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 Tong5

ABSTRACT

Background and purpose Early haematoma expansion is deterministic 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-temed 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 heterogeneity, 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 >6mL 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 supplement 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 χ² 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).

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.

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.\(^7\)\(^8\)\(^9\)\(^10\) To improve the performance of NCCT markers, Morotti et al\(^10\) 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

![Image](https://example.com/image.png)

**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.

### Table 2

Scores for models and NCCT markers of testing dataset

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

*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.
and early haematoma expansion have been reported.\textsuperscript{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.\textsuperscript{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, Y-jj, 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.

REFERENCES