diagnostic value. Quant Imaging Med Surg 2014;4(4):282–90.
[11] Davis SM, Donnan GA, Parsons MW, Levi C, Butcher KS, Peeters A, et al. Effects of alteplase beyond 3 h after stroke in the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET): a placebo-controlled randomised trial. Lancet Neurol 2008; 7(4):299–309.
[12] Lansberg MG, Straka M, Kemp S, Mlynash M, Wechsler LR, Jovin TG, et al. MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2): a prospective cohort study. Lancet Neurol 2012;11(10):860–7.
[13] Mishra NK, Albers GW, Davis SM, Donnan GA, Furlan AJ, Hacke W, et al. Mismatch-based delayed thrombolysis: a meta-analysis. Stroke 2010;41(1):e25–33.
[14] Campbell BC, Christensen S, Foster SJ, Desmond PM, Parsons MW, Butcher KS, et al. Visual assessment of perfusion-diffusion mismatch is inadequate to select patients for thrombolysis. Cerebrovasc Dis 2010;29(6):592–6.
[15] Rosso C, Colliot O, Pires C, Delmaire C, Valabregue R, Crozier S, et al. Early ADC changes in motor structures predict outcome of acute stroke better than lesion volume. J Neuroradiol 2011;38(2):105–12.
[16] Rosso C, Samson Y. The ischemic penumbra: the location rather than the volume of recovery determines outcome. Curr Opin Neurol 2014;27(1):35–41.
[17] Munsch F, Sagnier S, Asselineau J, Bigourdan A, Guttmann CR, Debruxelles S, et al. Stroke location is an independent predictor of cognitive outcome. Stroke 2016; 47(1):66–73.
[18] Wang Y, Liao X, Zhao X, Wang C, Liu L, Zhou Y, et al. Imaging-based thrombolysis trial in acute ischemic stroke-II (ITAIS-II). Int J Stroke 2009;4(1):49–53 [discussion 49].
[19] Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, et al. Bayesian analysis of neuroimaging data in FSL. Neuroimage 2009;45(1 Suppl):S173–86.
[20] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004;23(Suppl. 1):S208–19.
[21] Wu O, Ostergaard Lweisskoff RM, Benner T, Rosen BR, Sorensen AG. Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix. Magn Reson Med 2003;50(1):164–74.
[22] Zhang S, Tang H, Yu YN, Yan SQ, Parsons MW, Lou M. Optimal magnetic resonance perfusion thresholds identifying ischemic penumbra and infarct core: a Chinese population-based study. CNS Neurosci Ther 2015;21(3):289–95.
[23] Lansberg MG, Lee J, Christensen S, Straka M, De Silva DA, Mlynash M, et al. RAPID automated patient selection for reperfusion therapy: a pooled analysis of the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET) and the Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution (DEFUSE) Study. Stroke 2011;42(6):1608–14.
[24] Xia M, Wang J, He Y. BrainNet viewer: a network visualization tool for human brain connectomics. Plos One 2013;8(7):e68910.
[25] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16(4):385–95.
[26] Landwehr N, Hall M, Frank E. Logistic model trees. Mach Learn 2005;59(1–2): 161–205.
[27] Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–45.
[28] Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997;16(9):965–80.
[29] Bruce O, Jeffrey LS. Day-90 acute ischemic stroke outcomes can be derived from early functional activity level. Cerebrovasc Dis 2009;1:50.
[30] Gumbinger C, Reuter B, Stock C, Sauer T, Wietholter H, Bruder I, et al. Time to treatment with recombinant tissue plasminogen activator and outcome of stroke in clinical practice: retrospective analysis of hospital quality assurance data with comparison with results from randomised clinical trials. BMJ 2014;348:g3429.
[31] Singer OC, Du Mesnil De Rochemont R, Foerch C, Stengel A, Sitzer M, Lanfermann H, et al. Early functional recovery and the fate of the diffusion/perfusion mismatch in

patients with proximal middle cerebral artery occlusion. Cerebrovasc Dis 2004; 17(1):13–20.
[32] Hillis AE, Wityk RJ, Barker PB, Ulatowski JA, Jacobs MA. Change in perfusion in acute nondominant hemisphere stroke may be better estimated by tests of hemispatial neglect than by the National Institutes of Health Stroke Scale. Stroke 2003;34(10): 2392–6.
[33] Hacke W, Furlan AJ, Al-Rawi Y, Davalos A, Fiebach JB, Gruber F, et al. Intravenous desmoteplase in patients with acute ischaemic stroke selected by MRI perfusion-diffusion weighted imaging or perfusion CT (DIAS-2): a prospective, randomised, double-blind, placebo-controlled study. Lancet Neurol 2009;8(2):141–50.
[34] Venema E, Mulder M, Roozenbeek B, Broderick JP, Yeatts SD, Khatri P, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomised trials. BMJ 2017;357:j1710.
[35] Dorado L, Millan M, Davalos A. Reperfusion therapies for acute ischemic stroke: an update. Curr Cardiol Rev 2014;10(4):327–35.
[36] Kidwell CS, Jahan R, Alger JR, Schaewe TJ, Guzy J, Starkman S, et al. Design and rationale of the Mechanical Retrieval and Recanalization of Stroke Clots Using Embolectomy (MR RESCUE) Trial. Int J Stroke 2014;9(1):110–6.