European Journal of Radiology 167 (2023) 111081

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

# An artificial intelligence-based prognostic prediction model for hemorrhagic stroke

Yihao Chen [a,1], Cheng Jiang [b,1], Jianbo Chang [a,1], Chenchen Qin [b], Qinghua Zhang [c], Zeju Ye [d], Zhaojian Li [e,g], Fengxuan Tian [f], Wenbin Ma [a], Ming Feng [a], Junji Wei [a,*], Jianhua Yao [b,*], Renzhi Wang [a,*]

[a] *Department of Neurosurgery, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China*
[b] *Tencent AI Lab, Shenzhen, China*
[c] *Department of Neurosurgery, Shenzhen Nanshan Hospital, Shen Zhen, China*
[d] *Department of Neurosurgery, Dongguan People's Hospital, Guangdong Province, China*
[e] *Department of Neurosurgery, The Affiliated Hospital of Qingdao University, China*
[f] *Department of Neurosurgery, Qinghai Provincial People's Hospital, Qinghai Province, China*
[g] *Department of Medicine, Qingdao University, Qingdao, China*

A R T I C L E  I N F O

*Keywords:*
Intracerebral hemorrhage
Deep learning
Prognosis
ICH scale
Computed tomography

A B S T R A C T

*Purpose:* The prognosis following a hemorrhagic stroke is usually extremely poor. Rating scales have been developed to predict the outcomes of patients with intracerebral hemorrhage (ICH). To date, however, the prognostic prediction models have not included the full range of relevant imaging features. We constructed a clinic-imaging fusion model based on convolutional neural networks (CNN) to predict the short-term prognosis of ICH patients.

*Materials and methods:* This was a multi-center retrospective study, which included 1990 patients with ICH. Two CNN-based deep learning models were constructed to predict the neurofunctional outcomes at discharge; these were validated using a nested 5-fold cross-validation approach. The models' predictive efficiency was compared with the original ICH scale and the ICH grading scale. Poor neurological outcome was defined as a Glasgow Outcome Scale (GOS) score of 1–3.

*Results:* The training and test sets included 1599 and 391 patients, respectively. For the test set, the clinic-imaging fusion model had the highest area under the curve (AUC = 0.903), followed by the imaging-based model (AUC = 0.886), the ICH scale (AUC = 0.777), and finally the ICH grading scale (AUC = 0.747).

*Conclusion:* The CNN prognostic prediction model based on neuroimaging features was more effective than the ICH scales in predicting the neurological outcomes of ICH patients at discharge. The CNN model's predictive efficiency slightly improved when clinical data were included.

## 1. Introduction

Strokes have become one of the most frequent causes of death worldwide, second only to coronary heart disease [1]. Among stroke patients, those with a hemorrhagic stroke have one of the worst prognoses, with an in-hospital mortality rate of up to 20–30 % [2,3] and with up to 80 % requiring long-term care [4,5]. At present, it is important to determine the prognosis of ICH using risk stratification models, so that

*Abbreviations:* ICH, intracerebral hemorrhage; CNN, convolutional neural network; GOS, Glasgow Outcome Scale; AUC, area under the curve; ICH-GS, ICH grading scale; ANN, artificial neural network; LR, logistic regression; EHR, electronic health records; SICH, spontaneous intracerebral hemorrhage; CICHID, Chinese Intracranial Hemorrhage Image Database; NCCT, non-contrast computed tomography; CRF, case report form; GCS, Glasgow Coma Scale; GMU, gated multimodal unit; CMT, conservative medical treatment; CC, conventional craniotomy; MIA, minimally invasive approach; ROC, receiver operating characteristic curve; CI, confidence interval; mRS, modified ranking scale.

* Corresponding authors at: Department of Neurosurgery, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100730, China (R. Wang and J. Wei). Building 12A 28th floor, Ecological Park, Nanshan District, Tencent AI Lab, Shenzhen 518000, China (J. Yao).
*E-mail addresses:* weijunji@pumch.cn (J. Wei), Jianhua.yao@gmail.com (J. Yao), wangrz@126.com (R. Wang).

[1] Yihao Chen, Cheng Jiang, and Jianbo Chang contributed equally to this work and should be considered co-first authors.

https://doi.org/10.1016/j.ejrad.2023.111081
Received 14 February 2022; Received in revised form 22 June 2023; Accepted 4 September 2023
Available online 9 September 2023
0720-048X/

relevant information can be given to family members and clinicians can be guided in treatment planning.

Of these risk stratification models, the ICH scale is one of the most commonly used [6]. It has been shown to be both effective and straightforward with many different cohorts [7,8]. Variants of the original ICH scale have been developed, including the modified ICH scale [9] and the ICH grading scale (ICH-GS) [10], both of which are widely used. However, the traditional grading scales have several flaws that need to be addressed: (1) The relative simplicity of the index can lead to an overestimation of the disease severity. This may lead to certain treatments being withdrawn or not attempted because of an erroneous assumption that they would be ineffective [11]. (2) The risk factor screening may have led to certain essential variables being discounted, which would result in inaccurate classification and prediction. (3) The scales are known vary in terms of their predictive accuracy for different racial groups and clinical measures[12,13]. (4) Traditional approaches have focused on quantifiable clinical features and have neglected potentially significant imaging features (e.g., midline shift, perihematomal edema, compressed basal cistern).

Several different approaches have been adopted to address these issues. For instance, Lukić et al. developed an artificial neural network (ANN)-based model to predict ICH prognoses. The model was found to be significantly more accurate than logistic regression (LR) analysis (ANN 93.55 % compared with LR 79.32 %) [14]. In another study, Wang et al. used machine learning to develop the random forest ICH prognostic model. The accuracy in predicting ICH patient outcomes was found to be 83.1 % and 83.9 % at 1 and 6 months, respectively [15]. Xu et al. adopted an approach combining CT radiomics with machine learning methods to develop ICH prognostic models that included 18 CT imaging features. Of these, the random forest model showed the best performance (accuracy 92.7 %) for predicting ICH outcomes at 6 months [16]. Overall, these new models have been able to refine the selection of risk factors and include more imaging features. However, the training data were mainly clinical textual data derived from electronic health records (EHR) rather than the data from imaging files [14,15,17]. In addition, the models' performance was not ideal, and in each study, the sample size was limited.

Deep learning has shown great promise for predicting prognoses using imaging data [18–20]. It has been shown that deep learning using a convolutional neural network (CNN) is superior to machine learning methods [21]. The CNN combines a series of linear and non-linear layers to learn the imaging features and make predictions. In this study, we developed a CNN-based predictive model that fuses clinical data and the full range of imaging features, in an attempt to address the shortcomings of previous ICH prognostic models. We compared the predictive efficiency of the model with the ICH scale and the ICH-GS scale, concerning the neurological outcomes of ICH patients at the time of discharge.

## 2. Materials and methods

### 2.1. Study population

Clinical and imaging data were collected from the Chinese Intracranial Hemorrhage Image Database (CICHID) for retrospective analysis. The data collection was restricted to patients with a spontaneous intracerebral hemorrhage (SICH). The CICHID program began in 2019, and it aimed to use big data and artificial intelligence techniques to construct a precise evaluation system for ICH patients. Our data came from nine different medical centers, and it was used for model derivation and validation. Strict inclusion and exclusion criteria were applied; any patients with missing data (e.g., low quality imaging files, incomplete medical records) were excluded from the analyses (Fig. 1).

The study received ethical approval from the Ethics Committee of PUMCH (Ethics code: S-K1175). As this was a retrospective study,

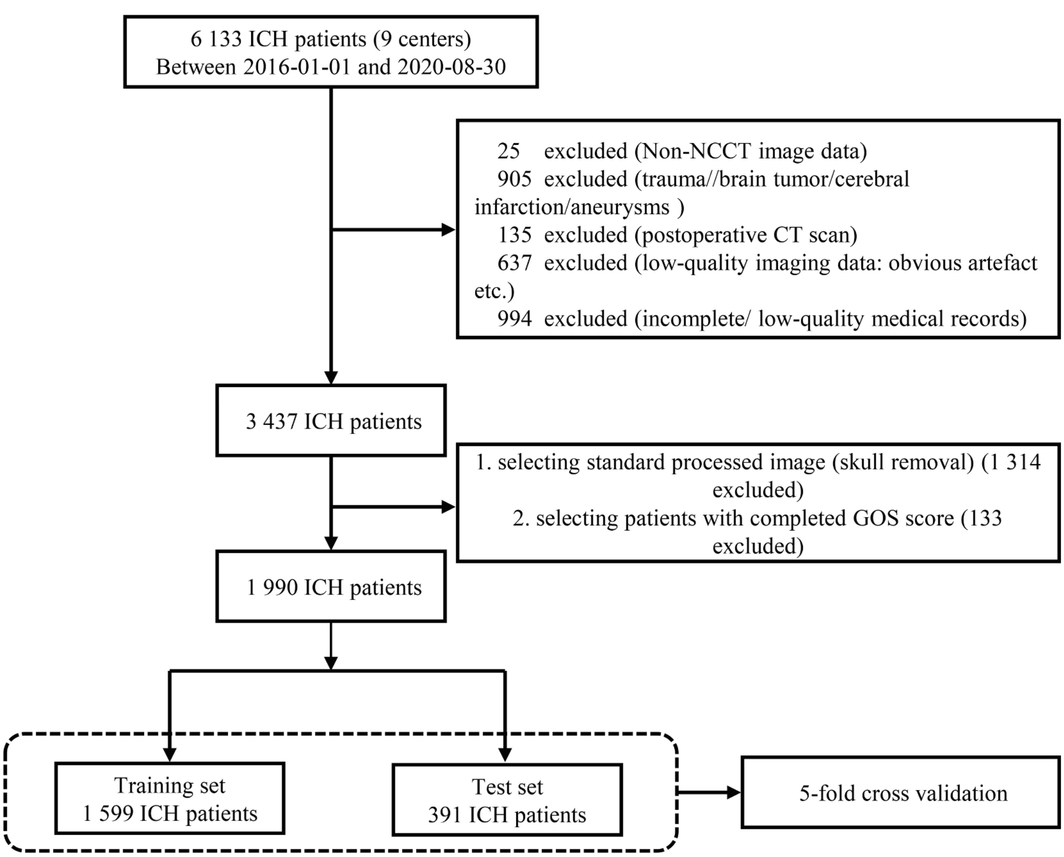


**Fig. 1.** Flowchart of the study population. ICH: intracerebral hemorrhage; NCCT: Non-contrast computed tomography.

requirements for written informed consent were waived.

### 2.2. Prediction target

We aimed to construct a CNN-based model that would predict the neuro-functional outcomes of ICH patients at the time of discharge. Neurological status was assessed using the Glasgow Outcome Scale (GOS); a GOS score of 1–3 was defined as a poor outcome [22,23]. The scores were determined by a neurosurgeon (Y.C., 5 years of experience) who had been specially trained to use the scale.

### 2.3. Demographic data and image processing

All patients received an initial non-contrast computed tomography (NCCT) scan during the first 24 h following admission to the hospital. The imaging and demographic data were fully anonymized prior to the analyses. The type of hemorrhage (intraventricular, subarachnoid, extradural, or subdural) was determined for each patient on the basis of a NCCT scan. The image below the foramen magnum was automatically clipped in preprocessing stage. Then these NCCT images with an average slice thickness of 5.0 mm were skull stripped. We set the window width to 90 HU and the window level to 45 HU. Data augmentation, random image cropping and random rotation were then performed. The image size was finally set at 192*192*80 pixels. The hematoma volumes were calculated for each patient, as this was required for the ICH/ICH-GS scores. To measure the volume, all of the CT slices that contained the hematoma were identified and delineated independently by one trained researcher and were subsequently manually checked by an experienced neuroimaging physician. These aforementioned-CT images with a standard DICOM format were carried out using "Insight Toolkit SNAP (ITK-SNAP)" software [24]. All delineated slices of each patient were integrated into a complete segmentation file, and then the hematoma volumes were automatically calculated by the ITK-SNAP software.

The demographic data were recorded independently by two researchers (neurosurgeon [Y.C., 5 years of experience; J.C., 6 years of experience]). A case report form (CRF) was designed and used for the clinical and demographic data collection. The consistency and completeness of the data were checked by another senior doctor (neurosurgeon [Z.Y., 10 years of experience]).

### 2.4. Assessment of the original ICH score and the ICH-GS score

An independent neurosurgeon (Y.C., 5 years of experience) determined the ICH score and the ICH-GS score for each patient. This involved calculating the total scores for the relevant items on each patient's CRF (e.g., age, Glasgow Coma Scale [GCS] score, ICH location, ICH volume, intraventricular extension of hemorrhage) [6,10]. Of note, the ICH volume was measured manually in mL (rounded to two decimal places).

### 2.5. The CNN model

For the visual feature extraction, we used DenseNet [25] because of its robustness and efficacy. We used its 3-D variant with a depth of 121. This architecture is known to make back-propagation easier because of the dense connections. With better use of visual features at different scales, it has been shown to have state-of-art performance for various image processing tasks. We chose to use a gated multimodal unit (GMU) [26] to combine the information from the CT images and the electronic health records. This unit can effectively merge the visual and textual information with an adaptive weight in their subspaces, which is more sophisticated than simple concatenation or a linear sum. The method used is shown in Fig. 2.

We performed 5-fold cross validation on the dataset. Four NVIDIA Tesla V100 GPUs with 32 GB memory were used together for training. We used Pytorch 1.5.1 for our experiments. The batch size was set at 32, and the learning rate was set at 1e-4. We adopted a total of 22 features from the EHR, including age, sex, GCS score, headache/emesis/coma symptoms, ICH extension to the ventricles, a history of hemorrhagic stroke/ischemic stroke/hypertension/diabetes mellitus/hyperlipidemia/coronary heart disease/heart failure/arrhythmia/anti-coagulant therapy/anti-platelet therapy/smoking/alcohol intake, systolic blood pressure, diastolic blood pressure, and treatment, including conservative medical treatment (CMT), conventional craniotomy (CC), and a minimally invasive approach (MIA; including minimally invasive puncture surgery and endoscopic surgery).

### 2.6. Statistical analyses

Data analyses were performed using SPSS 24.0 (IBM Corp., Armonk,

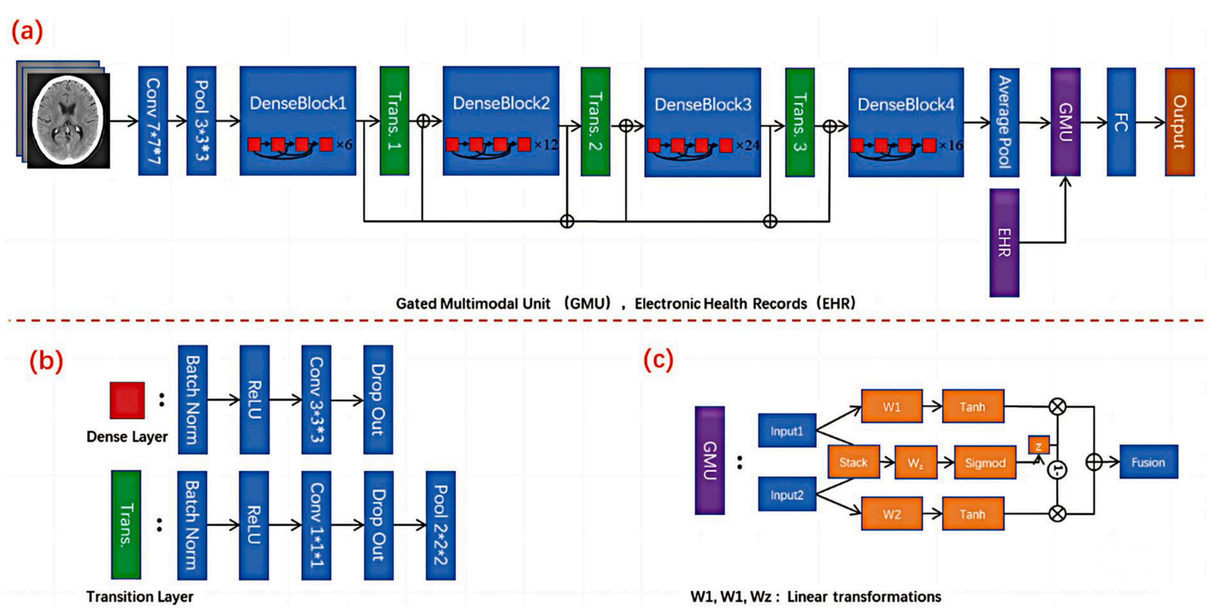


**Fig. 2.** The pipeline of clinic-imaging fusion-based deep learning. (a) is the main structure; (b) illustrates the Dense Layer and the Transition Layer in (a); and (c) is the detailed structure of the GMU block in (a).

NY, USA). The Kolmogorov-Smirnov test was used to test for normality. Normally distributed data were summarized as the mean (±standard deviation) and compared between groups using a Student's *t*-test or a one-way ANOVA. Data that were not normally distributed were summarized as the median (and interquartile range [IQR]) and compared between groups using a Mann–Whitney *U* test or Kruskal-Wallis H test. Categorical data were summarized as the number of cases, the rate, or the constituent ratio (%), and comparisons used the chi-square test. Factors that significantly predicted poor outcomes in univariate analyses were then included in a multivariate logistics regression analysis. A two-tailed $P < 0.05$ was considered statistically significant. A receiver operating characteristic (ROC) curve analysis was used to evaluate the predictive performance of the prognostic models (the two CNN models and the two traditional ranking scales), and the area under the curve (AUC), sensitivity, specificity, and 95 % confidence intervals (CI) were reported. The ROC was performed using GraphPad Prism software (version 9.0). Multiple ROC curves were compared using the DeLong method [27].

## 3. Results

### 3.1. Baseline characteristics

The study was conducted on data from patients with a spontaneous intracerebral hemorrhage. A total of 3437 ICH patients were identified as candidates for inclusion. A screening procedure was carried out, as described in Fig. 1, which led to a total of 1990 SICH patients being retained for the analyses. The patients were randomly allocated to either the training set (n = 1599) or the test set (n = 391). The baseline demographic data, imaging data, and clinical outcomes are presented in Table 1.

### 3.2. Multiple logistic analysis

Univariate analyses showed that the two groups (favorable outcome *vs* poor outcome) differed significantly in age, GCS/ICH/ICH-GS score on admission, systolic blood pressure, history of ischemic stroke/hemorrhagic stroke/anti-platelet therapy/alcohol intake, timing of the initial CT scan, initial hematoma volume, ICH extension to the ventricles, and treatment (all $P < 0.05$, see Table 1). Multivariate regression analyses showed that a poor prognosis was associated with advanced age (OR: 1.042, 95 % CI: 1.030–1.055, $P < 0.01$), a lower GCS score on admission (OR: 0.682, 95 % CI: 0.639–0.727, $P < 0.01$), a higher ICH-GS score on admission (OR: 0.718, 95 % CI: 0.608–0.848, $P < 0.01$), higher systolic blood pressure on admission (OR: 1.005, 95 % CI: 1.001–1.010, $P < 0.01$), a history of hemorrhagic stroke (OR: 1.909, 95 % CI: 1.209–3.017, $P < 0.01$), a larger initial hematoma volume (OR: 1.018, 95 % CI: 1.011–1.026, $P < 0.01$), ICH extension to the ventricles (OR: 1.425, 95 % CI: 1.006–2.018, $P = 0.046$), and surgical approaches involving conventional craniotomy (OR: 0.312, 95 % CI: 0.231–0.420, $P < 0.01$), as shown in Table 2.

### 3.3. Comparison between the models using the test data

The traditional grading scales and the new deep learning-based models all effectively predicted the prognosis of the ICH patients at discharge. The ROC curves for each model are shown in Fig. 3. The clinic-imaging fusion CNN model had the best performance (AUC = 0.903 ± 0.015, $P < 0.01$), followed by the imaging-based CNN model (AUC = 0.886 ± 0.016, $P < 0.01$), the ICH scale (AUC = 0.777 ± 0.022, $P < 0.01$), and finally the ICH-GS scale (AUC = 0.747 ± 0.024, $P < 0.01$; see Table 3). There was a statistically significant difference between the predictive ability of the CNN models and the two grading scales ($P < 0.01$); there was also a difference between the predictive ability of the clinic-imaging fusion CNN model and the imaging-based CNN model ($P = 0.027$).

**Table 1**
Patient Characteristics and Clinical Outcomes.

| | All patients n = 1990 | Favorable outcome (GOS 4-5) n = 712 | Poor outcome (GOS 1-3) n = 1 278 | *P* value |
|---|---|---|---|---|
| Age, years | 60(50–69) | 56(48–66) | 62(50–70) | <0.01* |
| Male sex, n (%) | 1289(64.77 %) | 481(67.55 %) | 808(63.22 %) | 0.052 |
| GCS on admission | 13(8–14) | 14(13–15) | 11(7–14) | <0.01* |
| ICH score on admission | 1(0–2) | 1(0–1) | 2(1–3) | <0.01* |
| ICH-GS score on admission | 7(6–9) | 7(6–7) | 8(7–9) | <0.01* |
| Comorbidities, n (%) | | | | |
| History of hemorrhagic stroke | 151(7.59 %) | 35(4.92 %) | 116(9.08 %) | <0.01* |
| History of ischemic stroke | 194(9.75 %) | 49(6.88 %) | 145(11.35 %) | <0.01* |
| Hypertension | 1345(67.59 %) | 469(65.87 %) | 876(68.54 %) | 0.222 |
| Diabetes mellitus | 166(8.34 %) | 55(7.72 %) | 111(8.69 %) | 0.458 |
| Hyperlipidemia | 20(1.01 %) | 5(0.70 %) | 15(1.17 %) | 0.312 |
| Coronary heart disease | 122(6.13 %) | 40(5.62 %) | 82(6.42 %) | 0.477 |
| Chronic heart failure | 3(0.15 %) | 1(0.14 %) | 2(0.16 %) | 0.930 |
| Cardiac arrhythmia | 20(1.01 %) | 5(0.70 %) | 15(1.17 %) | 0.312 |
| Anti-coagulant therapy | 16(0.80 %) | 3(0.42 %) | 13(1.02 %) | 0.154 |
| Anti-platelet therapy | 147(7.39 %) | 37(5.20 %) | 110(8.61 %) | <0.01* |
| Smoking | 457(22.96 %) | 155(21.77 %) | 302(23.63 %) | 0.344 |
| Alcohol intake | 447(22.46 %) | 133(18.68 %) | 314(24.57 %) | <0.01* |
| Physical examination, mmHg | | | | |
| Systolic blood pressure | 171(155–190) | 167(150–185) | 175(158–194) | <0.01* |
| Diastolic blood pressure | 100(88–110) | 100(90–110) | 100(87–110) | 0.301 |
| CT image | | | | |
| Onset to CT, hr | 3(2–5) | 3(2–6) | 2(2.5–4) | <0.01* |
| Initial hematoma volume, ml | 22.64 (11.74–39.48) | 15.6 (7.35–28.49) | 27.68 (15.18–46.44) | <0.01* |
| ICH extension to ventricles, n (%) | 614(30.85 %) | 147(20.65 %) | 467(36.54 %) | <0.01* |
| Treatment modality | | | | <0.01* |
| CMT | 1208(60.70 %) | 604(84.83 %) | 604(47.26 %) | – |
| CC | 118(5.93 %) | 18(2.53 %) | 100(7.82 %) | – |
| MIA | 664(33.37 %) | 90(12.64 %) | 574(44.91 %) | – |

Abbreviations: GOS, Glasgow outcome scale; GCS, Glasgow coma scale; ICH, Intracerebral hemorrhage; ICH-GS, ICH grading scale; CMI, conservative medical treatment; CC, conventional craniotomy; MIA, minimally invasive approach. Minimally invasive approach includes minimally invasive puncture surgery and endoscopic surgery.
* Indicates P value < 0.05.

### 3.4. Evaluation of the ICH score and ICH-GS score

Table 4 and Table 5 show the poor outcome rates for each ICH/ICH-GS score rank. It can be seen that the poor outcome rates increased with higher ICH and ICH-GS scores; all of the patients with an ICH score of 5 or an ICH-GS score of 13 were found to have poor neurological outcomes. Fig. 4 shows that the GOS scores decreased with higher ICH and ICH-GS scores. Using the whole data set (n = 1990), the two models were not found to differ significantly in terms of their predictive ability

**Table 2**
Multivariable logistic regression analysis of patients with poor prognosis.

| Variable | P value | OR | OR 95 % CI |
|---|---|---|---|
| Age | <0.01* | 1.042 | 1.030–1.055 |
| GCS on admission | <0.01* | 0.682 | 0.639–0.727 |
| ICH score on admission | 0.351 | 1.123 | 0.880–1.432 |
| ICH-GS score on admission | <0.01* | 0.718 | 0.608–0.848 |
| Systolic blood pressure | <0.01* | 1.005 | 1.001–1.010 |
| Comorbidities | | | |
| History of hemorrhagic stroke | <0.01* | 1.909 | 1.209–3.017 |
| History of ischemic stroke | 0.541 | 1.142 | 0.746–1.748 |
| Anti-platelet therapy | 0.294 | 1.300 | 0.796–2.121 |
| Alcohol intake | 0.263 | 1.172 | 0.888–1.547 |
| CT image | | | |
| Onset to initial CT | 0.244 | 0.999 | 0.998–1.000 |
| Initial hematoma volume | <0.01* | 1.018 | 1.011–1.026 |
| ICH extension to ventricles | 0.046* | 1.425 | 1.006–2.018 |
| Treatment modality (compared to CMI) | <0.01* | – | – |
| CC | <0.01* | 0.312 | 0.231–0.420 |
| MIA | 0.930 | 1.029 | 0.547–1.936 |

Abbreviations: GCS, Glasgow coma scale; ICH, Intracerebral hemorrhage; ICH-GS, ICH grading scale; CMI, conservative medical treatment; CC, conventional craniotomy; MIA, minimally invasive approach.

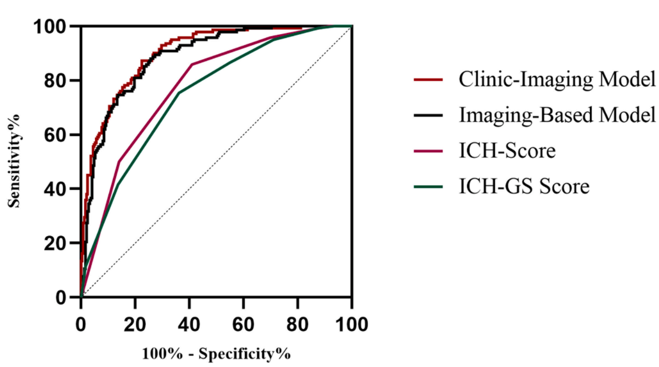


**Fig. 3.** Receiver operating characteristic curves of the different prognostic models for identifying patients with a poor outcome at discharge. ICH: intracerebral hemorrhage; ICH-GS: ICH grading scale.

(AUC: ICH-Score 0.733 compared with ICH-GS Score 0.723, P = 0.189; see ROC curves in Fig. 5).

## 4. Discussion

In this study, two CNN-based models were constructed to predict the neurological outcome of ICH patients at discharge. Nested 5-fold cross-validation demonstrated the accuracy and stability of the classification, as well as the generalizability of the results (see Supplemental materials). Both the clinic-imaging fusion model and the simple imaging-based model were shown to have superior predictive efficiency than the original ICH scale and the ICH-GS scale.

In recent years, an ICH prognostic prediction model has been developed using radiomics combined with machine learning methods [15,16]. However, no other studies have used deep-learning to construct a prognostic prediction model. To the best of our knowledge, this study is the first to show that a clinic-imaging fusion CNN model is better able to predict the prognosis of ICH patients than the traditional risk stratification scales. However, both the ICH scale and the ICH-GS were found to mirror prognostic trends in ICH patients to some extent, and the two did not differ significantly in terms of their predictive efficiency.

In this study, we're able to verify the predictive efficiency of the ICH scale and the ICH-GS for the short-term neurological outcomes of 1990 ICH patients. The AUCs were found to be 0.733 and 0.723, respectively, which is lower than in previous reports [28–30]. Our results are likely be more accurate than previous studies, as we were able to overcome several limitations: (1) the previous studies had limited sample sizes and lacked robust external validation; (2) the ICH volume was previously calculated using the traditional Tada formula, which can be inaccurate [31]; (3) the use of different targets (e.g., survival outcomes, short-term/long-term neurological outcomes) may have affected the previous results. These limitations may all have affected the accuracy of the traditional grading scale models [32].

Neuroimaging data contains important information for determining a patient's prognosis. These features include the severity of a midline shift [3], perihematomal edema [33,34], and enlarged ventricles [3]. In addition, markers of the ICH shape and density can predict hematoma expansion [35], which is associated with poor outcomes [36]. Traditional rating scales have typically been limited to one or two imaging features (e.g., ICH volume) that are easy to quantify. We reasoned that a predictive model that includes the full range of relevant image features would represent an improvement over the traditional grading scales. In this study, we successfully constructed a clinic-imaging fusion prognostic prediction model based on 1990 SICH patients. Our model is more suitable for emergency settings because it does not require manual segmentation of the ICH images. The model was better able to predict neurological outcomes at discharge than the traditional grading scales. This was also found for our imaging-based CNN model. Furthermore, the fusion model had a slightly, but not significantly higher AUC than the imaging-only model, thus demonstrating that the clinical information, as derived from the EHR, is important. Indeed, in an emergency setting,

**Table 4**
Poor outcome rates for the ICH score for each score rank.

| ICH Score | 0 n = 587 | 1 n = 593 | 2 n = 463 | 3 n = 257 | 4 n = 77 | 5 n = 13 |
|---|---|---|---|---|---|---|
| percentage of poor outcome cases, n (%) | 235 (40.03) | 357 (60.20) | 365 (78.83) | 234 (91.05) | 74 (96.10) | 13 (100) |

Poor outcome were defined as cases with GOS Score ≤ 3.

**Table 3**
The four models' performance on test set.

| Model | AUC 95 % CI | Sensitivity 95 % CI | Specificity 95 % CI | PPV 95 % CI | NPV 95 % CI | *P* Value |
|---|---|---|---|---|---|---|
| Clinic-Imaging Model | 0.903 (0.869–0.930) | 78 (72–83) | 87 (81–92) | 91.47 (87.39–94.32) | 68.89 (63.55–73.77) | <0.01 |
| Imaging-Based Model | 0.886 (0.850–0.916) | 74 (68–80) | 88 (82–93) | 91.58 (87.38–94.48) | 66.14 (61.06–70.87) | <0.01 |
| ICH-Score | 0.777 (0.733–0.818) | 59 (53–65) | 86 (79–91) | 88.02 (82.86–91.79) | 54.46 (50.39–58.48) | <0.01 |
| ICH-GS Score | 0.747 (0.701–0.790) | 64 (58–70) | 75 (67–82) | 81.96 (77.05–86.00) | 54.32 (49.58–58.98) | <0.01 |

Abbreviations: AUC indicates receiver operator characteristic area under the curve; CI, Confidence interval; PPV, Positive Predictive Value; NPV, Negative Predictive Value; ICH-GS, ICH grading scale.

**Table 5**
Poor outcome rates for the ICH-GS score for each score rank.

| ICH-GS Score | 5 n = 93 | 6 n = 426 | 7 n = 531 | 8 n = 335 | 9 n = 265 | 10 n = 190 | 11 n = 92 | 12 n = 50 | 13 n = 8 |
|---|---|---|---|---|---|---|---|---|---|
| percentage of poor outcome cases, n (%) | 33 (35.48) | 188 (44.13) | 277 (52.16) | 262 (78.20) | 213 (80.37) | 167 (87.89) | 82 (89.13) | 48 (96) | 8 (100) |

poor outcome were defined as cases with GOS Score ≤ 3.

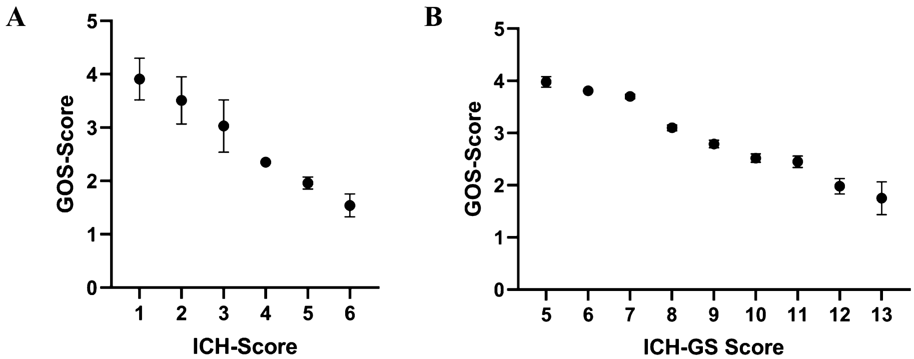


**Fig. 4.** Graphs showing the ICH-Score (A) and the ICH-GS Score (B) against the GOS Score for the whole cohort.

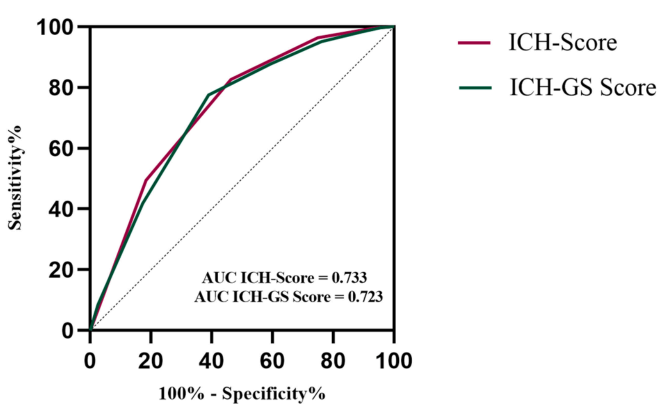


**Fig. 5.** Receiver operating characteristic curves of the ICH Score and the ICH-GS for identifying patients with a poor outcome at discharge, based on the whole cohort. ICH: intracerebral hemorrhage; ICH-GS: ICH grading scale.

it is well known that factors such as age and medical history play a vital role in resisting the shock of stressful events. Our fusion model used a GMU to combine the imaging and clinical data; this is a novel method for multimodal learning based on gated neural networks [26]. Whereas previous multimodal models mainly focused on mapping from one modality to another to construct a common representation [37,38], the GMU combines multiple sources of information to optimize the final goal, thus allowing it to perform different tasks in different networks. Compared to machine learning methods (or radiomics combined) that require accurate delineation of the region of interest for ICH, our model automatically preprocesses CT images by performing data clipping, skull removal, window width/level settings, data augmentation, and rotation. Using the DensNet network combined with the GMU module, our model automatically extracts and analyzes imaging and clinical features to output the probability of poor outcomes, completing the prediction in less 1 second per case. However, it is inevitably difficult to explain the image features selected by the deep learning model. Although visualization techniques may be helpful in understanding the reasoning behind the model's decision-making process, constructing a deep learning model with interpretability remains a significant challenge.

The present study showed that the main elements in the original ICH score (age, size of the hematoma, ICH extension to the ventricles, and GCS score) were independent risk factors for a poor neurological outcome at discharge, a finding that was not unexpected. The present study revealed that patients who accepted surgery (CC, but not MIA) had a poorer prognosis. This is consistent with the STICH I/II trials [39,40], which found that early surgery did not improve the overall neurological outcome in ICH patients. However, as the present study was retrospective, it is possible that different factors could have been confounded. With the rapid development of new surgical instruments (e.g., neuronavigation systems, the SurgiScope, the neuro evacuation device) [41] and updated surgical approaches (e.g., with a shorter time from onset to evacuation) [42], it can be anticipated that the potential value of minimally invasive surgery for ICH will soon be recognized.

Our new clinic-imaging fusion CNN model could be used clinically: (1) as a risk stratification tool, as the model can predict the short-term outcomes of ICH patients in emergency settings; (2) as a guide for clinical decision making, so that appropriate treatments can be chosen and medical resources can be optimally allocated (e.g., to determine the prognosis for a patient who accepts hematoma removal); (3) to promote communication between clinicians and patients, and so facilitate cooperation. However, the model could be further optimized. For example, the model could incorporate laboratory materials and data related to hematoma expansion, and separate models could be constructed for different time points following stroke onset (e.g., within 24 h of onset, at 24–72 h following onset, 72 h after onset, etc.). The model was based on data from patients following a strict screening procedure, which inevitably caused many patients to be excluded for poor-quality or missing data. Furthermore, using the GOS instead of the modified ranking scale (mRS) to assess the outcome might be a potential limit because the mRS is more common to report in hemorrhagic stroke. Given that the GOS assesses disability and social participation and is the profound cited outcome measure in studies on brain injury [43], we considered using the GOS are equally applicable for using the mRS. Finally, to ensure the robustness of our model, we used the traditional 5-fold cross-validation method instead of creating an external validation dataset. It would be better to divide the dataset by centers rather than using the current random division approach. In the future, it will be important to verify the clinic-imaging fusion model's convenience and accuracy in real-life clinical settings. This will better determine the potential clinical value of this novel prognostic prediction model, based on artificial intelligence technology.

## 5. Conclusions

In this work, our team used a deep learning-based method to construct a short-term prognostic prediction model for ICH patients. The predictive performance was compared with the original ICH scale and the ICH-GS using a large cohort. The results indicated that the clinic-imaging fusion CNN model has the best predictive efficiency. The model could be used in a clinical setting to improve the management of patients with ICH.

## Data availability

The used code and all data generated or analyzed during this study are available from the corresponding author on reasonable request.

## Ethics approval

This retrospective study was approved by the Institutional Review Board of Peking Union Medical College Hospital (PUMCH; Ethics code: S-K1175).

## Informed consent

For this study, formal consent was not required.

## Sources of funding

This work was supported by the National High Level Hospital Clinical Research Funding (2022-PUMCH-C-032).

## CRediT authorship contribution statement

**Yihao Chen:** Writing – original draft, Methodology, Conceptualization, Validation. **Cheng Jiang:** Visualization, Software, Investigation, Writing – original draft. **Jianbo Chang:** Conceptualization, Validation. **Chenchen Qin:** Data curation, Investigation. **Qinghua Zhang:** Resources. **Zeju Ye:** Resources. **Zhaojian Li:** Resources. **Fengxuan Tian:** Resources. **Wenbin Ma:** Supervision. **Ming Feng:** Supervision, Formal analysis. **Junji Wei:** Writing – review & editing, Supervision, Funding acquisition. **Jianhua Yao:** Writing – review & editing, Software, Methodology. **Renzhi Wang:** Writing – review & editing, Supervision, Project administration.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank the Chinese Intracranial Hemorrhage Image Database (CICHID) and the Group of Medical Data, Chinese Medical Doctor Association for data support and assistance. We gratefully acknowledge the financial supports by the National High Level Hospital Clinical Research Funding under grant numbers 2022-PUMCH-C-032.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrad.2023.111081.

## References

[1] M. Ng, T. Fleming, M. Robinson, et al., Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013, Lancet 384 (2014) 766–781.
[2] H.S. Chen, C.F. Hsieh, T.T. Chau, C.D. Yang, Y.W. Chen, Risk factors of in-hospital mortality of intracerebral hemorrhage and comparison of ICH scores in a Taiwanese population, Eur. Neurol. 66 (2011) 59–63.
[3] R. Bhatia, H. Singh, S. Singh, et al., A prospective study of in-hospital mortality and discharge outcome in spontaneous intracerebral hemorrhage, Neurol. India 61 (2013) 244–248.
[4] C.J. van Asch, M.J. Luitse, G.J. Rinkel, I. van der Tweel, A. Algra, C.J. Klijn, Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis, Lancet Neurol. 9 (2010) 167–176.
[5] J.J. Loan, A.B. Gane, L. Middleton, et al., Association of baseline hematoma and edema volumes with one-year outcome and long-term survival after spontaneous intracerebral hemorrhage: A community-based inception cohort study, Int. J. Stroke (2020) 322749238.
[6] J.R. Hemphill, D.C. Bonovich, L. Besmertis, G.T. Manley, S.C. Johnston, The ICH score: a simple, reliable grading scale for intracerebral hemorrhage, Stroke 32 (2001) 891–897.
[7] J.R. Hemphill, M. Farrant, T.J. Neill, Prospective validation of the ICH Score for 12-month functional outcome, Neurology 73 (2009) 1088–1094.
[8] F.A. Schmidt, E.M. Liotta, S. Prabhakaran, A.M. Naidech, M.B. Maas, Assessment and comparison of the max-ICH score and ICH score by external validation, Neurology 91 (2018) e939–e946.
[9] R.T. Cheung, L.Y. Zou, Use of the original, modified, or new intracerebral hemorrhage score to predict mortality and morbidity after intracerebral hemorrhage, Stroke 34 (2003) 1717–1722.
[10] J.L. Ruiz-Sandoval, E. Chiquete, S. Romero-Vargas, J.J. Padilla-Martinez, S. Gonzalez-Cornejo, Grading scale for prediction of outcome in primary intracerebral hemorrhages, Stroke 38 (2007) 1641–1644.
[11] K.J. Becker, A.B. Baxter, W.A. Cohen, et al., Withdrawal of support in intracerebral hemorrhage may lead to self-fulfilling prophecies, Neurology 56 (2001) 766–772.
[12] R. Faigle, E.B. Marsh, R.H. Llinas, V.C. Urrutia, R.F. Gottesman, Race-Specific Predictors of Mortality in Intracerebral Hemorrhage: Differential Impacts of Intraventricular Hemorrhage and Age Among Blacks and Whites, J. Am. Heart Assoc. 5 (2016).
[13] E. Heeley, C.S. Anderson, M. Woodward, et al., Poor utility of grading scales in acute intracerebral hemorrhage: results from the INTERACT2 trial, Int. J. Stroke 10 (2015) 1101–1107.
[14] S. Lukic, Z. Cojbasic, Z. Peric, et al., Artificial neural networks based early clinical prediction of mortality after spontaneous intracerebral hemorrhage, Acta Neurol. Belg. 112 (2012) 375–382.
[15] H.L. Wang, W.Y. Hsu, M.H. Lee, et al., Automatic Machine-Learning-Based Outcome Prediction in Patients With Primary Intracerebral Hemorrhage, Front. Neurol. 10 (2019) 910.
[16] X. Xu, J. Zhang, K. Yang, Q. Wang, X. Chen, B. Xu, Prognostic prediction of hypertensive intracerebral hemorrhage using CT radiomics and machine learning, Brain Behavior (2021).
[17] G. Celik, O.K. Baykan, Y. Kara, H. Tireli, Predicting 10-day mortality in patients with strokes using neural networks and multivariate statistical methods, J. Stroke Cerebrovasc. Dis. 23 (2014) 1506–1512.
[18] T. Jo, K. Nho, A.J. Saykin, Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data, Front. Aging Neurosci. 11 (2019) 220.
[19] J. Lao, Y. Chen, Z.C. Li, et al., A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme, Sci. Rep. 7 (2017) 10353.
[20] C. Jin, Y. Jiang, H. Yu, et al., Deep learning analysis of the primary tumour and the prediction of lymph node metastases in gastric cancer, Br. J. Surg. (2020).
[21] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[22] A. Sreekrishnan, J.L. Dearborn, D.M. Greer, et al., Intracerebral Hemorrhage Location and Functional Outcomes of Patients: A Systematic Literature Review and Meta-Analysis, Neurocrit. Care 25 (2016) 384–391.
[23] R. Behrouz, V. Misra, D.A. Godoy, et al., Clinical Course and Outcomes of Small Supratentorial Intracerebral Hematomas, J. Stroke Cerebrovasc. Dis. 26 (2017) 1216–1221.
[24] P.A. Yushkevich, J. Piven, H.C. Hazlett, et al., User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, Neuroimage 31 (2006) 1116–1128.
[25] G. Huang, Z Liu, L. van der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, 2016. http://arxiv.org/abs/1608.06993.
[26] J. Arevalo, T. Solorio, M. Montes-y-Gómez, F.A. González, Gated Multimodal Units for Information Fusion, 2017. http://arxiv.org/abs/1702.01992.
[27] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845.
[28] V.P. Gupta, A. Garton, J.A. Sisti, et al., Prognosticating Functional Outcome After Intracerebral Hemorrhage: The ICHOP Score, World Neurosurg. 101 (2017) 577–583.
[29] S.S. Bruce, G. Appelboom, M. Piazza, et al., A comparative evaluation of existing grading scales in intracerebral hemorrhage, Neurocrit. Care 15 (2011) 498–505.
[30] T. Gregorio, S. Pipa, P. Cavaleiro, et al., Assessment and Comparison of the Four Most Extensively Validated Prognostic Scales for Intracerebral Hemorrhage: Systematic Review with Meta-analysis, Neurocrit. Care 30 (2019) 449–466.

[31] X. Xu, X. Chen, J. Zhang, et al., Comparison of the Tada formula with software slicer: precise and low-cost method for volume assessment of intracerebral hematoma, Stroke 45 (2014) 3433–3435.
[32] J.R. Hemphill, S.M. Greenberg, C.S. Anderson, et al., Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association, Stroke 46 (2015) 2032–2060.
[33] T.Y. Wu, G. Sharma, D. Strbian, et al., Natural History of Perihematomal Edema and Impact on Outcome After Intracerebral Hemorrhage, Stroke 48 (2017) 873–879.
[34] S. Urday, L.A. Beslow, F. Dai, et al., Rate of Perihematomal Edema Expansion Predicts Outcome After Intracerebral Hemorrhage, Crit. Care Med. 44 (2016) 790–797.
[35] C.D. Barras, B.M. Tress, S. Christensen, et al., Density and shape as CT predictors of intracerebral hemorrhage growth, Stroke 40 (2009) 1325–1331.
[36] D. Dowlatshahi, A.M. Demchuk, M.L. Flaherty, M. Ali, P.L. Lyden, E.E. Smith, Defining hematoma expansion in intracerebral hemorrhage: relationship with patient outcomes, Neurology 76 (2011) 1238–1244.
[37] C. Tschoe, C.D. Bushnell, P.W. Duncan, M.A. Alexander-Miller, S.Q. Wolfe, Neuroinflammation after Intracerebral Hemorrhage and Potential Therapeutic Targets, J. Stroke 22 (2020) 29–46.
[38] E.B. Engler-Chiurazzi, K.L. Monaghan, E. Wan, X. Ren, Role of B cells and the aging brain in stroke recovery and treatment, Geroscience 42 (2020) 1199–1216.
[39] A.D. Mendelow, B.A. Gregson, E.N. Rowan, G.D. Murray, A. Gholkar, P. M. Mitchell, Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): a randomised trial, Lancet 382 (2013) 397–408.
[40] A.D. Mendelow, B.A. Gregson, H.M. Fernandes, et al., Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial, Lancet 365 (2005) 387–397.
[41] A.B. Carpenter, J. Lara-Reyna, T. Hardigan, T. Ladner, C. Kellner, K. Yaeger, Use of emerging technologies to enhance the treatment paradigm for spontaneous intraventricular hemorrhage, Neurosurg. Rev. (2021).
[42] C.P. Kellner, R. Song, M. Ali, et al., Time to Evacuation and Functional Outcome After Minimally Invasive Endoscopic Intracerebral Hemorrhage Evacuation, Stroke 52 (2021) e536–e539.
[43] B. Jennett, M. Bond, Assessment of outcome after severe brain damage, Lancet 1 (1975) 480–484.