SVN Stroke and Vascular Neurology

Deep learning for automatically predicting early haematoma expansion in Chinese patients

Jia-wei Zhong ,[1] Yu-jia Jin,[1] Zai-jun Song,[1] Bo Lin,[2] Xiao-hui Lu,[3] Fang Chen,[4] Lu-sha Tong[1]

To cite: Zhong J, Jin Y, Song Z, *et al*. Deep learning for automatically predicting early haematoma expansion in Chinese patients. *Stroke & Vascular Neurology* 2021;**0**. doi:10.1136/svn-2020-000647

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/svn-2020-000647).

Received 26 September 2020
Revised 8 November 2020
Accepted 25 November 2020

[1]Department of Neurology, Zhejiang University School of Medicine Second Affiliated Hospital, Hangzhou, China
[2]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[3]State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University School of Mechanical Engineering, Hangzhou, China
[4]Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Correspondence to
Dr Lu-sha Tong;
2310040@zju.edu.cn

Dr Fang Chen;
chenfang@nuaa.edu.cn

ABSTRACT

Background and purpose Early haematoma expansion is determinative in predicting outcome of intracerebral haemorrhage (ICH) patients. The aims of this study are to develop a novel prediction model for haematoma expansion by applying deep learning model and validate its prediction accuracy.

Methods Data of this study were obtained from a prospectively enrolled cohort of patients with primary supratentorial ICH from our centre. We developed a deep learning model to predict haematoma expansion and compared its performance with conventional non-contrast CT (NCCT) markers. To evaluate the predictability of this model, it was also compared with a logistic regression model based on haematoma volume or the BAT score.

Results A total of 266 patients were finally included for analysis, and 74 (27.8%) of them experienced early haematoma expansion. The deep learning model exhibited highest C statistic as 0.80, compared with 0.64, 0.65, 0.51, 0.58 and 0.55 for hypodensities, black hole sign, blend sign, fluid level and irregular shape, respectively. While the C statistics for swirl sign (0.70; p=0.211) and heterogenous density (0.70; p=0.141) were not significantly higher than that of the deep learning model. Moreover, the predictive value for the deep learning model was significantly superior to that of the logistic model of haematoma volume (0.62; p=0.042) and the BAT score (0.65; p=0.042).

Conclusions Compared with the conventional NCCT markers and BAT predictive model, the deep learning algorithm showed superiority for predicting early haematoma expansion in ICH patients.

INTRODUCTION

Intracerebral haemorrhage (ICH) remains a devastating disease with high mortality and morbidity.[1] Early expansion of haematoma is a determinative factor in predicting outcome.[2] A wealth of studies has been carried out to seek an advanced way of early identifying patients with haematoma expansion. Spot sign was a well-established imaging marker associated with haematoma expansion, based on CT angiography (CTA).[3–5] However, in large-scaled clinical trials, the majority of ICH patients did not undergo CTA when admitted.[6] Imaging signs more feasible based on non-contrast CT (NCCT) included black hole sign, swirl sign, island sign, blend sign and so on.[7–10] However, these signs usually bared a poor sensitivity since their probability of occurrence is relatively low.[11 12] Recently, multi-itemed scores for predicting haematoma expansion expansion have been proposed in order to increase prediction efficacy, such as the BRAIN (B for baseline ICH volume, R for recurrent ICH, A for anticoagulation with warfarin, I for intraventricular haemorrhage and N for numbers of hours from onset to CT), HEAVN (H for heterogenity, E for peripheral edema, A for anticoagulation use, V for volume>30ml on initial CT, and N for Niveau formation), BAT (B for blend sign, A for hypodensity presence, and T for time from onset to NCCT). Yet none of these scores exhibits a C statistic >0.8, which means unsatisfactory prediction value.[12]

Nowadays, deep learning is recognised as the most effective machine learning algorithm and is cutting a striking figure in processing multi-categorical data. Some predicting models based on deep learning have been successfully used to predict clinical outcome. Convolutional neural networks (CNN) is a deep learning statistical method commonly used in image recognition.[13] U-net is a network and training strategy which relies on strong usage of data augmentation, and thereby requires fewer data and shorter time to produce an ideal method.[14] We hypothesised that CNN-derived U-net-supported modelling based on topological and morphological imaging features on NCCT might provide an advanced predicting model which retains both better prediction efficacy and convenient clinical application. For this purpose, a novel prediction model for ICH patients with early haematoma expansion is developed and tested compared with NCCT signs, as well as haematoma volume and BAT score.

METHODS

The data that support the findings of the study are available from the corresponding author on reasonable request. Retrospective analysis of a prospectively and consecutively collected cohort of spontaneous ICH patients in The Second Affiliated Hospital of Zhejiang University between February 2012 to October 2019. Patients were enrolled if they: (1) were over 18 years old, (2) were obtained baseline CT imaging within 8 hours from onset and (3) obtained follow-up CT scan after 20–24 hours after baseline. We excluded patients with infratentorial haematoma, ventricular haemorrhage only, critical deterioration or surgical operation before the follow-up CT imaging or poor image quality. From the prospectively longitudinal cohort, patients between February 2012 and August 2018 were selected for training dataset and those between September 2018 and October 2019 were selected as testing dataset. Early haematoma expansion was defined as an increase of >6 mL absolute volume or 33% relative volume of haematoma in the follow-up CT compared with initial CT.

Labelling and calculating the volume of haematoma

The baseline CT images for this study were obtained with four different CT scanners. The acquisition parameters were described in online supplemental table 1. Manual segmentations for haematoma were performed on the CT scans by a single radiologist with more than 10 years of experience, and 20 randomly selected cases were re-evaluated after a minimal interval of 7 days by another skilled radiologist to assess inter-rater reliability. The inter-rater agreement for the segmentation of haematoma was 0.978 for semi-automated segmentation method. Based on the binary label map, the haematoma volume was then calculated.

Assessment of the NCCT markers and the BAT score

The NCCT signs were evaluated by two authors, each with >10 years of experience. The inter-rater reliability was assessed using the whole dataset with a dichotomy (κ=0.886, 0.906, 0.848, 0.951, 0.892, 0.859, 0.932 for hypodensities, black hole sign, swirl sign, blend sign, fluid level, irregular shape, heterogeneous density, respectively). In addition, disagreements were settled by consensus between the two authors. The BAT score was based on the assessment of NCCT markers. The process of assessing the NCCT markers and the BAT score was shown in online supplementa tables 2 and 3.

Data preprocessing and augmentation

The baseline CT images were skull stripped and then normalised for signal intensity. The images were all resampled to a uniform field of view of 112×112×160 mm and matrix size of 256×256×32. Considering the limited data size, data augmentation was performed via three-dimensional rotation, translation and scaling. Thereby the size of original dataset was augmented 10-fold (see online supplemental file.

Training of the deep learning model

A two-output deep learning model was designed to segment the haematoma to acquire high-level imaging features and predict early haematoma expansion.

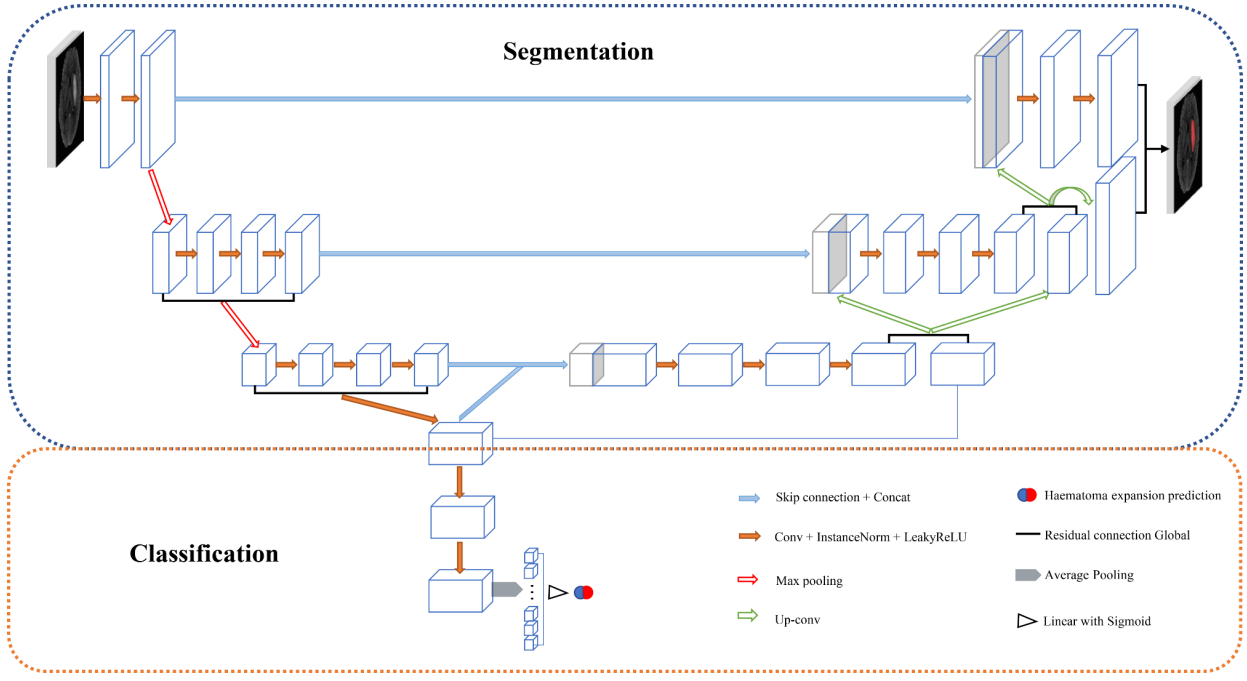


Figure 1 The concept of the model in this study: (1) the model had a single input (CT imaging data) and two outputs for segmentation and prediction; (2) based on the U architecture, the high-level image information derived from the bridge layer of U were treated as biomarkers for haematoma expansion prediction.

U-net provides technical support for basic architecture design for segmentation. The most aggregated contextual information was deposited in the middle layer of the model (the deepest convolutional layer of the U) and extracted as the high-level imaging feature used for a binary classification for haematoma expansion (figure 1). Detail of the model architecture is shown in online supplemental figure 1. For training, fivefold cross-validation was used for adjusting hyperparameters. These processes only involved CT data. No clinical information was applied as input. See online supplemental file.

Logistic regression models based on the haematoma volume and the BAT score

To compare the performance of the deep learning model, univariate logistic regression models were respectively developed based on the haematoma volume and the BAT score from the training dataset.

Evaluation of the models

Each model was applied in the testing dataset. The sensitivity, specificity, likelihood ratio and receiver operator characteristic (ROC) area under the curve (AUC) were calculated for each model and NCCT markers. Dice coefficient was calculated to evaluate the performance of segmentation.

Statistical analysis

Continuous variables are expressed as the mean with the CI of the mean or SD, or as the median with IQR. Continuous variables were compared using Mann-Whitney U test and Student's t-test as appropriate, and categorical variables were compared using Pearson's χ^2 test. AUC of deep learning model was compared with that of the other models and NCCT signs using Delong test. The statistical analysis was performed with R statistical software (V.4.0.1).

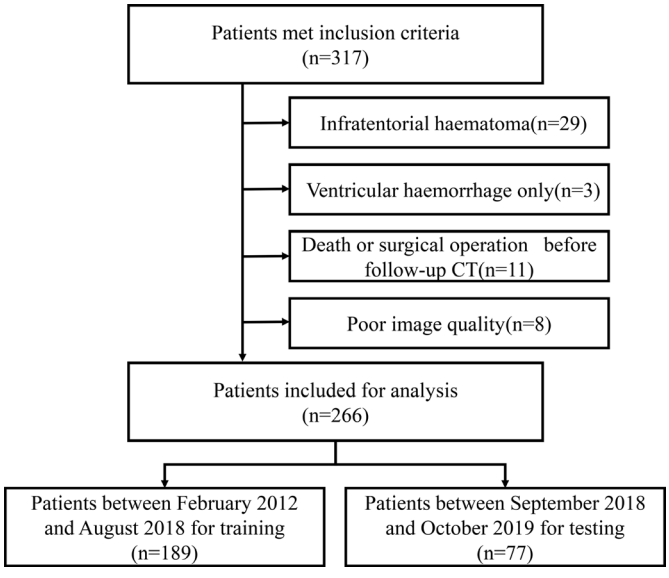


Figure 2 Flow chart illustrating patient selection for training dataset and testing dataset.

Table 1 Patientcharacteristics grouped by training and testing datasets

Variable name (and type)	Training dataset	Testing dataset	P value*
Sample size (n)	189	77	
Age, years, mean±SD	62.2±13.4	63.3±12.0	0.521
Sex, male, n (%)	132 (69.8)	54 (70.1)	0.963
Hypertension, n (%)	140 (74.9)	62 (80.5)	0.325
Diabetes mellitus, n (%)	19 (10.2)	17 (22.1)	0.010
Prestroke, n (%)	13 (6.9)	11 (14.3)	0.056
Antiplatelet history, n (%)	13 (6.9)	7 (9.1)	0.535
Anticoagulation history, n	2†	0	N/A
Time to baseline CT, hours, mean±SD	3.4±2.0	3.9±2.2	0.096
Baseline NIHSS score, median (IQR)	9 (5–12)	8 (4–14)	0.718
Baseline haematoma volume, mL, mean±SD	17.4±15.3	17.8±18.8	0.850
Intraventricular haemorrhage, n (%)	70 (37.0)	31 (40.3)	0.623
NCCT markers			
Hypodensities	132 (69.8)	44 (57.1)	0.047
Black hole sign	28 (14.8)	12 (15.6)	0.873
Swirl sign	116 (61.4)	44 (57.1)	0.523
Blend sign	24 (12.7)	13 (16.9)	0.371
Fluid level	13 (6.9)	9 (11.7)	0.196
Irregular shape	143 (75.7)	50 (64.9)	0.075
Heterogeneous density	100 (52.9)	34 (44.2)	0.195
BAT score, median (IQR)	2 (2–4)	2 (0–3)	0.065
Haematoma expansion, n (%)	52 (27.5)	22 (28.6)	0.861

*Continuous variables were compared using Mann-Whitney U test and Student's t-test as appropriate and categorical variables were compared using Pearson's χ^2 test.
†Two patients with warfarin history for atrial fibrillation had no haematoma expansion.
NCCT, non-contrast CT; NIHSS, National Institute of Health Stroke Scale.

RESULTS

A total of 317 patients fulfilled our inclusion criteria. Fifty-one patients were excluded due to the following reasons, infratentorial haematoma (n=29), only ventricular haemorrhage (n=3), died or had surgical operation before follow-up CT (n=11), or with poor image quality (n=8). A total of 266 patients were included for analysis, and divided into training dataset (n=189) and testing dataset (n=77) according to the admission date (figure 2). Baseline patient characteristics were compared between the training and testing dataset in table 1.

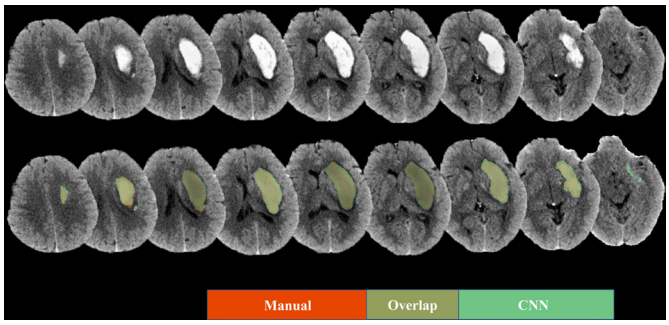


Figure 3 An illustrative case of the segmentation result: the haematoma segmented by the convolutional neural networks (CNN) model was in green, and the segmentation in the manual method was in red.

The Dice coefficient of the CNN model was 0.96±0.01 with the training dataset and 0.87±0.15 with the testing dataset. Representative images are shown in figure 3.

Of the 266 patients, 74 (27.8%) patients had haematoma expansion. Patients of training dataset and testing dataset were not different in the haematoma volume (p=0.850) and the BAT score (p=0.065). The sensitivity, specificity, likelihood ratio and AUCs of the NCCT markers and the three models were shown in table 2, and the ROC curves were shown in online supplemental figure 2.

For the NCCT markers, the CNN model exhibited highest AUC: 0.80 (95% CI 0.70 to 0.90) compared with 0.64 (95% CI 0.53 to 0.75), 0.65 (95% CI 0.54 to 0.75), 0.51 (95% CI 0.41 to 0.61), 0.58 (95% CI 0.48 to 0.67) and 0.55 (95% CI 0.44 to 0.67) for hypodensities, black hole sign, blend sign, fluid level and irregular shape, while the AUCs of swirl sign (0.70 (95% CI 0.61 to 0.80); p=0.211) and heterogenous density (0.70 (95% CI 0.59 to 0.81); p=0.141) were not significantly higher than that of the deep learning model. For the three models, the CNN model superior predictive accuracy than the haematoma volume (AUC 0.62 (95% CI 0.46 to 0.78); p=0.042) and the BAT score (0.65 (95% CI 0.53 to 0.78); p=0.042) according logistic regression models. In addition, the CNN model represented the lowest negative likelihood ratio (0.06 (95% CI 0.02 to 0.24)).

DISCUSSION

In this study, we developed a CNN-derived predictive model based on topological and morphological imaging features on NCCT to predict early haematoma expansion in ICH patients. According to the sensitivity, specificity, positive likelihood ratio, negative likelihood ratio and C statistics, comparing with other existing prediction models or NCCT markers, CNN model exhibited superior prediction efficacy.

Most recently, several NCCT markers were proposed for predicting early haematoma expansion in ICH patients, including blend sign, swirl sign and black hole sign.[7 8 10]

Table 2 Scores for models and NCCT markers of testing dataset

	Sensitivity	Specificity	Positive likelihood ratio*	Negative likelihood ratio*	AUC	P value†
Hypodensities	0.77 (0.54 to 0.91)	0.51 (0.37 to 0.64)	0.63 (0.40 to 0.98)	0.18 (0.08 to 0.40)	0.64 (0.53 to 0.75)	0.026
Black hole sign	0.36 (0.18 to 0.59)	0.93 (0.82 to 0.98)	2.00 (0.82 to 4.89)	0.27 (0.17 to 0.44)	0.65 (0.54 to 0.75)	0.006
Swirl sign	0.86 (0.64 to 0.96)	0.54 (0.41 to 0.68)	0.76 (0.50 to 1.16)	0.1 (0.03 to 0.30)	0.70 (0.61 to 0.80)	0.211
Blend sign	0.18 (0.06 to 0.41)	0.84 (0.71 to 0.92)	0.44 (0.18 to 1.08)	0.39 (0.26 to 0.59)	0.51 (0.41 to 0.61)	<0.001
Fluid level	0.23 (0.09 to 0.46)	0.93 (0.82 to 0.98)	1.25 (0.49 to 3.19)	0.33 (0.22 to 0.51)	0.58 (0.48 to 0.67)	0.002
Irregular shape	0.73 (0.50 to 0.88)	0.38 (0.261 to 0.52)	0.47 (0.30 to 0.74)	0.29 (0.14 to 0.59)	0.55 (0.44 to 0.67)	0.002
Heterogeneous density	0.73 (0.50 to 0.88)	0.67 (0.53 to 0.79)	0.89 (0.55 to 1.43)	0.16 (0.08 to 0.34)	0.70 (0.59 to 0.81)	0.141
Haematoma Volume	0.50 (0.29to 0.71)	0.84 (0.71 to 0.92)	1.22 (0.65 to 2.29)	0.24 (0.14 to 0.41)	0.62 (0.46 to 0.78)	0.042
BAT score	0.36 (0.18to 0.59)	0.76 (0.62 to 0.86)	0.61 (0.32 to 1.17)	0.33 (0.21 to 0.53)	0.65 (0.53 to 0.78)	0.042
CNN	0.91 (0.69to 0.98)	0.58 (0.44 to 0.71)	0.87 (0.57 to 1.33)	0.06 (0.02 to 0.24)	0.80 (0.70 to 0.90)	N/A

*Positive likelihood ratio and negative likelihood ratio were weighted by prevalence.
†AUC of CNN model was compared with AUC of the other models and NCCT signs using Delong test.
AUC, receiver operator characteristic area under the curve; CNN, convolutional neural network; N/A, not applicable; NCCT, non-contrast CT.

However, most of the NCCT markers were with relatively low sensitivity and low incidence.[11 12] To improve the performance of NCCT markers, Morotti *et al*[15] reported the BAT score based on the NCCT signs, and shown that their BAT score had a C-statistic of 0.65–0.70 for validation cohorts, which was in accordance with results of this study. Size of their dataset was sufficiently large, whereas, the comparison between the BAT score and the conventional NCCT markers was not performed. In addition, correlations between the initial haematoma volume and early haematoma expansion have been reported.[16] Therefore, in the present study, the logistic model only based on haematoma volume was built and its discriminative ability was similar to the BAT score.

Because of the excellent performance of the artificial intelligence technology in clinical, Liu *et al*[17] proposed a support vector machine (SVM) to predict early haematoma growth with an external validation AUC of 0.85. However, in this model, not only the NCCT markers but a bundle of clinical information was required as the input of SVM. By Only the initial NCCT images can be more convenient in clinical practice, while no previous study applied the deep learning technology in determining haematoma expansion. Therefore, we first developed a two-output CNN model merely based on the baseline NCCT image, and its performance was superior to the conventional NCCT markers and the BAT score.

There are several limitations in our study. First, data applied was from a single centre, though collected prospectively. Second, the test data set was also from our centre, the extrapolation of the model is thus limited due to the lack of external testifying from other centres. Thus, we should be cautious when the model is applied to other cohort with different clinical characteristics, and more external validation is needed. However, we divided the data into training and testing dataset according to the admission time of patients, to distinguish the training dataset from testing dataset in a longitudinal data, comparable with a trained model testifying by another prospective cohort. Third, since the data were collected from a single centre, only the patients from neighbouring regions can be sent to this centre in a super-early time after ICH ictus, thus the cohort may not represent the general ICH patients. Moreover, as an artificial intelligence technology, deep learning requires strong support from hardware and software, thus the application of this method for the rural hospitals may be limited.

In conclusion, our study developed an advanced prediction model using deep learning to predict early haematoma expansion in ICH patients. This CNN model exhibited a superior predicting ability compared with other prediction models aforementioned, therefore provides a more accurate method for predicting early haematoma expansion.

Correction notice This article has been corrected since it was published Online First. The funding statement has been updated with the specific number of source of funding.

Contributors Concept and design: L-sT, J-wZ. Acquisition, analysis or interpretation of data: J-wZ, Y-jJ, Z-jS, BL, X-hL. Drafting of the manuscript: L-sT, J-wZ, Yj-J, Z-jS. Critical revision of the manuscript for important intellectual content: L-sT, FC. Statistical analysis: J-wZ, Y-jJ. Obtained funding: L-sT. Administrative, technical or material support: L-sT, FC, BL, X-hL. Supervision: L-sT, FC. J-wZ and Y-jJ contributed equally to the manuscript; L-sT and FC are both corresponding authors.

Funding This study was supported by the National Natural Science Foundation of China (NSFC 81971155).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by the Institutional Review Board of The Second Affiliated Hospital of Zhejiang University.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The data that support the findings of the study are available from the corresponding author upon reasonable request.

ORCID iD

Jia-wei Zhong http://orcid.org/0000-0002-6175-7707

REFERENCES

1 Qureshi AI, Tuhrim S, Broderick JP, *et al*. Spontaneous intracerebral hemorrhage. *N Engl J Med* 2001;344:1450–60.
2 Dowlatshahi D, Demchuk AM, Flaherty ML, *et al*. Defining hematoma expansion in intracerebral hemorrhage: relationship with patient outcomes. *Neurology* 2011;76:1238–44.
3 Wada R, Aviv RI, Fox AJ, *et al*. CT angiography "spot sign" predicts hematoma expansion in acute intracerebral hemorrhage. *Stroke* 2007;38:1257–62.
4 Demchuk AM, Dowlatshahi D, Rodriguez-Luna D, *et al*. Prediction of haematoma growth and outcome in patients with intracerebral haemorrhage using the CT-angiography spot sign (predict): a prospective observational study. *Lancet Neurol* 2012;11:307–14.
5 Orito K, Hirohata M, Nakamura Y, *et al*. Leakage sign for primary intracerebral hemorrhage: a novel predictor of hematoma growth. *Stroke* 2016;47:958–63.
6 Morotti A, Brouwers HB, Romero JM, *et al*. Intensive blood pressure reduction and spot sign in intracerebral hemorrhage: a secondary analysis of a randomized clinical trial. *JAMA Neurol* 2017;74:950–60.
7 Li Q, Zhang G, Xiong X, *et al*. Black hole sign: novel imaging marker that predicts hematoma growth in patients with intracerebral hemorrhage. *Stroke* 2016;47:1777–81.
8 Selariu E, Zia E, Brizzi M, *et al*. Swirl sign in intracerebral haemorrhage: definition, prevalence, reliability and prognostic value. *BMC Neurol* 2012;12:109.
9 Li Q, Liu Q-J, Yang W-S, *et al*. Island sign: an imaging predictor for early hematoma expansion and poor outcome in patients with intracerebral hemorrhage. *Stroke* 2017;48:3019–25.
10 Li Q, Zhang G, Huang Y-J, *et al*. Blend sign on computed tomography: novel and reliable predictor for early hematoma growth in patients with intracerebral hemorrhage. *Stroke* 2015;46:2119–23.
11 Morotti A, Boulouis G, Romero JM, *et al*. Blood pressure reduction and noncontrast CT markers of intracerebral hemorrhage expansion. *Neurology* 2017;89:548–54.
12 Yogendrakumar V, Moores M, Sikora L, *et al*. Evaluating hematoma expansion scores in acute spontaneous intracerebral hemorrhage: a systematic scoping review. *Stroke* 2020;51:1305–8.
13 Cichy RM, Kaiser D. Deep neural networks as scientific models. *Trends Cogn Sci* 2019;23:305–17.
14 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Lect Notes Comput Sc* 2015;9351:234–41.
15 Morotti A, Dowlatshahi D, Boulouis G, *et al*. Predicting intracerebral hemorrhage expansion with noncontrast computed tomography: the bat score. *Stroke* 2018;49:1163–9.
16 Li Q, Huang Y-J, Zhang G, *et al*. Intraventricular hemorrhage and early hematoma expansion in patients with intracerebral hemorrhage. *Sci Rep* 2015;5:11357.

17 Liu J, Xu H, Chen Q, *et al*. Prediction of hematoma expansion in spontaneous intracerebral hemorrhage using support vector machine. *EBioMedicine* 2019;43:454–9.

SUPPLEMENTAL MATERIAL

Deep Learning for Automatically Predicting Early Hematoma Expansion in Chinese Patients

Image acquisition

The baseline CT images from our study were obtained with four CT scanners: GE Optima CT540, SIEMENS SOMATOM Definition Flash, SIEMENS SOMATOM Force, SIEMENS SOMATOM Perspective. The acquisition parameters were as follows: slice thickness, 5.0mm; axial slice number, 27-34; voxel size, 0.3906-0.5566×0.3906-0.5566×5.0mm; matrix size, 512×512; field of view, 200-285×200-285mm; window width, 90HU; window level, 35HU (**Supplemental Table1**).

Labeling and calculation of the hematoma volume

Manual segmentations for hematoma were performed on the CT scans by a single author with more than 10 years of experience. To assess interrater reliability, repeat manual segmentations in 20 randomly selected cases were performed after a minimal interval of 7 days. Labels were manually painted on each 2-dimensional slice of each CT image applying the open-source software ITK-SNAP[1] (http://www.itksnap.org/). Comparing to the contralateral hemisphere, intracerebral hemorrhage was differentiated from intraventricular hemorrhage or subarachnoid hemorrhage. Based on the binary label map, the hematoma volume was then calculated.

Data preprocessing

The CT images of our study were collected as a Digital Imaging and Communication in Medicine (DICOM) image series, and then were transformed to Neuroimaging Informatics Technology Initiative format. Each CT image was skull skipped by using Otsu`s method.[2] The CT images were resampling to a field of view of 112 × 112 × 160 mm and matrix size of 256 × 256 × 32 by applying a bicubic interpolation algorithm, and then the images were windowed with a threshold of 0 to 100 HU. After this, normalization was performed by subtracting the mean value within the skull-stripped brain region and dividing by the standard deviation of the signal intensity of the region, and negative values were set to zero.

Data augmentation

Data augmentation were performed by applying 3-dimensional image transformation with scaling, translation, and rotation. For the training dataset, each CT image was randomly transformed by applying these three transformations using a linear interpolation algorithm. The scaling and translation were performed between –10% and 10% of the image size and the rotation between –5 and 5 degrees. The size of the training dataset was increased 10 times with these techniques.

Model architecture

In this study, we build a two-task model based on a deep convolution neural network. The semantic segmentation of hematoma and the prediction of hematoma expansion were simultaneously working. Hematoma expansion outcome was binarized according to the volume of 24h follow-up CT image compared to the baseline CT image(⩾6 mL or ⩾33%). Detail of the model architecture is shown in **Supplemental Figure 1**.

Segmentation network

The segmentation network was based on U-Net[3], an encoder-decoder network for 3-dimentional image segmentation. Our network for segmentation had 4-level architecture with 2 down-sampling, 2 up-sampling and 1 convolution operation for bridging layer(Supplemental Figure 2): 32 × 256 × 256 (16 channels) → 16 × 128 × 128 (32 channels) → 8 × 64 × 64 (64 channels) → 8 × 64 × 64 (128 channels) → 16 × 128 × 128 (64 channels) → 32 × 256 × 256 (32 channels). The down-sampling of encoding path was performed by a 3 × 3 × 3 three-dimensional convolution layer with 2 × 2 × 2 strides, and the up-sampling of decoding path was performed with a size of 2 × 2 × 2. Padding was applied in steps above. The final output of segmentation was performed by applying 1 × 1 × 1 convolution with sigmoid function, reducing the channel number to one.

Classification network

The classification network was added to the bridging layer of the segmentation network (**Supplemental Figure 1**). Two 3 × 3 × 3(with 128 filters) three-dimensional convolution operations with padding were performed and then feature channels(size 128 × 8 × 64 × 64) were processed with three-dimensional global average pooling to be 128 units. The flattened features were connected to a unit of output applying the Sigmoid activation function for the binary classification work.

Training process

The deep neural network model had 3,802,578 parameters. The Adam optimizer was applied with back propagation. The loss function for segmentation was Dice loss (1 – Dice coefficient), and the loss function for classification was binary cross-entropy. The segmentation and classification were simultaneously trained with the cost function in the ratio 9:1 (segmentation: classification). The network was trained for 40 epochs totally with a batch size of 1. A multi-step learning rate schedule was performed with an initial learning rate of 1×10^{-4}, reduced by a factor of 10 at 20th and 30th epoch.

The training process was performed with an 11-GB graphics processing unit (NVIDIA GeForce RTX 2080Ti). Training process took about 14 hours. The code of convolution neural network was written in Python 3.7(https://www.python.org/) and implemented in open-source deep learning framework Pytorch 1.4.0(https://pytorch.org/).

Model evaluation

For the segmentation task, the Dice coefficient was calculated to evaluate the segmentation results.

For the classification task, the sensitivity, specificity, likelihood ratio weighted by prevalence, and area under the curve (AUC) were calculated based on receiver operating characteristic (ROC) curves (**Supplemental Figure 2**).

Figures and Tables

Supplemental Table1 Acquisition parameters for CT scanners

Scanner	GE Optima CT540	SIEMENS SOMATOM Definition Flash	SIEMENS SOMATOM Force	SIEMENS SOMATOM Perspective
Slice thickness, mm	5.0	5.0	5.0	5.0
Axial slice number	28~34	26~28	27	27
Voxel size, mm	0.4883×0.4883×5.0	0.3906~0.5566×0.3906~0.5566×5.0	0.3906~0.4453×0.3906~0.4453×5.0	0.4492×0.4492×5.0
Matrix size	512×512	512×512	512×512	512×512
Field of view, mm	250×250	200~285×200~285	200~228×200~228	230×230
Window width	90	90	90	90
Window level	35	35	35	35

CT indicates computed tomography.

Supplemental Table2 Diagnostic criteria for NCCT markers[4]

Marker	Criteria
Hypodensity	Any hypodense region strictly encapsulated within the hemorrhage with any shape, size, and density.
Black hole sign	Hypoattenuating area with a density difference >28HU compared with the surrounding hematoma. No connection with surface outside the hematoma.
Swirl sign	Rounded, streak-like, or irregular region of hypo- or isoattenuation compared with the brain parenchyma. Does not have to be encapsulated in the ICH.
Blend sign	Relatively hypoattenuating area next to a hyperattenuating area of the hematoma, with a well-defined margin and a density difference >18HU between the two areas.
Fluid level	Presence of an area hypodense to the brain above and one hyperattenuating area below a discrete straight line of separation, irrespective of its density appearance.
Irregular shape	Two or more focal hematoma margin irregularities, joined or separate from the hematoma edge on the axial NCCT slice with largest hematoma area.

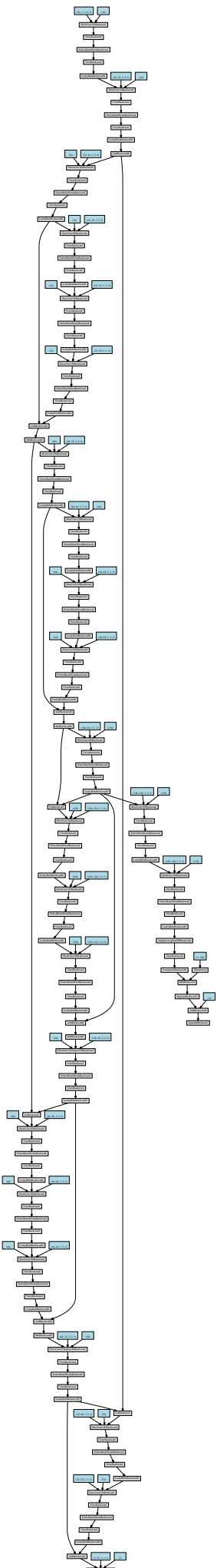
NCCT indicates non-contrast computed tomography.

Supplemental Table3 Criteria for the BAT score[5]

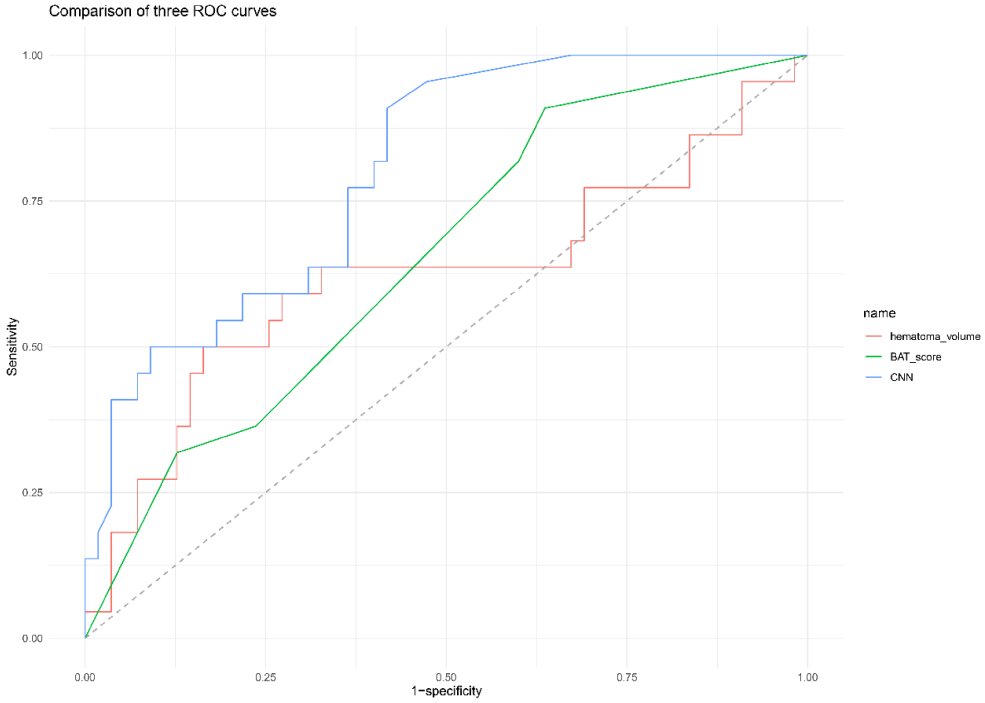
	Points
Blend sign	
Present	1
Absent	0
Any hypodensity	
Present	2
Absent	0
Time from onset to NCCT	
<2.5h	2
≥2.5h	0

NCCT indicates non-contrast computed tomography.

Supplemental Figure 1 The architecture of the model



Supplemental Figure 2 ROC curves of testing



The receiver operating characteristic curves of the CNN model, the hematoma volume model and the BAT score model were displayed.

References

1. Yushkevich PA, Piven J, Hazlett HC, *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-28.
2. Otsu N. Threshold Selection Method from Gray-Level Histograms. *Ieee T Syst Man Cyb*. 1979;9(1):62-6.
3. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect Notes Comput Sc*. 2015;9351:234-41.
4. Morotti A, Boulouis G, Dowlatshahi D, *et al.* Standards for Detecting, Interpreting, and Reporting Noncontrast Computed Tomographic Markers of Intracerebral Hemorrhage Expansion. *Ann Neurol*. 2019;86(4):480-92.
5. Morotti A, Dowlatshahi D, Boulouis G, *et al.* Predicting Intracerebral Hemorrhage Expansion With Noncontrast Computed Tomography: The BAT Score. *Stroke*. 2018;49(5):1163-9.

SUPPLEMENTAL MATERIAL

Deep Learning for Automatically Predicting Early Hematoma Expansion in Chinese Patients

Image acquisition

The baseline CT images from our study were obtained with four CT scanners: GE Optima CT540, SIEMENS SOMATOM Definition Flash, SIEMENS SOMATOM Force, SIEMENS SOMATOM Perspective. The acquisition parameters were as follows: slice thickness, 5.0mm; axial slice number, 27-34; voxel size, 0.3906-0.5566×0.3906-0.5566×5.0mm; matrix size, 512×512; field of view, 200-285×200-285mm; window width, 90HU; window level, 35HU (**Supplemental Table1**).

Labeling and calculation of the hematoma volume

Manual segmentations for hematoma were performed on the CT scans by a single author with more than 10 years of experience. To assess interrater reliability, repeat manual segmentations in 20 randomly selected cases were performed after a minimal interval of 7 days. Labels were manually painted on each 2-dimensional slice of each CT image applying the open-source software ITK-SNAP[1] (http://www.itksnap.org/). Comparing to the contralateral hemisphere, intracerebral hemorrhage was differentiated from intraventricular hemorrhage or subarachnoid hemorrhage. Based on the binary label map, the hematoma volume was then calculated.

Data preprocessing

The CT images of our study were collected as a Digital Imaging and Communication in Medicine (DICOM) image series, and then were transformed to Neuroimaging Informatics Technology Initiative format. Each CT image was skull skipped by using Otsu`s method.[2] The CT images were resampling to a field of view of 112 × 112 × 160 mm and matrix size of 256 × 256 × 32 by applying a bicubic interpolation algorithm, and then the images were windowed with a threshold of 0 to 100 HU. After this, normalization was performed by subtracting the mean value within the skull-stripped brain region and dividing by the standard deviation of the signal intensity of the region, and negative values were set to zero.

Data augmentation

Data augmentation were performed by applying 3-dimensional image transformation with scaling, translation, and rotation. For the training dataset, each CT image was randomly transformed by applying these three transformations using a linear interpolation algorithm. The scaling and translation were performed between –10% and 10% of the image size and the rotation between –5 and 5 degrees. The size of the training dataset was increased 10 times with these techniques.

Model architecture

In this study, we build a two-task model based on a deep convolution neural network. The semantic segmentation of hematoma and the prediction of hematoma expansion were simultaneously working. Hematoma expansion outcome was binarized according to the volume of 24h follow-up CT image compared to the baseline CT image($\geqslant$6 mL or $\geqslant$33%). Detail of the model architecture is shown in **Supplemental Figure 1**.

Segmentation network

The segmentation network was based on U-Net[3], an encoder-decoder network for 3-dimentional image segmentation. Our network for segmentation had 4-level architecture with 2 down-sampling, 2 up-sampling and 1 convolution operation for bridging layer(Supplemental Figure 2): 32 × 256 × 256 (16 channels) → 16 × 128 × 128 (32 channels) → 8 × 64 × 64 (64 channels) → 8 × 64 × 64 (128 channels) → 16 × 128 × 128 (64 channels) → 32 × 256 × 256 (32 channels). The down-sampling of encoding path was performed by a 3 × 3 × 3 three-dimensional convolution layer with 2 × 2 × 2 strides, and the up-sampling of decoding path was performed with a size of 2 × 2 × 2. Padding was applied in steps above. The final output of segmentation was performed by applying 1 × 1 × 1 convolution with sigmoid function, reducing the channel number to one.

Classification network

The classification network was added to the bridging layer of the segmentation network (**Supplemental Figure 1**). Two 3 × 3 × 3(with 128 filters) three-dimensional convolution operations with padding were performed and then feature channels(size 128 × 8 × 64 × 64) were processed with three-dimensional global average pooling to be 128 units. The flattened features were connected to a unit of output applying the Sigmoid activation function for the binary classification work.

Training process

The deep neural network model had 3,802,578 parameters. The Adam optimizer was applied with back propagation. The loss function for segmentation was Dice loss (1 – Dice coefficient), and the loss function for classification was binary cross-entropy. The segmentation and classification were simultaneously trained with the cost function in the ratio 9:1 (segmentation: classification). The network was trained for 40 epochs totally with a batch size of 1. A multi-step learning rate schedule was performed with an initial learning rate of 1×10^{-4}, reduced by a factor of 10 at 20th and 30th epoch.

The training process was performed with an 11-GB graphics processing unit (NVIDIA GeForce RTX 2080Ti). Training process took about 14 hours. The code of convolution neural network was written in Python 3.7(https://www.python.org/) and implemented in open-source deep learning framework Pytorch 1.4.0(https://pytorch.org/).

Model evaluation

For the segmentation task, the Dice coefficient was calculated to evaluate the segmentation results.

For the classification task, the sensitivity, specificity, likelihood ratio weighted by prevalence, and area under the curve (AUC) were calculated based on receiver operating characteristic (ROC) curves (**Supplemental Figure 2**).

Figures and Tables

Supplemental Table1 Acquisition parameters for CT scanners

Scanner	GE Optima CT540	SIEMENS SOMATOM Definition Flash	SIEMENS SOMATOM Force	SIEMENS SOMATOM Perspective
Slice thickness, mm	5.0	5.0	5.0	5.0
Axial slice number	28~34	26~28	27	27
Voxel size, mm	0.4883×0.4883×5.0	0.3906~0.5566×0.3906~0.5566×5.0	0.3906~0.4453×0.3906~0.4453×5.0	0.4492×0.4492×5.0
Matrix size	512×512	512×512	512×512	512×512
Field of view, mm	250×250	200~285×200~285	200~228×200~228	230×230
Window width	90	90	90	90
Window level	35	35	35	35

CT indicates computed tomography.

Supplemental Table2 Diagnostic criteria for NCCT markers[4]

Marker	Criteria
Hypodensity	Any hypodense region strictly encapsulated within the hemorrhage with any shape, size, and density.
Black hole sign	Hypoattenuating area with a density difference >28HU compared with the surrounding hematoma. No connection with surface outside the hematoma.
Swirl sign	Rounded, streak-like, or irregular region of hypo- or isoattenuation compared with the brain parenchyma. Does not have to be encapsulated in the ICH.
Blend sign	Relatively hypoattenuating area next to a hyperattenuating area of the hematoma, with a well-defined margin and a density difference >18HU between the two areas.
Fluid level	Presence of an area hypodense to the brain above and one hyperattenuating area below a discrete straight line of separation, irrespective of its density appearance.
Irregular shape	Two or more focal hematoma margin irregularities, joined or separate from the hematoma edge on the axial NCCT slice with largest hematoma area.

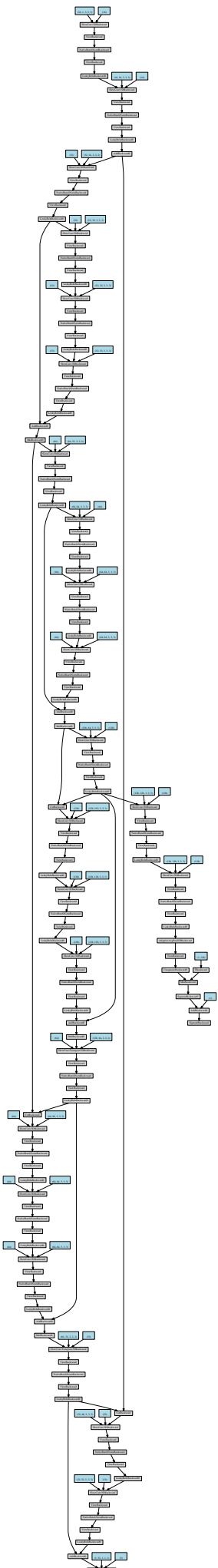
NCCT indicates non-contrast computed tomography.

Supplemental Table3 Criteria for the BAT score[5]

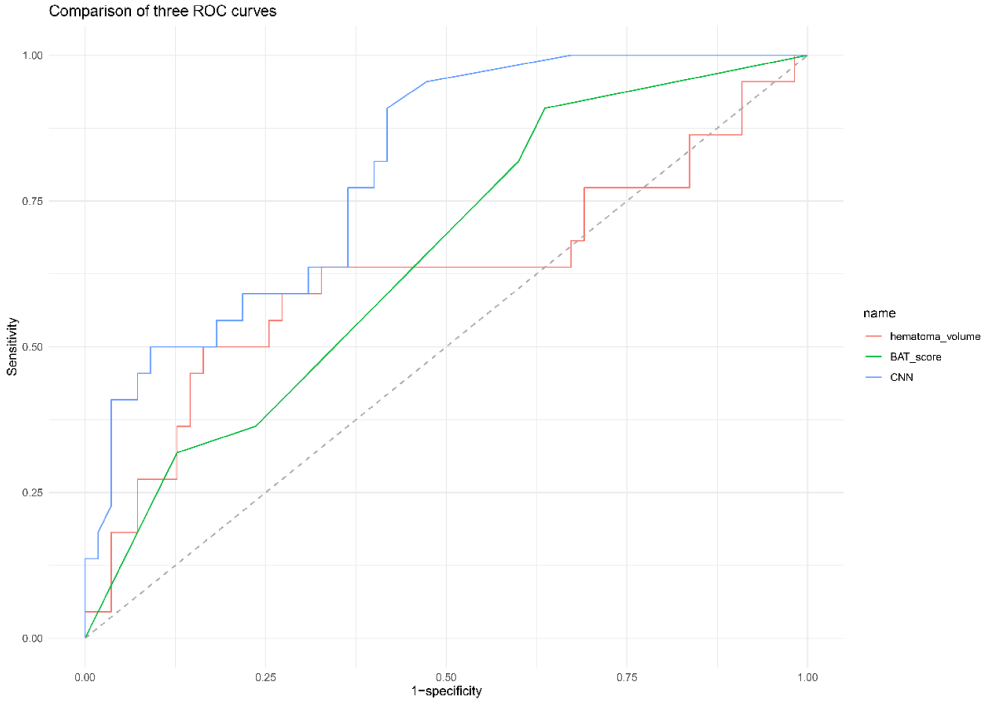
	Points
Blend sign	
Present	1
Absent	0
Any hypodensity	
Present	2
Absent	0
Time from onset to NCCT	
<2.5h	2
≥2.5h	0

NCCT indicates non-contrast computed tomography.

Supplemental Figure 1 The architecture of the model



Supplemental Figure 2 ROC curves of testing



The receiver operating characteristic curves of the CNN model, the hematoma volume model and the BAT score model were displayed.

References

1. Yushkevich PA, Piven J, Hazlett HC, *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-28.
2. Otsu N. Threshold Selection Method from Gray-Level Histograms. *Ieee T Syst Man Cyb*. 1979;9(1):62-6.
3. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect Notes Comput Sc*. 2015;9351:234-41.
4. Morotti A, Boulouis G, Dowlatshahi D, *et al.* Standards for Detecting, Interpreting, and Reporting Noncontrast Computed Tomographic Markers of Intracerebral Hemorrhage Expansion. *Ann Neurol*. 2019;86(4):480-92.
5. Morotti A, Dowlatshahi D, Boulouis G, *et al.* Predicting Intracerebral Hemorrhage Expansion With Noncontrast Computed Tomography: The BAT Score. *Stroke*. 2018;49(5):1163-9.